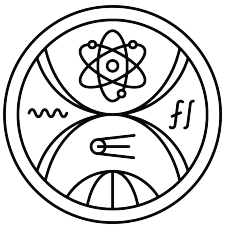
Comenius University in Bratislava
Faculty of Mathematics, Physics and Informatics



Hybrid gene expression models

Dissertation Thesis

Mgr. Iryna Zabaikina 2024

Comenius University in Bratislava
Faculty of Mathematics, Physics and Informatics

Hybrid gene expression models

Dissertation Thesis

Study program:	Applied mathematics
Study field:	22. - mathematics
Department:	Department of applied mathematics and statistics
Supervisor:	doc. Mgr. Pavol Bokes, PhD.

Mgr. Iryna Zabaikina

Bratislava, 2024

I would like to express my gratitude to my supervisor, doc. Mgr. Pavol Bokes, PhD., for his guidance, advice, and patience over the past five years. I am also deeply thankful to my family and my husband, Pavlo, for their love and unwavering support throughout this period. Finally, I would like to thank Loki, who has been my emotional support on this journey.

Comenius University in Bratislava
Faculty of Mathematics, Physics and Informatics

THESIS ASSIGNMENT

Name and Surname:	Mgr. Iryna Zabaikina
Study programme:	Applied Mathematics (Single degree study, Ph.D. III. deg., full time form)
Field of Study:	Mathematics
Type of Thesis:	Dissertation thesis
Language of Thesis:	English
Secondary language:	Slovak

Title: Hybrid gene expression models

Annotation: In living cells, genes contain information that is expressed into product molecules, such as RNAs and proteins. Some of these molecular species are present in low copy numbers, whereas others can be produced in large quantities. Such disparities in abundances of molecular species make the use of hybrid models in gene expression particularly attractive. A hybrid model combines a set of discrete species that is subject Markovian stochastic dynamics and a set of continuous species governed by deterministic dynamics. The aim of the PhD project will be to examine hybrid models of specific systems, in particular ones that include feedback control via transcriptional as well as post-transcriptional regulation. A combination of stochastic simulation and mathematical analysis will be put to use to characterise the behaviour of such models and draw conclusions with respect to the underlying biological phenomena.

Tutor:	doc. Mgr. Pavol Bokes, PhD.
Department:	FMFI.KAMŠ - Department of Applied Mathematics and Statistics
Head of department:	prof. RNDr. Marek Fila, DrSc.

Electronic version available:
archív

Assigned: 07.01.2020

Approved: 08.01.2020 prof. RNDr. Daniel Ševčovič, DrSc.
Guarantor of Study Programme

.....
Student

.....
Tutor

Abstract

ZABAIKINA, Iryna. *Hybrid gene expression models* [dissertation thesis]. Comenius University Bratislava. Faculty of Mathematics, Physics and Informatics; Department of applied mathematics and statistics. Supervisor: doc. Mgr. Pavol Bokes, PhD.. Bratislava, 2024.

Regulation of gene expression is represented by a variety of control motifs, mathematical models of which can provide a theoretical estimate of the process parameters. In this project, we study three particular examples of regulatory networks. The first one is negative feedback when mRNA indirectly inhibits its production. The second one is an incoherent feed-forward loop, which is represented by the interaction between mRNA and antagonistic microRNA. We construct a generalized hybrid model using a Markovian drift-jump framework with random production bursts and continuous degradation. Combined with the Chapman-Kolmogorov equation, it provides the means to determine the probability distribution of mRNA concentration. We derive the mean steady-state concentration of mRNA for both models. Subsequently, we show that it is less sensitive to the production rate in the feed-forward loop than in the negative feedback. In addition, it turns out that in presence of the low noise, FFL maintains the concentration of mRNA at a steady level despite disturbance in production rate, i.e. is perfectly adaptating. Finally, the third one is the positive feedback on dilution when the protein inhibits cell growth. We model a single cell using the drift-jump framework, then develop a population model using a measure-valued Markov process combined with the population balance equation. We show that this type of regulation causes a difference between single-cell and population protein distributions. We also demonstrate that the nature of the division mechanism, whether stochastic or deterministic (sizer), does not affect the protein distribution.

Keywords: gene expression, hybrid model, negative feedback, feed-forward loop, perfect adaptation, feedback on dilution, population model

Abstrakt

ZABAIKINA, Iryna. *Hybrid gene expression models* [dizertačná práca]. Univerzita Komenského v Bratislave. Fakulta matematiky, fyziky a informatiky; Katedra aplikovanej matematiky a štatistiky. Vedúci práce: doc. Mgr. Pavol Bokes, PhD.. Bratislava, 2024.

Regulácia génovej expresie je reprezentovaná rôznymi kontrolújucimi motívmi, takže ich matematické modely môžu poskytnúť teoretický odhad parametrov procesu. V tomto projekte skúmame tri konkrétne príklady regulačných sietí. Prvým príkladom je negatívna spätná väzba, keď mediátorová RNA (mRNA) nepriamo inhibuje vlastnú syntézu. Druhým príkladom je nekoherentná dopredná slučka, ktorá je reprezentovaná interakciou medzi mRNA a antagonistickou mikroRNA. Pomocou Markovovskeho drift-jump frameworku zostrojíme zovšeobecnený hybridný model s náhodnými produkčnými pulzmi a spojitou degradáciou. V kombinácii s Chapmanovou-Kolmogorovovou rovnicou, taký model poskytuje pravdepodobnostné rozdelenie koncentrácie mRNA. Odvodíme priemernú stacionárnu koncentráciu mRNA pre obidva modely. Následne ukazujeme, že dopredná slučka je menej citlivá na rýchlosť produkcie ako negatívna spätná väzba. Okrem toho sa ukazuje, že v prítomnosti nízkeho šumu dopredná slučka udržiava koncentráciu mRNA na stabilnej úrovni napriek výrazným kolísaniam rýchlosti produkcie, t.j. sa prispôsobuje dokonale. Nakoniec, tretím príkladom je pozitívna spätná väzba na zriedenie, keď proteín inhibuje rast buniek. Pre modelovanie jednotlivej bunky použijeme drift-jump framework, následne zostrojíme model bunkovej populácie pomocou merateľného Markovovho procesu v kombinácii s rovnicou rovnováhy populácie. Ukazujeme, že tento typ regulácie spôsobuje rozdiel medzi rozdelením koncentrácie proteínu v jednotlivej bunke a v populácii. Tiež demonštrujeme, že typ bunkového delenia, či stochastický alebo deterministický (sizer), neovplyvňuje rozdelenie proteínu.

Kľúčové slová: génová expresia, hybridný model, negatívna spätná väzba, dopredná slučka, dokonalé prispôsobenie, spätná väzba na zriedenie, model populácie

Acronyms

DNA	deoxyribonucleic acid
RNA	ribonucleic acid
RNAp	RNA polymerase
mRNA	messenger RNA
miRNA	microRNA
TF	transcriptional factor
UTR	untranslated region
FFL	feed forward loop
IFFL	incoherent feed forward loop
NFB	negative feedback
PFB	positive feedback
ODE	ordinary differential equation
PDE	partial differential equation
PDF	probability density function
CCDF	complementary cumulative distribution function
SC	single cell framework
POP	population framework
LT	Laplace transform
PBE	population balance equation
CKDE	Chapman-Kolmogorov differential equation

List of notations

The following list compiles the most commonly used notation in the main text.

Constants and functions

$\Lambda(x)$	protein-dependent burst frequency
α	(constant) burst frequency
β	mean burst size
δ	hazard rate of mRNA-miRNA interaction
γ	(constant) degradation rate
$\gamma(x)$	protein-dependent degradation rate
λ	population growth rate
k	feedback intensity

Chemical kinetics

$\emptyset$	Species are degraded or do not affect the studied system
B	Burst production
X	The species of interest
Y	The auxiliary species (e.g. miRNA)

Contents

Introduction

The cell is the fundamental building block of every living organism, and biochemical reactions that occur inside it define life at a molecular level. All information about these reactions is stored encrypted in DNA – a long amino acid chain, which contains functional units called *genes*. Each gene provides instructions for a functional product (protein or RNA), which support cell development, adaptation to inner and outer conditions, proliferating, and overall functionality of a cell and, thus, an organism. The central process in a cell is *gene expression*, which is decoding genes and synthesis of corresponding functional product. It is complex and strictly regulated mechanism, which is yet to be fully understood.

The aim of this work is to study the gene expression regulation. This dynamic process is crucial for cellular function and adaptability, and can be observed in the form of different regulatory circuits. These circuits play a key role in the modulation of gene activity, allowing cells to respond efficiently to both internal and external changes. We focus on two types of regulatory mechanisms. First one is a feedback loop, where the output of a process governs its own activity either inhibiting it (negative feedback) or enhancing it (positive feedback), which arises at transcriptional stage [1]–[3], post-translational stage [4], or embedded in cell size control [5]. Second one is a feed-forward loop, in which input signal enhances output signal through one path and dampens it through another one.

Advance of experimental techniques allows to expand our knowledge of the intracellular reactions. First, they improve experimental data quality, which allows to either build more precise models or validate existing ones. Second, they made possible to reveal previously unknown aspects of cellular mechanisms. Finally, they provide not only theoretical data, but also practical information on disease progression and treatment outcomes at the cell level. Thus, understanding and altering regulation of gene expression are useful for such applied fields as synthetic biology and pharmacogenomics, which are based on results from both mathematical models and practical sides of research of gene expression. Synthetic biology uses predictions by mathematical models to validate hypotheses before costly empirical tests. At the same time, mathematical models can reveal features and characteristics, which may help in identifying regulatory loops in experimental data. Gene expression has

significant role in in drug response [6], [7]; understanding the influence of regulation on both single cell and cell colony can help explain how diseases affect expression of certain genes, predict population-wide treatment outcomes [8], and address drug tolerance issues [9]. Overall, integrative approach not only advances theoretical understanding but also enhances laboratory practices by informing experiment design and interpretation.

Extensive research has already been conducted in the field of gene expression regulation, employing a variety of modelling techniques to uncover different aspects of regulatory networks. Fundamentally, there are the following principal approaches. The deterministic approach uses chemical kinetics to provide the framework for quantifying the rates of chemical reactions within these models; it results into systems of ordinary differential equations. Dynamical systems are used to capture key features of the process based on experimental data [10] and steady state behaviour [11], [12]. The key assumption in these models is that the involved species are well-mixed and molecular interactions occur uniformly throughout the cellular environment. It is also their weak side, since it might not be true in cases of spatial heterogeneity or low molecule numbers where stochastic effects are significant.

Discrete stochastic models allow to explore gene expression at the level of individual events and interactions, which is valuable in case of gene expression, where stochastic effects play a critical role. The temporal evolution of species in such systems is expressed in terms of a chemical master equation (CME) [13], [14]. A major solution method is the Linear Noise Approximation (LNA), which is widely used for quantifying gene expression noise [15]–[17].

Finally, in this work we focus on hybrid models, in particular on ones based on the piecewise deterministic Markov process [18]. They typically model the process as mainly deterministic with random discrete events that change the system's state, like gene switching [19], [20], and bursty production [21]. This approach can also be successfully extended to explore cell populations [22]–[24] without a drastic expansion of the model.

The work is structured as follows. In Chapter 1 we provide a detailed overview of the biological side of the work; we describe biochemical mechanisms that underlie such regulatory circuits as transcriptional autoregulation, feed-forward loop, and feedback on cellular growth dilution. In the Chapter 2, we develop the mathematical background for the following work. This chapter begins with mathematical constructions, which are frequently used throughout the work. Specifically, Sections 2.2 and 2.1 are about the Laplace transform and special functions, respectively. In Sections 2.3–2.5 we build the general framework. We start with an evolution of a probability density function in a deterministic system. Afterwards, we extend it with random jumps and derive a general master equation for a single cell. Next, we

introduce population framework and the population balance equation. The following chapters are dedicated to specific regulatory circuits. In Chapter 3, we study the feedback on burst frequency in single cell and population models. In Chapter 4, we study the behaviour of the feed-forward loop in a low-noise regime (when the species are produced in small and frequent bursts). This is followed by an analysis of a general model in a steady state, and a conclusion that summarises our current results. Finally, in last two chapters we study the model of the cell, in which the cell growth rate is affected by the presence of a certain protein. In Chapter 5, we study the case of a protein that inhibits the cell growth. The higher protein concentrations is, the slower the cell expansion becomes, which preserves higher protein concentrations; hence, positive feedback occurs. We construct and analyse both standard model of protein concentration and a bivariate model, which includes the cell volume. In Chapter 6, we consider the opposite case with negative feedback on dilution. In last chapter, we discuss results, limitations and future directions of the research.

Chapter 1

Biological background

Gene expression

A cell life cycle is primarily governed by changes in a wide variety of *functional products*, which are predominantly proteins, along with various types of ribonucleic acids (RNAs). These molecules are crucial for carrying out essential activities such as cell growth, division, response to external changes, storing and production of energy, etc.

Deoxyribonucleic acid (DNA) is central molecule in each living cell that stores all information about the structure of functional products and instructions for their usage. Structurally, DNA is a double helix, each strand of which consists of building bases called *nucleotides*: adenine (A), cytosine (C), guanine (G), and thymine (T). The base pairs (A) $\rightleftarrows$ (T) and (C) $\rightleftarrows$ (G) are mutually complementary, e.g. if one strand contains a sequence (ACTG), then the other one contains (TGAC).

A segment of DNA that contains the complete instructions for building a specific protein or RNA is called a *gene*. This region (or regions) includes all elements necessary to encode a functional molecule: a directly coding region, regulatory sequences (including promoter and termination signal), and auxiliary non-coding parts [25], [26]; a scheme of a gene is shown in Figure 1.1.

Gene expression is the process, during which the information in the gene is decoded and used to manage the synthesis of a gene product. This process is intricate; some optional steps can differ significantly across different organisms, cell types, external and internal conditions [27]. However, the following three consistent steps are mainly obligatory [28] for protein synthesis:

- transcription (where separate copy of a gene – mRNA – is produced);
- translation (where mRNA serves as a template for protein synthesis);
- post-translational modifications (where protein becomes active).

Let us take a closer look at each of aforementioned processes. The transcription begins at a promoter, which determines the starting site, as shown in Figure 1.1. It

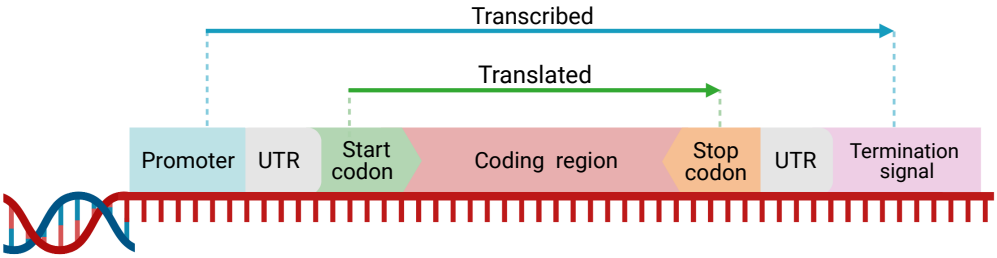


Figure 1.1: Structural parts of the gene in DNA, which are involved in different steps of producing proteins (the figure was created with BioRender.com)

can typically be in two states: active, where it is accessible to the specific transcription machinery called RNA polymerase (RNAp), or inactive, where it is inaccessible and transcription. When RNA polymerase binds to DNA, gene expression is initiated. First, RNAp unwinds the DNA strands, providing access to the single strand, which serves as a template for further synthesising a complementary RNA chain. Then the elongation begins: RNAp moves along the DNA strand, reads one base and adds one complementary base at a time. The complementary rules are similar to ones above, except it adds urasil (U) instead of (T). Behind the advancing RNA polymerase, DNA naturally rewinds back into its stable structure of a double helix. When the termination signal is reached, RNAp detaches from DNA and releases a produced molecule – messenger RNA (mRNA) – into a cellular space.

After transcription is finished, mRNA is delivered via mobile export receptors to a ribosome, where translation is performed [29]. This means that the ribosome uses mRNA as a template to produce an amino acid chain of the gene product. More precisely, only a part of mRNA between start and stop codons is directly translated, it is marked with the green arrow in Figure 1.1). untranslated regions (UTRs) play critical roles in regulating the efficiency and stability of this process, thereby influencing the amount of protein produced. Intense translation of the mRNA molecule lasts until it degradation. These short periods are called translational bursts, throughout which a random number of the amino acid chains are synthesised.

The last step is a post-translational modification, within which the amino acid chains undergo changes affecting their structure, location, or stability. After the protein or RNA is used follows their elimination. The process is called degradation, during which the functional product is split into single amino acids or smaller amino acid chains because of time or enzymes.

Each of these steps is tightly regulated by a complex network of signalling pathways and regulatory proteins, ensuring that the right genes are expressed at the right times and in the right amounts.

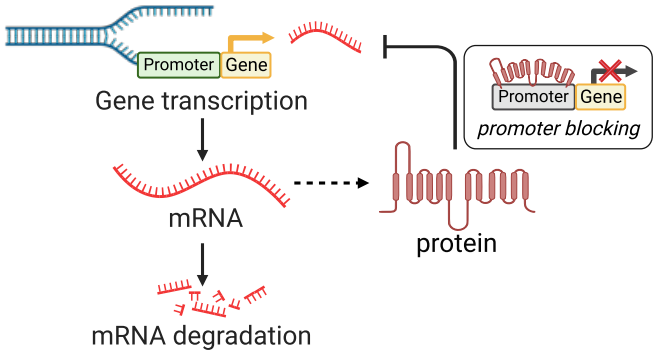


Figure 1.2: Scheme of the protein synthesis with implemented negative autoregulation; bold arrows correspond to the set reactions studied in Section 3 (inhibiting arrow corresponds to the regulatory function $\Lambda(x)$).

Regulation using transcriptional factor (TF)

Direct way to regulate the amount of protein is to control its mRNA concentration in a cell by changing accessibility of the promoter to RNA polymerase. There are two principal ways of implementation: altering chemical structure of the promoter or using auxiliary molecules, which can influence promoter accessibility. We focus on a second one, specifically on proteins called *transcriptional factors* (TFs); depending on the type, they can enhance, reduce, or stop gene transcription. They implement a regulatory mechanism called feedback; its fundamental principle is self-regulation. At the transcriptional stage, positive feedback (PFB) implies that a protein enhances its own transcription to increase its concentration in the cell; negative feedback (NFB) implies the capability of the protein to reduce the rate of its own expression by preventing transcription [30]. This dynamic balance ensures that protein levels within the cell can quickly response to changing internal and external conditions.

We focus on a particular case of proteins that are inhibitory transcriptional factors for their own gene promoter; thereby, production explicitly depends on the current concentration of the protein in the cell (Figure 1.2).

Many properties, which are provided by the feedback loop (speed of response, sustained mean concentration, etc.), are characterised by its response function, which determines the transcription rate of the gene as a function of the protein concentration in the cell. Many studies use a sigmoid response function, such as Hill function [31]. The feedback loop may be implemented in different ways; the most widely arising are reducing the frequency of bursts or mean size of bursts [32]. For example, the response function affects the concentration via frequency in the following way: if the protein concentration is currently high, then the frequency of bursts decreases or even becomes zero. Otherwise, as concentration decreases, bursts occur more frequently.

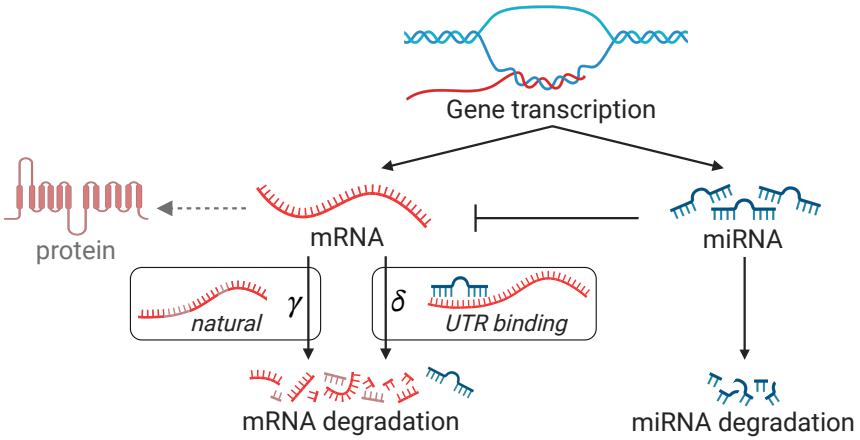


Figure 1.3: Scheme of the protein life cycle, which is regulated by miRNA; bold arrows correspond to the set reactions studied in Section 4 (the figure was created with BioRender.com).

Regulation using feed forward loop (FFL)

Another type of regulatory circuits was found after qualitative improvement of laboratory technologies, which gave specific information about presence of auxiliary gene products [33], [34]. In particular, it was discovered that some RNAs, called non-coding RNA, are released into the cell immediately after transcription, avoiding further stages of processing. Mainly, they have support functions such as transferring amino acids (tRNA), forming ribosomes (rRNA), supporting splicing (snRNA), maintaining spermatogenesis (piRNA) [35]–[37].

Our topic of interest is micro RNA (miRNA) – short (approximately 21 nucleotides) non-coding RNA, which is engaged in transcriptional and post-transcriptional regulation of gene expression [38], [39]. Each miRNA is complementary, i.e. has perfect pairing, to a specific type of an untranslated region (UTR) of an mRNA molecule [40], [41]. MiRNA recognises the target UTR and binds to it, causing degradation or suppressing further translation of the mRNA molecule (Fig. 1.3). In essence, microRNAs can be classified into two fundamental types based on the genomic location of their coding sequence. Intergenic miRNAs are located in the non-coding regions in between protein-coding sequences and have their own promoter; thus, they are usually transcribed independently [42], [43]. Intragenic miRNAs, which are located in the intron (intronic miRNA) or the exon (exonic miRNA) of host genes, are co-transcribed in a long transcript precursor (pre-mRNA) and then spliced into separate molecules [44]. In this work, we focus on the intronic miRNAs, the specific role of which is to directly regulate the production of the host gene [45], [46].

A feed forward loop (FFL) is a frequently arising control motif in gene regulatory networks [47]. Its basic principle is to anticipate external disturbance in a controlled

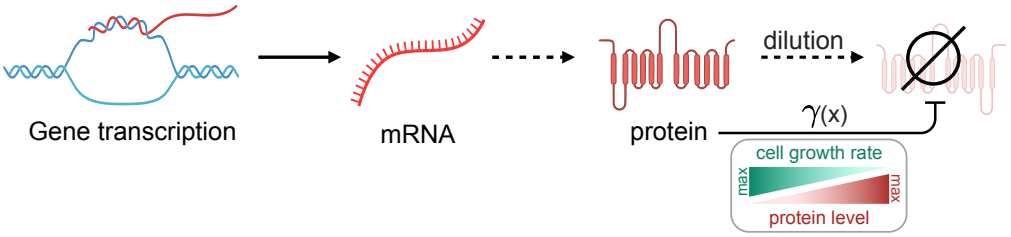


Figure 1.4: Scheme of the protein life cycle, which is positively regulated by dilution rate; bold arrows correspond to the set reactions studied in Section 5.

quantity and immediately provide a counter balancing response. A subtype of FFL, called incoherent FFL (IFFL), appears when a controlled quantity Z is amplified by X and dampened by Y indirectly through X. The miRNA-mRNA interaction is also an IFFL, since an upstream transcriptional activator amplifies the level of mRNA directly and at the same time dampens it indirectly through amplification of its miRNA antagonist [48]. In particular, it is present in transcriptional networks of *S. cerevisiae* [49] and *E. coli* [50], [51]. This control motif was modelled as a deterministic system and proved to be perfectly adaptating [19], [52], i.e IFFL retains expression at a required level regardless of the strength of an upstream signal.

Regulation on cell growth (dilution)

The aforementioned regulatory mechanisms are based on intermolecular interactions between functional products during gene expression process. Studies show that decrease in the protein amount can also be regulated. There are two principal approaches: degrade (or deactivate) molecule of a protein and affect the protein concentration in a cellular environment.

The first approach has different implementations. Damaged proteins can be tagged for degradation, then they become target for proteasome – special molecular mechanism that efficiently degrades proteins [53], [54]. The cell can isolate and deactivate unwanted cellular contents using double lipid membranes. These mechanisms are usually applied to short-lived, heavily damaged [55], [56], and harmful molecules [57].

The second approach to control decay of protein concentration is to manage the cell growth rate and thus rate of protein dilution. The main idea is that the intense protein production affects the rate at which the cell can grow or expand its volume. Typical assumption is that cell is growing exponentially, but the level of a certain protein impacts cell volume growth so it becomes almost linear.

We consider on the case where high protein levels inhibit cell growth. This effect can have various causes, such as protein burden [5], protein-induced stress response [58], and exhaustion of metabolic machinery (i.e., less energy and materials

are available for growth) [59]–[61].

This effect is modelled as a positive feedback mechanism: the more of this protein there is, the slower the cell grows, preserving higher protein concentration. Finally, in this positive feedback scenario, low-protein cells proliferate faster than high-protein cells; it is therefore important to distinguish between the single cell (genealogy) framework and the population framework [62]–[64].

Chapter 2

Mathematical background

In this chapter we introduce some preliminary mathematical background for this work. Section 2.1 contains definitions and properties of the special functions, which might be useful for understanding some parts of a solution approach; in Section 2.2 we provide information about the Laplace transform and double Laplace transform. Throughout Sections 2.3–2.5 we consequentially develop the the single cell and population frameworks. We proceed from the evolution of a probability density function in dynamical systems, then drift-jump framework and finally to the population balance equation. In the last Section 2.6, we present the basic model of the unregulated gene expression, which is fundamental model for each of the following chapters.

2.1 Special functions

In this section we provide definitions, relations and/or integral representations of functions that are mentioned in the text. Special functions do not have a formal definition, but it is a common name for a set of different classes of functions that occur as solutions of both theoretical and applied mathematical problems. They cannot be represented using elementary functions, but as power series, infinite products, integrals, repeated differentiation, etc.

2.1.1 Gamma function and related concepts

Definition 2.1. *Let z be a complex number with positive real part. Then the gamma function is absolutely convergent improper integral:*

$$\Gamma(z) = \int_0^\infty t^{z-1} e^{-t}\,\mathrm{d}t\,,\ z \in \mathbb{C},\ \mathrm{Re}(z) > 0. \tag{2.1}$$

Gamma function is the generalisation of the integer factorial, so it satisfies the recurrent relation $\Gamma(z+1) = z\Gamma(z)$. This relation also provides an analytical contin-

uation of the domain of $\Gamma(z)$ to the whole complex plane except negative integers, where it has simple poles. In case $z \in \mathbb{N}$, it reduces to the factorial function, i.e., $\Gamma(z) = (z-1)!$.

Definition 2.2. *Let $a \in \mathbb{C} \setminus \mathbb{Z}_{\leq 0}$ and $k \in \mathbb{N}$. Then we define the Pochhammer symbol:*

$$(a)_k = \frac{\Gamma(a+k)}{\Gamma(a)}, \quad (a)_0 = 1. \tag{2.2}$$

Definition 2.3. *Let s be a complex number with positive real part and let x be a non-negative real number. Then lower incomplete gamma function $\gamma(a,x)$ and the upper (complementary) incomplete gamma function $\Gamma(a,x)$ are defined as follows:*

$$\gamma(a,x) = \int_0^x t^{a-1}e^{-t}\,\mathrm{d}t\,, \tag{2.3}$$

$$\Gamma(a,x) = \int_x^\infty t^{a-1}e^{-t}\,\mathrm{d}t\,. \tag{2.4}$$

The functions $\gamma(a,x)$ and $\Gamma(a,x)$ are generalisations of the gamma function $\Gamma(x)$ obtained by splitting integral limits in (2.1) at a point $x \geq 0$; clearly, it yields $\Gamma(a,x) + \gamma(a,x) = \Gamma(a)$.

Property 2.1. *The incomplete gamma function has the following properties [65]:*

(a) Recurrence relation:

$$\gamma(a+1,x) = a\gamma(a,x) - x^a e^{-x},$$
$$\Gamma(a+1,x) = a\Gamma(a,x) + x^a e^{-x},$$

which allows to extend domain for parameter a to $\mathbb{C} \setminus \mathbb{Z}_{\leq 0}$.

(b) Special values: $\Gamma(a,0) = \Gamma(a)$, $\gamma(1,x) = 1 - e^{-x}$, and $\Gamma(1,x) = e^{-x}$.

(c) Differentiation with respect to x:

$$\frac{\partial\gamma(a,x)}{\partial x} = x^{a-1}e^{-x}, \quad \frac{\partial\Gamma(a,x)}{\partial x} = -\frac{\partial\gamma(a,x)}{\partial x}.$$

(d) Series expansion:

$$\gamma(a,x) = x^a\Gamma(a)e^{-x}\sum_{k=0}^{\infty}\frac{x^k}{\Gamma(x+k+1)}. \tag{2.5}$$

Definition 2.4. *Let z be any complex number with positive real part. Then the exponential integral is defined as:*

$$E_1(z) = \int_z^\infty \frac{e^{-t}}{t}\,\mathrm{d}t\,, \quad |\operatorname{Arg}(z)| < \pi. \tag{2.6}$$

The exponential integral of a real non-negative value x:

$$Ei(x) = -\int_{-x}^{\infty} \frac{e^{-t}}{t}\,\mathrm{d}t\,,\ x \in \mathbb{R} > 0. \tag{2.7}$$

From (2.6) *and* (2.7) *follows the relation* $E_1(x) = -Ei(-x)$.

Comparing definitions (2.4) and (2.6), it is clear that the exponential integral is a special case of the upper incomplete gamma function:

$$E_1(z) = \Gamma(0, z). \tag{2.8}$$

Thus, it inherits all of aforementioned properties of $\Gamma(a, x)$.

2.1.2 Bessel functions

Definition 2.5. *The Bessel differential equation of order ν is following:*

$$z^2\frac{\mathrm{d}^2 w}{\mathrm{d}z^2} + z\frac{\mathrm{d}w}{\mathrm{d}z} + w(z^2 - \nu^2) = 0, \quad \nu \in \mathbb{C}. \tag{2.9}$$

The Frobenius method is used to solve this equation, i.e., the solution has the power series form:

$$w(x) = \sum_{r=0}^{\infty} a_r z^{\alpha+r}, \quad a_0 \neq 0, \tag{2.10}$$

substitution of which into (2.9) provides that $\alpha = \pm\nu$, a_0 is any constant, and further coefficients a_r are given by recurrent relation:

$$(2\nu + 1)a_1 = 0, \quad a_{r-2} = -r(2\nu + r)a_r, \forall r \geq 2. \tag{2.11}$$

Then we obtain the first fundamental solution of (2.9) called the Bessel function of the first kind.

Definition 2.6. *The Bessel function of the first kind, denoted as $J_\nu(z)$, is defined for a complex variable z by convergent power series [66]:*

$$J_\nu(z) = \sum_{r=0}^{\infty} \frac{(-1)^r (z/2)^{\nu+2r}}{r!\Gamma(\nu + r + 1)}, \quad \nu \in \mathbb{C}. \tag{2.12}$$

If ν is not an integer, J_ν and $J_{-\nu}$ are linearly independent and the solution of (2.9):

$$w(z) = C_1 J_\nu + C_2 J_{-\nu}.$$

Otherwise, as $\nu \in \mathbb{Z}$, they are linearly dependent, because $J_{-\nu} = (-1)^n J_\nu$. The second fundamental solution of (2.9) is the Bessel function of the second kind.

Definition 2.7. *The Bessel function of the second kind, denoted as $Y_\nu(z)$ is defined as follows:*

$$Y_\nu(z) = \frac{\cos(\nu\pi)J_\nu(z) - J_{-\nu}(z)}{\sin(\nu\pi)}, \quad \forall \nu \in \mathbb{C} \setminus \mathbb{Z}. \tag{2.13}$$

In case of integer order n, $Y_n(z)$ is defined by the limit $Y_n(z) = \lim_{\nu\to n} Y_\nu(z)$.

Then the general solution of the Bessel's equation of any order ν (2.9) is following:

$$w(z) = C_1 J_\nu + C_2 Y_\nu.$$

Although the Bessel functions are convergent, they are oscillatory. Thus were developed the modified Bessel functions, which are particularly suited to applications where the solutions are real and involve exponential growth or decay. They arise from the modified Bessel differential equation [67]:

$$z^2\frac{\mathrm{d}^2 w}{\mathrm{d}z^2} + z\frac{\mathrm{d}w}{\mathrm{d}z} - w(z^2 + \nu^2) = 0, \quad \nu \in \mathbb{C}. \tag{2.14}$$

The modified Bessel function of the first kind of order ν is

$$I_\nu(z) := i^{-\nu} J_\nu(iz) = \sum_{r=0}^{\infty} \frac{(z/2)^{\nu+2r}}{r!\Gamma(\nu + r + 1)}. \tag{2.15}$$

$I_\nu(z)$ is defined for all any $z \in \mathbb{C}$. The function $I_\nu(z)$ is always positive for real positive z and grows exponentially as x increases. For any positive ν, $I_\nu(0) = 0$ and has the following asymptotic behaviour for small z:

$$I_\nu(z) \approx \frac{(z/2)^\nu}{\Gamma(\nu + 1)}.$$

Note that I_ν and $I_{-\nu}$ are essentially multiples of J_ν and $J_{-\nu}$, respectively. They are linearly independent in case of non-integer order and form a set of fundamental solutions of (2.14). Otherwise, for the same reason as in (2.13), we present the modified Bessel function of the second kind $K_\nu(z)$ and the general solution of (2.14) is given by

$$w(z) = C_1 I_\nu(z) + C_2 K_\nu(z).$$

Definition 2.8. *The modified Bessel function of the second kind, denoted as $K_\nu(z)$ is defined as follows:*

$$K_\nu(z) = \frac{\pi}{2}\frac{I_{-\nu}(z) - I_\nu(z)}{\sin(\nu\pi)}, \quad \forall \nu \in \mathbb{C} \setminus \mathbb{Z}. \tag{2.16}$$

In case of integer order n, $K_n(z)$ is defined by the limit $K_n(z) = \lim_{\nu\to n} K_\nu(z)$.

The function K_ν is symmetrical with respect to ν ($K_\nu = K_{-\nu}$) and always positive

for positive real z. It has singularity at $z = 0$ and converges to zero, as z increases.

In Chapter 6, we are interested in such integral representations of K_ν that can be expressed as a Laplace image of a real-valued function according to Definition 2.11; we provide the following:

$$K_\nu(z) = \frac{(z/2)^\nu}{2} \int_0^\infty t^{-\nu-1} e^{-t-z^2/4t} \, \mathrm{d}t \,, \tag{2.17}$$

which is the fundamental one in this context, because further representations are mainly based on its transformations (see [68] based on [69]).

2.1.3 Hypergeometric function

Definition 2.9. *The Gauss' hypergeometric differential equation is a second-order linear ordinary differential equation given by*

$$x(1-x)\frac{\mathrm{d}^2 y}{\mathrm{d}x^2} + (c - (a+b+1)x)\frac{\mathrm{d}y}{\mathrm{d}x} - aby = 0, \quad a, b, c \in \mathbb{C},\ c \notin \mathbb{Z}_{\leq 0}. \tag{2.18}$$

The following solution of the hypergeometric equation is obtained using the Frobenius method (2.10), as it was done for the Bessel equation in the previous section.

Definition 2.10. *The hypergeometric function, denoted as ${}_2F_1(a, b; c; x)$, is defined for a complex variable x, where $|x| < 1$, by the power series:*

$${}_2F_1(a, b; c; x) = \sum_{n=0}^{\infty} \frac{(a)_n (b)_n}{(c)_n} \frac{x^n}{n!}, \quad a, b, c \in \mathbb{C},\ c \notin \mathbb{Z}_{\leq 0},$$

where $(\cdot)_n$ is the Pochhammer symbol (2.2)*.*

This series converges absolutely for $|x| < 1$. However, the function ${}_2F_1(a, b; c; x)$ can be analytically continued for $|x| \geq 1$, extending its domain beyond the unit circle in the complex plane.

Property 2.2. *The hypergeometric function ${}_2F_1$ has the following properties [69]:*

(a) Symmetry: ${}_2F_1(a, b, c; x) = {}_2F_1(b, a, c; x)$.

(b) Argument transformation:

$$(1-x)^{a+b-c} \, {}_2F_1(a, b, c; x) = {}_2F_1(c-a, c-b, c; x).$$

(c) Differentiation with respect to x:

$$\frac{\mathrm{d}^n}{\mathrm{d}x^n} \, {}_2F_1(a, b, c; x) = \frac{(a)_n (b)_n}{(c)_n} \, {}_2F_1(a+n, b+n, c+n; x).$$

(d) Integral representation:

$$ {}_2F_1(a,b,c;x) = \frac{1}{B(b,c-b)} \int_0^1 t^{b-1}(1-t)^{c-b-1}(1-tx)^{-a}\,\mathrm{d}t\,, $$

where the beta function B given by the integral $B(a,b) = \int_0^1 t^{a-1}(1-t)^{b-1}\,\mathrm{d}t$.

(e) Special values:

$$ {}_2F_1(0,b,c;x) = 1, $$
$$ {}_2F_1(-1,b,c;x) = 1 - \frac{bx}{c}. $$

If $c = b+1$, the integrand in Property 2.2(d) contains only two power functions and the beta function becomes a fraction, i.e., $B(b,1) = 1/b$. Then we perform substitutions $y = tu$ and $x = -\kappa u$ and obtain [69]:

$$ {}_2F_1(a,b,b+1;-\kappa u) = \frac{b}{u^b} \int_0^u \frac{y^{b-1}}{(\kappa y+1)^a}\,\mathrm{d}y\,, \qquad \mathrm{Re}\{b\} > 0,\ |\arg(\kappa u+1)| < \pi. \tag{2.19} $$

2.2 Laplace transform

The principal difficulty in studying drift-jump dynamics appearing in Sections 2.4–2.5 stems from the nonlocal character of the influx term J_{in}, which in many applications takes the form of an n-dimensional convolution. This clearly motivates the use of the Laplace transform (LT), under which the (nonlocal) convolution is turned into a (local) multiplication operator. In this section we review the well-known and lesser-known properties of the Laplace transform, which are used in this work.

2.2.1 Single Laplace transform

Definition 2.11. *Let $f(t)$ be a integrable function on $[0,\infty)$. The (unilateral) Laplace transform of $f(t)$ is the function $F(s)$, which is given by [70]:*

$$ F(s) = \mathcal{L}\{f(t)\}(s) = \int_0^\infty f(t)e^{-st}\,\mathrm{d}t\,. $$

The Laplace transform is a unique relation between real-valued function $f(t)$, co-called the original function and its image $F(s)$; this and further details concern only functions, for which Laplace transform exists. Let us introduce possible ways to determine it [71].

Theorem 2.1. *Sufficient conditions of the Laplace transform existence.*
The Laplace transform of $f(t)$ exists in sense of absolute convergence, if:

(a) $\int_0^\infty |f(t)|\,\mathrm{d}t$ *exists;*
(b) $f(t)$ *is of exponential order* $O(e^{at})$ *as* $t \to \infty$, *provided* $\mathrm{Re}(s) > a$.

Property 2.3. *Let $f(t)$ and $g(t)$ be integrable and bounded functions on $[0, \infty)$ and their Laplace transforms $F(s)$ and $G(s)$ exist in sense of absolute convergence. Then the following properties are valid for $a, b \in \mathbb{R}$:*

(a) Linearity:

$$\mathcal{L}\{af(t) + bg(t)\}(s) = a\mathcal{L}\{f(t)\}(s) + b\mathcal{L}\{g(t)\}(s).$$

(b) Time scaling:

$$\mathcal{L}\{f(at)\}(s) = \frac{1}{a}\mathcal{L}\{f(t)\}(\frac{s}{a}).$$

(c) Complex (image) shifting:

$$\mathcal{L}\{f(t)e^{-at}\}(s) = \mathcal{L}\{f(t)\}(s + a).$$

(d) Laplace transform derivative:

$$\mathcal{L}\{t^n f(t)\}(s) = (-1)^n \mathcal{L}^{(n)}\{f(t)\}(s).$$

Property 2.4. *The Laplace transform of probability density function of selected distributions [72].*

(a) Let $p(x)$ be a probability density function (PDF) of an exponential distribution with mean β:

$$p(x) = \frac{1}{\beta}e^{-x/\beta},$$

then

$$P(s) = \mathcal{L}[p](s) = \frac{1}{s\beta + 1}.$$

(b) Let $p(x)$ be a PDF of a gamma distribution with shape a and scale b:

$$p(x) = \frac{1}{\Gamma(a)b^a}e^{-x/b}x^{a-1},$$

then

$$P(s) = \mathcal{L}\{p(x)\} = \frac{1}{(sb + 1)^a}.$$

Definition 2.12. *A function $f(t)$ is called the inverse Laplace transform if it is continuous on $[0, \infty)$ and for a given $F(s)$ it satisfies:*

$$F(s) = \mathcal{L}[f(t)](s).$$

It is then denoted as $f(t) = \mathcal{L}^{-1}\{F(s)\}(t)$.

Definition 2.13. *The convolution of two functions* $f(t)$ *and* $g(t)$ *with non-negative support is denoted as* $(f * g)(t)$ *and defined as a proper integral:*

$$(f * g)(t) = \int_0^t f(t-\tau)g(\tau)\,\mathrm{d}\tau\,, \quad f, g : [0, \infty) \to \mathbb{R}.$$

Theorem 2.2. *Let* $f(t)$ *and* $g(t)$ *be integrable and bounded functions on* $[0, \infty)$*, whose Laplace transforms are absolutely convergent, then*

$$\mathcal{L}\{(f * g)(t)\}(s) = \mathcal{L}\{f(t)\}(s)\mathcal{L}\{g(t)\}(s).$$

2.2.2 Double Laplace transform

Since there are two state variables in our model (which is described in details in Section 4), it is reasonable to introduce the double Laplace transform, which is a natural extension of the single-variable Laplace transform.

Definition 2.14. *Let* $f(x, y)$ *be a function of two variables* x *and* y *defined in the first quadrant of* $\mathbb{R}^2$*. The double Laplace transform of* $f(x, y)$ *is defined by the double integral in the form:*

$$F(\phi, \psi) = \mathcal{L}_2\{f(x, y)\}(\phi, \psi) = \int_0^\infty \int_0^\infty f(x, y)e^{-x\phi - y\psi}\,\mathrm{d}x\,\mathrm{d}y\,.$$

A sufficient condition for existence of the integral in Definition 2.14 is following: if $f(x, y)$ is a continuous function in every finite intervals $(0, X)$ and $(0, Y)$ and of exponential order $O(e^{ax+by})$, then $F(\phi, \psi)$ exists for all ϕ and ψ provided $\mathrm{Re}(\phi) > a$ and $\mathrm{Re}(\psi) > b$ [73]. If this condition is met, by analogy with Theorem 2.1, it also implies absolute convergence of the integral.

As follows from the definition, $F(\phi, \psi)$ could be equivalently obtained through consistently applying the single Laplace transform twice; that is why it inherits Property 2.3. The remaining part of the section is devoted to lesser known properties, which will be needed.

Property 2.5. *Let* $f(x, y)$ *be a function of two non-negative variables, such that* $\mathcal{L}_2 f(x, y)(\phi, \psi)$ *exists, then the following properties are valid [74]:*

(a) Double Laplace transform of an integral:

$$\mathcal{L}_2\left\{\int_0^x \int_0^y f(u, v)\,\mathrm{d}u\,\mathrm{d}v\right\}(\phi, \psi) = \frac{1}{\phi\psi}F(\phi, \psi),\ \phi > 0,\ \psi > 0.$$

(b) Partial derivative of double Laplace transform:

$$\mathcal{L}_2\{x^n y^m f(x,y)\}(\phi,\psi) = (-1)^{n+m}\frac{\partial^{n+m}}{\partial\phi^n\psi^m}F(\phi,\psi),\ \ n,m\in\mathbb{N}.$$

The convolution of $f(x,y)$ and $g(x,y)$ is denoted as $(f**g)(x,y)$ is and defined as a double integral:

$$(f**g)(x,y) = \int_0^x\int_0^y f(x-\xi,y-\nu)g(\xi,\nu)\,\mathrm{d}\nu\,\mathrm{d}\mu\,.$$

Theorem 2.3. *Let $f(x,y)$ and $g(y)$ be integrable in the positive quadrant functions, such that $\mathcal{L}_2\{f(x,y)\}$ and $\mathcal{L}_2\{g(x,y)\}$ are absolutely converging, then*

$$\mathcal{L}_2\left\{(f**g)(x,y)\right\}(\phi,\psi) = \mathcal{L}\{f\}(\phi,\psi)\mathcal{L}\{g\}(\phi,\psi).$$

We define a convolution about axis as the following proper integral [75]:

$$f\underset{a\ b}{*}g = \int_0^M f(x-a\nu,y-b\nu)g(\nu)\,\mathrm{d}\nu\,,\ \text{where}\ M=\min\{\frac{x}{a},\frac{y}{b}\},$$

where parameters a and b are positive constants.

Theorem 2.4. *Let $f(x,y)$ and $g(y)$ be integrable in the positive quadrant functions, such that $\mathcal{L}_2\{f(x,y)\}$ converges boundedly and $\mathcal{L}\{g(x)\}$ converges absolutely, then*

$$\mathcal{L}_2\left\{f\underset{a\ b}{*}g\right\}(\phi,\psi) = \mathcal{L}[f](\phi,\psi)\mathcal{L}[g](a\phi+b\psi).$$

2.3 Liouville equation

Piecewise deterministic processes will be described, in the section to follow, in term of integro-differential equations for the underlying probability distributions. As the first step toward formulating such equations, we investigate in this section the density evolution of purely deterministic processes.

Consider an initial value problem modeling deterministic decay

$$\dot{x} = -x,\ \ x(0)=x_0,$$

with a solution

$$x(t) = x_0e^{-t}.$$

In the present context, the variable x can represent concentration of a protein that is continually depleted by the action of enzymes, which break down proteins. If x_0 is drawn randomly from a continuous distribution with a probability density

function $f_0(x)$, then $x(t)$ is random too, and its probability density function, by the transformation rule, is given by

$$f(x,t) = e^t f_0(e^t x).$$

Direct calculations confirm that $f(x,t)$ satisfies:

$$\frac{\partial}{\partial t} f(x,t) - \frac{\partial}{\partial x}(x f(x,t)) = 0,\ \ f(x,0) = f_0(x).$$

More generally, the density evolution for

$$\dot{x} = a(x,t),\ \ x(0) = x_0$$

is given by the Liouville equation [76]:

$$\frac{\partial}{\partial t} f(x,t) + \frac{\partial}{\partial x}\left(a(x,t) f(x,t)\right) = 0,\ \ f(x,0) = f_0(x). \tag{2.20}$$

We proceed with generalization of the equation (2.20) to an $n-$dimensional space. If $\mathbf{x} = (x_1, \ldots, x_n) \in \mathbb{R}^n$ and $\mathbf{A} : \mathbb{R}^n \times \mathbb{R} \to \mathbb{R}^n$, then the densities in a dynamical system:

$$\dot{\mathbf{x}} = \mathbf{A}(\mathbf{x}, t),\ \ \mathbf{x}(0) = \mathbf{x_0}$$

are evolved as per

$$\frac{\partial f}{\partial t} + \nabla \cdot \mathbf{A}(\mathbf{x},t) f = 0,\ \ f(\mathbf{x},0) = f_0(\mathbf{x}).$$

2.4 Drift jump framework

We consider a Markov process with drift and jumps to be the most fitting framework for modeling gene expression, it captures stochastic and deterministic mechanics of the process and suitable for implementing various regulatory circuits. First, we assume that a deterministic motion $\mathbf{x} = \mathbf{x}(t)$ in the process is driven by a dynamical system:

$$\dot{\mathbf{x}}(t) = \mathbf{A}(\mathbf{x}(t), t),\ \ \mathbf{A} : \mathbb{R}^n \times [0, +\infty) \to \mathbb{R}^n. \tag{2.21}$$

On the other hand, random jumps are governed by a random process $\mathbf{B}(t)$, such that a conditional probability to jump from $\mathbf{x}'$ to $\mathbf{x} = \mathbf{x}' + \mathbf{b}$ in a infinitesimal time interval of length Δt is

$$Pr\{\mathbf{B}(t) \in [\mathbf{b}; \mathbf{b} + d\mathbf{b}] \,|\, \mathbf{x}(t) = \mathbf{x}'\} = \omega(\mathbf{b}|\mathbf{x}', t)\, \mathrm{d}\mathbf{b}\,, \tag{2.22}$$

which we will refer to as a jump kernel. Then the trajectory of the described process is [76]:

$$\mathbf{x}(t+\Delta t) = \begin{cases} \mathbf{x}(t) + \mathbf{A}(\mathbf{x},t)\Delta t, \text{ w.p. } 1-\lambda(\mathbf{x},t)\Delta t \\ \mathbf{x}(t) + \mathbf{A}(\mathbf{x},t)\Delta t + \mathbf{B}(t), \text{ w.p. } \lambda(\mathbf{x},t)\Delta t, \end{cases} \tag{2.23}$$

where $\lambda(\mathbf{x},t)$ is a function of a jump rate. Thus, particles can either proceed drifting along solutions of $\mathbf{A}$ or perform the random jump. Then a joint probability density function of the vector $\mathbf{x}(t)$ is governed by the Chapman-Kolmogorov differential equation (CKDE) [77]:

$$\frac{\partial}{\partial t}p(\mathbf{x},t) = -\sum_{i=1}^{n}\frac{\partial}{\partial x_i}\left(A_i(\mathbf{x},t)p(\mathbf{x},t)\right) - \lambda(\mathbf{x},t)p(\mathbf{x},t) + J_{\text{in}}(\mathbf{x},t), \tag{2.24}$$

where $J_{\text{in}}(\mathbf{x},t)$ is given by

$$J_{\text{in}}(\mathbf{x},t) = \int_{\mathbb{R}^n} \lambda(\mathbf{x}',t)\omega(\mathbf{x}-\mathbf{x}'|\mathbf{x}',t)p(\mathbf{x}',t)\ \mathrm{d}\mathbf{x}'\,. \tag{2.25}$$

The integral term gives the probability that the process jumps from state a $\mathbf{x}'$ to the state $\mathbf{x}$ in an infinitesimal interval of time; in the notation $J_{\text{in}}(\mathbf{x},t)$, the subscript indicates that the term relates to the influx of probability.

It is usual that the coefficients of (2.24) do not explicitly depend on time, i.e. we have $A_i(\mathbf{x},t) = A_i(\mathbf{x})$ for the drift, $\omega(\mathbf{x}-\mathbf{x}'|\mathbf{x}',t) = \omega(\mathbf{x}-\mathbf{x}'|\mathbf{x}')$ for the burst kernel, and $\lambda(\mathbf{x},t) = \lambda(\mathbf{x})$ for the burst frequency. Moreover, in the next section we consider a possibility that the jump kernel also does not explicitly depend on outgoing state $\mathbf{x}'$, in which case we have $\omega(\mathbf{x}-\mathbf{x}'|\mathbf{x}',t) = \omega(\mathbf{x}-\mathbf{x}')$.

In the next chapter, the formulated equation will be used in an applied problem, in which restriction is imposed on permissible values of the vector $\mathbf{x}(t)$. Therefore, let us further assume that the dynamical system (2.21) maintains non-negativity of trajectories $\mathbf{x}(t)$:

$$\forall t \geq 0, \forall \mathbf{x} = (x_1,\ldots,x_n) \in \mathbb{R}^n_+, \forall i \in 1,\ldots n : x_i = 0 \Rightarrow A_i(\mathbf{x},t) \geq 0, \tag{2.26}$$

where $\mathbb{R}^n_+$ is the non-negative orthant of $\mathbb{R}^n$. We also assume that jumps are non-negative, i.e. $\omega(\mathbf{b}|\mathbf{x}',t) = 0,\ \forall x \notin \mathbb{R}^n_+$; then the influx in (2.24) becomes a definite integral:

$$J_{\text{in}}(\mathbf{x},t) = \int_{R(\mathbf{x})} \lambda(\mathbf{x}',t)\omega(\mathbf{x}-\mathbf{x}'|\mathbf{x}',t)p(\mathbf{x}',t)\ \mathrm{d}\mathbf{x}'\,, \tag{2.27}$$

where the integration domain $R(\mathbf{x})$, where $\mathbf{x} = (x_1,\ldots,x_n)$ is n-dimensional rect-

angle given as Carthesian product:

$$R(\mathbf{x}) = \prod_{i=1}^{n} [0, x_i] \subset \mathbb{R}^n.$$

In this situation, equation (2.24) holds for $x \in \mathbb{R}_n^+$, whereas $p(\mathbf{x}, t) = 0$ for $x \notin \mathbb{R}_n^+$.

It is sometimes useful to use the jump size $\mathbf{b}$ instead of outgoing state $\mathbf{x}' = \mathbf{x} - \mathbf{b}$ as the integration variable; performing this substitution in (2.27) gives an alternative expression of the influx term

$$J_{\text{in}}(\mathbf{x}, t) = \int_{R(\mathbf{x})} \lambda(\mathbf{x} - \mathbf{b}, t)\omega(\mathbf{b}|\mathbf{x} - \mathbf{b}, t)p(\mathbf{x} - \mathbf{b}, t)\ \mathrm{d}\mathbf{b}\,, \tag{2.28}$$

which will be made use of below.

Finally, the time dependent distribution $p(\mathbf{x}, t)$ converges to the stationary distribution:

$$p(\mathbf{x}, t) \sim p(\mathbf{x}), \quad t \to \infty. \tag{2.29}$$

This convergence holds in general for Markov processes which have a stationary distribution and which are irreducible (can travel between any two states) and aperiodic (do not exhibit deterministic oscillations); clearly, processes studied in this thesis satisfy these conditions.

2.5 Population framework

2.5.1 Measure-valued Markov Process

In order to extend the model to the population level, we need to describe the cell cycle mechanism of a single cell. Let t_b and t_e denote the beginning and the end of the cell cycle. For the original cell $t_b = 0$; for the cells derived from the original cell, t_b is equal to the end of the cell cycle of its mother. The cell cycle's end t_e is sampled as follows: we assume that the cell has a protein-dependent propensity to end its cell cycle, and that this propensity is equal to the dilution rate:

$$Prob[t_e \in (t, t + dt)|t_e > t] = \gamma(x(t))dt + o(dt), \quad t > t_b. \tag{2.30}$$

The end of the cell cycle triggers a branching (division) event: the current process (the mother cell) is terminated and replaced with two new processes (daughter cells), which inherit the mother's protein concentration, and half of the mother's cell volume. The daughter processes evolve henceforth independently of each other.

The above construction leads to a sequence of Markov processes

$$x_i(t), \quad t_b^i \leq t < t_e^i, \quad i = 1, 2, \ldots,$$

where t_b^i and t_e^i denote the beginning and the end of the cell cycle of the ith cell, and $x_i(t)$ gives the protein concentration at time t of the ith cell. The mother cell is indexed by $i = 1$. The ordering of the rest of the sequence depends on the algorithmic implementation of the branching model (see Appendix A) and is immaterial for the purposes of this section.

The composition of the population at time t can be represented by the empirical population density (or empirical measure)

$$m(x,t) = \sum_{i:t_b^i<t<t_e^i} \delta(x - x_i(t)). \tag{2.31}$$

This construction places a unit mass on top of each existing cell's concentration. The sum is finite because the number of cells that exist at any given time is finite. The measure $m(x,t)$ is a random empirical measure: it is empirical because it is based on a finite number of observations $x_i(t)$; it is random because these observations are random. As a function of time t, $m(x,t)$ is a measure-valued Markov process: it is Markovian because the future dynamics of the population can be sampled given the present state of the measure irrespective of the past. Measure-valued Markov processes have been widely used to model the stochastic dynamics of multi-particle systems [78], [79].

The empirical measure is not normalised. In fact,

$$N(t) = \int_0^\infty m(x,t)dx = \#\{i : t_b^i < t < t_e^i\}$$

gives the number of cells that exist at time t.

Let us consider the expected value of the empirical population density

$$h(x,t) = \mathbb{E}\, m(x,t), \tag{2.32}$$

which we will refer to as the (average) population density.

Aforementioned can be generalised to the cell population, where each ith cell has n independent properties, i.e., the cell is a multivariate Markov process $\mathbf{x}_i(t)$.

2.5.2 Population balance equation

In general, population balance equation (PBE) is an equation that describes the mean-field behaviour of a population of particles with various possible internal fea-

tures, spatial movements, and interaction mechanics.

Each particle is defined by external coordinates of its centroid $\mathbf{r} = (r_1, r_2, r_3)^T$ with domain $\Omega_{\mathbf{r}}$ and internal coordinates $\mathbf{x} = (x_1, \ldots, x_d)^T$ with domain $\Omega_{\mathbf{x}}$, which determines possible values of the internal characteristics of the particle.

Let $m_1(\mathbf{x}, \mathbf{r}, t)$ be an empirical population density, then we define the (average) population density function:

$$h_1(\mathbf{x}, \mathbf{r}, t) = \mathbb{E}\, m_1(\mathbf{x}, \mathbf{r}, t), \tag{2.33}$$

which is a sufficiently smooth function to allow differentiation with respect of any argument required number of times. Since in this work we use the population balance equation (PBE) exclusively to model protein concentration within a cell population, we ensure that it meets the following conditions:

- the particle states are continuous non-negative variables as in (2.26), then the domain $\Omega_{\mathbf{x}} = \mathbb{R}^+_d \cup \{0\}$;
- the environment does not affect the particle inner state vector $\mathbf{x}$;
- particles are distributed in space uniformly and thus the (average) population density $h_1(\mathbf{x}, \mathbf{r}, t)$ does not depend on external coordinates.

These requirements on a single particle (cell) behaviour are identical to ones discussed in Section 2.4. Then each cell is described by a piecewise deterministic Markov process $\mathbf{x}(\mathbf{t})$, in which the deterministic motion is given by (2.21), random jumps are governed by random process $\mathbf{B}(t)$ (2.22), and the time trajectory is (2.23). The cell population then becomes a measure-valued Markov process (Section 2.5.1), implying $m_1(\mathbf{x}, \mathbf{r}, t) \equiv m(\mathbf{x}, t)$ and the average density becomes $h_1(\mathbf{x}, \mathbf{r}, t) \equiv h(\mathbf{x}, t)$ as per (2.31) and (2.32), respectively.

At this point, without birth and death mechanism, the time evolution of the initial population of the size h_0 is governed by population balance equation (PBE):

$$\frac{\partial}{\partial t} h(\mathbf{x}, t) + \boldsymbol{\nabla} \cdot \mathbf{A} h(\mathbf{x}, t) - (J_{\text{in}}(\mathbf{x}, t) - \lambda(\mathbf{x}, t) h(\mathbf{x}, t)) = 0,$$

which is identical to CKDE and shows that the population is conserved.

Finally, we introduce the net rate of particles generation (population growth) $G(\mathbf{x}, t)$, which is combined effect of all birth-death processes. The change in number of particles with internal coordinates in the range $[\mathbf{x}, \mathbf{x} + \mathrm{d}\mathbf{x}]$ is given by $G(\mathbf{x}, t)\, \mathrm{d}\mathbf{x}$. Here we assume that for given internal state $\mathbf{x}$ the growth net rate is proportional to number of particles in state $\mathbf{x}$, i.e., $G(\mathbf{x}, t) = g(\mathbf{x}, t) h(\mathbf{x}, t)$. Then we obtain the final form of population balance equation [80]:

$$\frac{\partial}{\partial t} h(\mathbf{x}, t) + \boldsymbol{\nabla} \cdot \mathbf{A} h(\mathbf{x}, t) - (J_{\text{in}}(\mathbf{x}, t) - \lambda(\mathbf{x}, t) h(\mathbf{x}, t)) = g(\mathbf{x}, t) h(\mathbf{x}, t). \tag{2.34}$$

In this work, the net rate function $g(\mathbf{x}, t)$ is a rate of a cell population proliferation, and we consider two types of them: constant ($g(\mathbf{x}, t) \equiv \text{const}$) and protein-dependent ($g(\mathbf{x}, t) = \gamma(\mathbf{x}(t))$ as per (2.30)).

At the initial time $t = 0$, we have a single cell with non-random initial values of internal coordinates $\mathbf{x_0} = (x_{0,1}...x_{0,d})^T$, the initial condition for (2.34) is

$$h(\mathbf{x}, 0) = m(\mathbf{x}, 0) = \prod_{i=1}^{d} \delta(x_i - x_{0,i}).$$

Depending on the birth-death mechanism, there may arise boundary conditions (as in Section 5.3) and regulatory conditions (see [81], Chapter 2.11 in [80]).

2.5.3 Large-time solution of PBE

The population growth term has important consequences. The large-time behaviour of both equations (2.24) and (2.34) is characterised by their principal eigenvalue λ and the associated eigenfunction $p(x)$ and adjoint eigenfunction $w(x)$. For the Chapman–Kolmogorov equation, the principal eigenfunction $p(x)$ is the stationary distribution, the principal eigenvalue is $\lambda = 0$, and the adjoint eigenfunction is trivially $w(x) \equiv 1$. For the population balance equation, these characteristics can be found using he application of spectral decomposition:

$$h(x, t) \sim w(x_0)e^{\lambda t}p(x), \quad t \to \infty. \tag{2.35}$$

The adjoint eigenfunction is thereby chosen so as to satisfy the biorthogonality condition $\int_0^\infty w(x)p(x)dx = 1$. The adjoint eigenfunction determines the dependence of the large time behaviour on the initial condition. This is not too important for us and we will not calculate the adjoint eigenfunction (or at least not yet).

We are interested in the actual size and composition of the population rather than its expectation. A classical result for supercritical (i.e. with $\lambda > 0$) branching processes comes to the rescue, which says that the expectation can be removed from the left hand side of (2.35), with the caveat that $w(x_0)$ be replaced by a random variable [82], [83]. The empirical population density thus satisfies

$$m(x, t) \sim W(x_0)e^{\lambda t}p(x), \quad t \to \infty, \tag{2.36}$$

where $W(x_0) > 0$ is a random variable such that $\mathbb{E}\, W(x_0) = w(x_0)$. Equation (2.36) has been rigorously proven for multitype branching processes with finite number of types [82], [83]. Eigenfunctions are then simple eigenvectors. Here, instead of finite number of types, we have a continuum of possible concentration levels, hence eigenfunctions.

Equation (2.36) is equivalent to

$$n(t) \sim W(x_0)e^{\lambda t}, \quad \frac{m(x,t)}{N(t)} \sim p(x), \quad t \to \infty. \tag{2.37}$$

The total population increases exponentially. The influence of the initial condition and the initial low population noise is encompassed in the random pre-exponential factor $W(x_0)$. The normalised empirical population density converges to the principal eigenfunction of the population balance equation.

2.6 Basic model of unregulated gene expression

In the absence of feedback, production events (or *bursts*) occur randomly in time with constant stochastic rate (or *frequency*) α per unit time. This means that inter-arrival times of synthesis events are i.i.d. exponentially distributed random variables with mean $1/\alpha$. Whenever a burst occurs, the protein concentration is discontinuously increased by the burst size, which is randomly drawn from the exponential distribution with mean β. Between bursts protein degrades deterministically, i.e, in this one dimensional model the dynamical system (2.21) is a simple ordinary differential equation (ODE):

$$\dot{x} = -x.$$

Note that the degradation rate is equal to one, implying that time is measured in units of protein lifetime. Thus the parameter α assumes the meaning of normalized burst frequency, and gives the expected number of bursts occurring per typical protein lifetime.

Finally, for the unregulated gene expression with an arbitrary burst kernel w the Chapman-Kolmogorov equation (2.24) becomes:

$$\frac{\partial p(x,t)}{\partial t} = \frac{\partial}{\partial x}\left(xp(x,t)\right) - \alpha\left(p(x,t) - \int_0^x b(x-y)p(y,t)\,\mathrm{d}y\right),$$

where last two terms can be represented as following derivative using Leibnitz integral rule:

$$-\int_0^x b(x-y)p(y,t)\,\mathrm{d}y + p(x,t) = \frac{\mathrm{d}}{\mathrm{d}x}\int_0^x \bar{\mathrm{B}}(x-y)p(y,t)\,\mathrm{d}y\,. \tag{2.38}$$

Typically, it is assumed that the burst size has an exponential distribution with mean β, then the burst kernel is given by:

$$\omega(x-y|y) = \frac{1}{\beta}e^{-(x-y)/\beta}.$$

Afterwards, we substitute $\omega(\cdot)$, integrate the equation with respect to x, and set $\frac{\partial p}{\partial t} = 0$. Since the process is aperiodic and irreducible, we use (2.29) and obtain the equation of stationary probability density function:

$$xp(x) = \alpha \int_0^x e^{-\frac{x-y}{\beta}} p(y)\,\mathrm{d}y\,, \tag{2.39}$$

which is the Volterra integral equation with a difference kernel, for which one can obtain $p(x)$ using Laplace transform approach [84]. First, we define the Laplace image $P(s)$ of pdf $p(x)$, i.e., $P(s) = \mathcal{L}\{p(x)\}(s)$. Next, we transform the integral equation into a differential equation:

$$-\frac{\mathrm{d}P(s)}{\mathrm{d}s} = \alpha P(s)\frac{1}{s+1/\beta},$$

solution of which is $P(s) = (bs+1)^{-a}$. The original function of this $P(s)$ is the PDF of gamma distribution (Property 2.4):

$$p(x) = \frac{1}{\beta^\alpha \Gamma(\alpha)} x^{\alpha-1} e^{-x/\beta}. \tag{2.40}$$

It means that in the unregulated gene expression the protein concentration $x \sim \Gamma(\alpha, \beta)$, i.e. it is gamma-distributed with shape α and scale β [18]. From this follow, in particular, steady state mean and variance of the protein concentration:

$$\mathbb{E}(x) = \alpha\beta, \quad \mathrm{Var}(x) = \alpha\beta^2. \tag{2.41}$$

Chapter 3

Feedback on transcriptional frequency

In this chapter, we explore gene expression within the context of a feedback on a transcriptional rate. We consider a Hill-type negative autoregulation, implying that the rate of transcription for the protein of interest is inversely dependent on its current concentration. This concept is implemented into the basic model of unregulated gene expression in a single cell (see Section 2.6).

Next, we consider an autoregulatory function and extended study to a model of a cell population (see Section 2.5). Although the autoregulatory loop modifies the protein distribution relative to scenarios of unregulated expression, we demonstrate that the protein distributions within a single cell and across a population are, in fact, identical.

This chapter is partially based on the article previously published in European Control Conference [85], the content of which was expanded by Section 3.2.

3.1 Single cell model

In this section we consider a model with negative autoregulation, which includes only the protein X itself. This nodel is an extension of the fundamental model, given in Section 2.6. The gene expression is studied in terms of a piecewise-deterministic Markov process [18] in a continuous time and state space, in which the protein X decays deterministically but is produced stochastically in bursts. Continuous space state is achieved by studying protein concentration – number of protein molecules per unit of volume – as opposed to the number of protein molecules [86]. The process

can be expressed as a sequence of chemical reactions:

$$R_1: \quad \emptyset \xrightarrow{\Lambda(x)} \mathrm{B} \times \mathrm{X}$$
$$R_2: \quad \mathrm{X} \xrightarrow{1} \emptyset$$

where X is produced in random bursts with frequency governed by response function $\Lambda(x)$ in the reaction R_1 and naturally degrades in R_2. We suppose that the burst size is drawn from exponential distribution with mean β. The value one of the degradation rate constant corresponds to measuring time in units of the protein lifetime, which can be done without any loss of generality. The empty set symbol on the left-hand side of R_1 indicates that X is created from an inexhaustible source of molecules; the stoichiometric coefficient B on the right-hand side gives the burst size. The empty set on the right-hand side represents a sink of degraded molecules, so that a product of decomposition is out of our interest.

According to the reaction channel R_2, the dynamical system (2.21) is given by

$$\dot{x} = A(x) = -x,$$

i.e. the protein concentration trajectory $x(t)$ is proportional to e^{-t} inside any time interval which does not contain a burst event. The value one of the degradation rate constant corresponds to measuring time in units of the protein lifetime, which can be done without any loss of generality.

The (stationary) distribution of protein concentration is obtained as a steady-state solution of the Chapman-Kolmogorov equation (2.24) associated to the process described above. The master equation here

$$\frac{\partial p(x,t)}{\partial t} = \frac{\partial}{\partial x}\left(x p(x,t)\right) - \Lambda(x) p(x,t) + J_{\mathrm{in}}(x,t), \tag{3.1}$$

where as J_{in} we denoted the probability influx according to (2.27), which is here given by

$$J_{\mathrm{in}}(x,t) = \int_0^x \omega(x-y|y)\,\Lambda(y) p(y,t)\,\mathrm{d}y\,.$$

In the probability flux $J_{\mathrm{in}}(x,t)$, describing an instant growth of the protein level in bursts, $\omega(x-y|y)$ is a jump kernel (2.22), which defines the conditional probability that concentration will jump from a given protein level y to x in infinitesimal period of time, and $\Lambda(x)$ is the function of the jump rate. Equation (3.1) is understood to hold for non-negative concentrations of $(x \geq 0)$, whereas $p(x,t) = 0$ for any $x < 0$.

Let us introduce an equivalent kernel, which we will refer to as a *jump kernel* given by the product of the burst rate and a survival function correspond-

ing to $\omega(x-y|y)$

$$B(x|y) = \Lambda(y) \int_x^\infty \omega(z-y|y)\,\mathrm{d}z\,, \tag{3.2}$$

which expresses the probability of burst size is larger than the difference $x-y$ at frequency $\Lambda(x)$. For the jump kernel the following equalities are true:

$$B(x|x) = \Lambda(x) Pr\{B(t) > 0\} = \Lambda(x), \tag{3.3}$$

$$\frac{\partial}{\partial x} B(x|y) = -\Lambda(y)\omega(x-y|y), \tag{3.4}$$

which allow us to use the Leibniz integral rule and collapse the last two terms in (3.1) into a derivative of an integral with a variable bound, i.e

$$\frac{\partial}{\partial x}\left(\int_0^x B(x|y)p(y,t)\,\mathrm{d}y\right) = \Lambda(x)p(x,t) - \int_0^x \Lambda(y)\omega(x-y|y)p(y,t)\,\mathrm{d}y\,. \tag{3.5}$$

Equation (3.1) takes the form of a probability conservation law [87]

$$\frac{\partial p(x,t)}{\partial t} + \frac{\partial J(x,t)}{\partial x} = 0,$$

in which the probability flux is now given by [88]

$$J(x,t) = -xp(x,t) + \int_0^x B(x|y)p(y,t)\,\mathrm{d}y\,.$$

Like previous studies (e.g. [18], [88]), we focus on the steady-state behavior of the model: we set $\partial p(x,t)/\partial t = 0$, from which it follows that the probability flux $J(x,t)$ must be constant with respect to x; and since we do not admit a nonzero flux of probability mass from infinity, this constant has to be equal to zero. Thus, we obtain a Volterra integral equation of the second kind [84] with respect to $p(x,t) = p(x)$:

$$xp(x) = \int_0^x B(x|y)p(y)\,\mathrm{d}y\,. \tag{3.6}$$

Next, we implementing negative feedback, meaning that we consider a non-increasing family of functions $\Lambda(x)$. Specifically, we declare $\Lambda(x)$ as the Hill function [89]

$$\Lambda_H(x) = \frac{\alpha}{1+(x/K)^H}, \tag{3.7}$$

where α, H, and K are parameters explained below. The Hill function has been widely used in literature and justified in terms of cooperative binding of a protein molecule to its gene's promoter. The parameter α here takes the role of maximal burst frequency, which is achieved in the absence of the self-repressing protein. The parameter K gives the concentration required to reach half the maximum burst

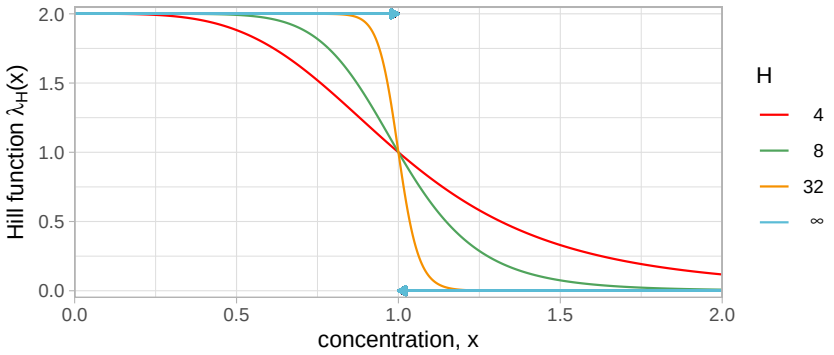


Figure 3.1: Examples of Hill functions (3.7) with different choices of Hill coefficients H. As H increases to infinity, the function develops at $x = K$ a jump discontinuity (marked with blue arrows). The other parameters of (3.7) are set to α and $K = 1$.

frequency; it corresponds to the dissociation constant of the protein–promoter binding [90]. Interestingly, by taking K to infinity one recovers the unregulated model as introduced earlier in this section. The parameter H, which is referred to as the Hill coefficient, indicates how steeply feedback reacts to changes in protein concentration (as shown in Figure 3.1), and directly corresponds to the cooperativity in the underlying promoter–protein interaction.

We focus specifically on very large binding cooperativities; taking the Hill coefficient to infinity, we find that (cf. Fig. 3.1)

$$\Lambda_\infty(x) = \lim_{H\to\infty} \Lambda_H(x) = \begin{cases} \alpha, & x < K, \\ \frac{\alpha}{2}, & x = K, \\ 0, & x > K, \end{cases} \tag{3.8}$$

which we hereafter refer to as bang–bang feedback. In case of bang–bang type feedback, the protein exponentially degrades, and bursts cannot occur, as long as the concentration level is higher than K; once x falls beneath K, bursts occur with a constant frequency α.

The negative regulation kernel is given by the $\Lambda(y)$ and the probability of a burst exceeding the size of $x - y$, i.e.

$$B(x|y) = \Lambda(y)\bar{B}(x - y), \tag{3.9}$$

where $\bar{B}(x-y)$ is a complementary cumulative distribution function (CCDF) of the exponential distribution, i.e., $\bar{B}(x - y) = e^{-\frac{x-y}{\beta}}$. Note that while this kernel is no longer a difference kernel, it is still a product kernel, and, as such, the associate

integral equation

$$xp(x) = \int_0^x \Lambda(y) e^{-\frac{x-y}{\beta}} p(y)\,\mathrm{d}y$$

admits an explicit solution [32]. In order to find it, we pull out from under the integral sign the exponential function $e^{-x/\beta}$, then differentiate the equation with respect to x and apply the Leibnitz integral rule; this yields an ODE

$$(e^{x/\beta} x p(x))' = \Lambda(x) e^{x/\beta} p(x),$$

the general solution of which is

$$p(x) = C x^{-1} e^{-x/\beta} \exp\left(\int \frac{\Lambda(x)}{x}\,\mathrm{d}x\right), \tag{3.10}$$

in which C is an integration constant. The primitive function in the argument of the exponential can easily be solved for the piecewise constant $\Lambda(x)$ in case of of the bang–bang regulation (3.8). Since the bang–bang response function features a discontinuity at $x = K$, the primitive function in the exponential, as well as the PDF $p(x)$ itself, are nonsmooth at $x = K$; to the left and to the right of the point of smoothlessness the density is given by separate expressions

$$p(x) = \begin{cases} C K^{-\alpha} e^{-x/\beta} x^{\alpha-1}, & x < K, \\ C e^{-x/\beta} x^{-1}, & x \geq K. \end{cases} \tag{3.11}$$

The condition that the distribution density must be normalised to one fixes the value of the normalisation constant to

$$C = \left(\frac{\gamma(\alpha, K/\beta)}{(K/\beta)^\alpha} + E_1(K/\beta)\right)^{-1}, \tag{3.12}$$

where $\gamma(\cdot,\cdot)$ is the lower incomplete gamma function and $E_1(\cdot)$ is the exponential integral defined on the complex plane [65]. Integrating the density multiplied by the factor x, we obtain

$$\mathbb{E}(x) = \frac{\alpha\beta\gamma(\alpha, K/\beta)}{\gamma(\alpha, K/\beta) + (K/\beta)^\alpha E_1(K/\beta)} \tag{3.13}$$

for the expected value of protein concentration at steady state. Considering a limiting case with the infinitely high frequency, the mean protein concentration is

$$\mathbb{E}(x) = \beta e^{-K/\beta} E_1^{-1}(K/\beta). \tag{3.14}$$

3.2 Population model

In this section, we explore the difference between the single cell and population model. Here the protein concentration in a single cell has similar dynamics as one in the previous section: protein produced instantly in random bursts; the burst size is drawn from the exponential distribution with mean β. Between the bursts the protein degrades with constant rate γ. Bursts occur randomly, but now we consider a general form of protein-dependent stochastic rate $\Lambda(x)$.

We extend model to describe a cell population: each cell evolves independently and follows the same dynamics as in single cell model. During cell division, a mother cell splits into two identical daughter cells and dies; each daughter cell inherits half of the mother's volume and protein content. Immediately after division, the protein levels in the daughter cells are identical to the mother's, ensuring uniformity in protein concentration in descendants. For the biologically correct model (finite non-zero mean cell volume), the division event is considered to be the Poisson process with rate γ.

Then the time-evolution of the protein concentration pdf $p(x,t)$ is governed by the Chapman-Kolmogorov equation:

$$\frac{\partial p(x,t)}{\partial t} = \frac{\partial}{\partial x}(\gamma x p(x,t)) + \int_0^x b(x-y)\Lambda(y)p(y)\,\mathrm{d}x - \Lambda(x)p(x), \tag{3.15}$$

where $b(x) = e^{-x/\beta}/\beta$ is probability density function of the exponential distribution. In order to find the stationary distribution, we write it as a probability conservation equation. It is done by gathering the last two terms of (3.15) as per Leibniz integral rule (3.5).

Finally, the steady state of the system implies that distribution does not change over time, i.e., $\partial p(x,t)/\partial t = 0$. We obtain:

$$\frac{\mathrm{d}}{\mathrm{d}x}(\gamma x p(x)) = \frac{\mathrm{d}}{\mathrm{d}x}\left(\bar{B} * \tilde{p}\right)(x), \tag{3.16}$$

where an auxiliary function $\tilde{p}(y) = \Lambda(y)p(y)$ is used for convenient representation of the convolution; $\bar{B} = e^{-x/\beta}$ is denoted CCDF of the exponential distribution.

We proceed with the model of the population, where the lifecycle of each cell is identical to one described at the beginning of the chapter. Its composition is given by the function $h(x,t)$ – the average number of cells with given concentration x at the time t (see Section 2.5). The time evolution of $h(x,t)$ is described by the population balance equation (2.34):

$$\frac{\partial h(x,t)}{\partial t} = \frac{\partial}{\partial x}(\gamma x h(x,t)) + \gamma h(x,t) - \frac{\partial}{\partial x}\left(\int_0^x \bar{\mathrm{B}}(x-y)\Lambda(y)p(y)\,\mathrm{d}y\right). \tag{3.17}$$

where the net rate function $g(x,t) \equiv \gamma$ and then the population growth term $\gamma h(x,t)$.

By substitution, we can prove that the solutions of PBE (3.17) satisfy:

$$h(x,t) = p(x)e^{\gamma t},$$

where $p(x)$ is the solution of the master equation (3.15).

From which it follows, in particular, that

$$h(x,t) \sim p(x)e^{\gamma t}, \quad t \to \infty,$$

where $p(x)$ solves the stationary problem (3.17). This leads to the conclusion that the stationary distributions of the protein concentration $p(x)$ in both frame- works are identical.

Chapter 4

Feed forward model

In this chapter we present a comprehensive model for understanding the interactions in a feed forward loop (FFL). It controls the species of interest (stable mRNA molecules) via co-produced auxilliary species (unstable miRNA molecules), which are able to deactivate or degrade mRNA molecules. We show that in a low noise regime (small and frequent production bursts) it is perfectly adaptating, i.e., it perfectly counterbalances changes in a production frequency and keeps the mean mRNA concentration at the constant level. Next we investigate how the inclusion of a moderate noise affects this perfect adaptation property. We derive the stationary moments of mRNA and show that FFL is less volatile then the negative feedback loop. Finally, we extend our research to the case of an unstable mRNA and derive the stationary moments.

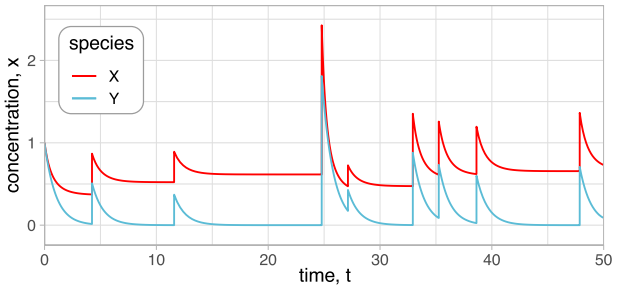
This chapter includes material based on an article "Maintenance of steady state mRNA levels by a microRNA-based feed forward loop in the presence of stochastic gene expression noise" that has been accepted for publication in European Journal of Applied Mathematics.

4.1 Model statement

We study with a chemical system with two species X and Y (mRNA and miRNA), the detailed biological explanation of which is provided in Chapter 1. We formalise it as a set of chemical reactions given in Figure 4.1a. The bursty production of both species in R_1 is modeled by stochastic jumps, and since they share the same gene, they are synthesized in equally sized bursts. Inhibition and degradation reactions R_2 and R_3 are modeled deterministically. We assume that X has a relatively long lifetime, so we neglect its spontaneous degradation. Therefore, a molecule of X can be eliminated only via the interaction R_2. While a molecule of Y survives in R_2, it degrades independently due to natural decay via the reaction channel R_3. As it follows from R_3, all reactions rates are normalised to the lifetime of Y; also the

$$
\begin{aligned}
R_1 &: \quad \emptyset \xrightarrow{\alpha} \mathrm{B} \times (\mathrm{X} + \mathrm{Y}) \\
R_2 &: \mathrm{X} + \mathrm{Y} \xrightarrow{\delta} \mathrm{Y} \\
R_3 &: \quad \mathrm{Y} \xrightarrow{1} \emptyset
\end{aligned}
$$

(a)



(b)

Figure 4.1: (a) the studied model in terms of chemical kinetics; the empty set symbol indicates that either reactants, or product of reaction are out of our interest. (b): Sample trajectories of X and Y concentrations. Between stochastic bursts, X and Y decay deterministically as per (4.1). Despite the fact that X degrades faster than Y, once level of Y is close to zero, X also stops decreasing; this indicates that the only way for X to degrade is interaction with Y. In the simulation, parameters are following: $\alpha = 0.25$, $\delta = 1$, $\beta = 1$.

hazard rate of the interaction between X and Y is constant and equal to δ. The model, in which mRNA is unstable and degrades at a constant rate γ, is further studied in the end of the section.

Since two species X and Y are involved, a state of their concentrations is given by a two-dimensional vector $\mathbf{x} = (x, y)$. The deterministic motion $\mathbf{x}(\mathbf{t})$ (2.21) is given by the mass-action kinetics of reactions R_2 and R_3, the decay of concentrations between bursts is described by the following ODEs:

$$\dot{x} = -\delta xy, \quad \dot{y} = -y. \tag{4.1}$$

The evolution of the probability density function (pdf) $p(x, y, t)$ is given by the equation:

$$\begin{aligned}
\frac{\partial p(x,y,t)}{\partial t} = \frac{\partial}{\partial x}\left(\delta xy p(x,y,t)\right) + \frac{\partial}{\partial y}\left(y p(x,y,t)\right) \\
+ J_{\mathrm{in}}(x,y,t) - \Lambda(x,y,t) p(x,y,t),
\end{aligned} \tag{4.2}$$

where

$$J_{\mathrm{in}}(x,y,t) = \Lambda(x,y,t) \int_0^x \int_0^y \omega(b_x, b_y) p(x - b_x, y - b_y, t)\, \mathrm{d}b_x\, \mathrm{d}b_y\,,$$

gives the probability influx, $\Lambda(x, y, t)$ is the burst rate and $\omega(b_x, b_y)$ is the bivariate burst jump kernel, which will be made explicit below. Note that we do not incorporate feedback in our model; the regulation is only of feed forward type. The integro-differential equation (4.2) can be treated as a special two-dimensional case of the Chapman-Kolmogorov equation, which is reviewed in Section 2.4.

According to the model requirements, bursts satisfy the following:

- the burst rate is time-homogeneous and does not depend on the either state variable: $\Lambda(x, y, t) \equiv \alpha$;
- the burst size $b_z = z - z'$ is also independent of the time; it is drawn from any distribution, the probability density function $f(b_z)$ of which has the non-negative support.
- both species are increased simultaneously and by the same amount, so that the jump kernel takes the form:

$$\omega(b_x, b_y) = f(b_x)f(b_y)\delta(F(b_x) - F(b_y)),$$

where $F(b_z)$ is the cumulative distribution function corresponding to $f(b_z)$ and $\delta(\cdot)$ is the Dirac delta function. Furthermore, a property of Dirac delta function, having an argument that is a differentiable function [91]:

$$\delta(g(x)) = \sum_i \frac{\delta(x - x_i)}{|g'(x_i)|}, \quad x_i \in \{x|g(x) = 0\}, \quad \forall x : g'(x) \neq 0,$$

allows us simplify $\delta(F(b_x) - F(b_y))$:

$$\omega(b_x, b_y) = f(b_x)\delta(b_x - b_y).$$

In our problem b_z is drawn from exponential distribution with mean β, which leads to $f(b_z) = e^{-b_z/\beta}/\beta$.

Application of the conditions above to (4.2) leads to the influx of the form:

$$J_{\text{in}}(x, y, t) = \frac{\alpha}{\beta} \int_0^x \int_0^y e^{-b_y/\beta} p(x - b_x, y - b_y, t)\delta(b_x - b_y)\, db_y\, db_x. \tag{4.3}$$

We proceed with simplification of the influx and use a property of the delta function: an integral $\int_A f(x)\delta(x - x_0)dx = f(x_0)$, only if $x_0 \in A$; in the inner integral of (4.3) with respect to b_y, we set the root of delta function $x_0 = b_x$ and obtain:

$$J_{\text{in}}(x, y, t) = \frac{\alpha}{\beta} \int_0^x \theta(y - b_x)e^{-b_x/\beta} p(x - b_x, y - b_x, t)\, db_x. \tag{4.4}$$

The appearance of the Heaviside step function $\theta(\cdot)$ can be explained as follows: if a burst size of one RNA is greater than the final value of another one, then the condition of equally sized bursts cannot be satisfied. Thereby permissible values of b_x are in the interval $(0, y]$ and, at the same time, the integration interval is $(0, x]$. After we eliminate $\theta(\cdot)$, the upper limit becomes minimum of x and y.

Combining production jumps with decay drift, we obtain the dynamics of Markovian drift-jump process (its sample paths are shown in Fig. 4.1b); an integro-

differential equation for $p(x, y, t)$ is following:

$$\begin{aligned}\frac{\partial}{\partial t}p(x,y,t) =& \frac{\partial}{\partial x}(\delta x y\ p(x,y,t)) + \frac{\partial}{\partial y}(y p(x,y,t)) - \alpha p(x,y,t) \\ &+ \frac{\alpha}{\beta}\int_0^M e^{-b_x/\beta} p(x-b_x, y-b_x, t)\, db_x.\end{aligned} \tag{4.5}$$

where $M = \min\{x, y\}$ gives the upper bound. Note that the equation will remain the same, if we change the order of integration in (4.3) and then let the root of the delta function x_0 be equal to b_x. This is due to the evenness of the Dirac delta function and the above-mentioned relation between $\theta(\cdot)$ and the integration limit M.

Proceeding to solve (4.5): we apply a double Laplace transform to $p(x, y, t)$ with respect to variables x and y, and obtain its image as a function of Laplace variables ϕ and ψ (definition and required properties of the double Laplace transform are provided in Section 2.2.2):

$$P(\phi, \psi, t) = \int_0^\infty \int_0^\infty e^{-\phi x - \psi y} p(x, y, t)\, dx dy. \tag{4.6}$$

Note that the integral term in (4.5) is a convolution about an axis by Definition 2.4:

$$f \underset{a}{*}\underset{b}{} g = \int_0^M f(x - a\nu, y - b\nu) g(\nu)\, d\nu, \text{ where } M = \min\left\{\frac{x}{a}, \frac{y}{b}\right\},$$

where parameters a and b are equal to β, $f(\cdot)$ is joint pdf $p(\cdot)$, and $g(\cdot)$ is an exponential function. Hence, its Laplace transform is:

$$\mathcal{L}\left[p \underset{\beta}{*}\underset{\beta}{} e^{-\nu/\beta}\right](\phi, \psi) = \mathcal{L}[p](\phi, \psi)\mathcal{L}[e^{-\nu/\beta}](\beta\phi + \beta\psi).$$

This allows us to apply a double Laplace transform to (4.5), which results in a PDE:

$$\frac{\partial P}{\partial t} = \delta\phi \frac{\partial^2 P}{\partial \phi \partial \psi} - \psi \frac{\partial P}{\partial \psi} - P\frac{\alpha\beta(\phi + \psi)}{\beta(\phi + \psi) + 1}, \tag{4.7}$$

which is, in comparison with (4.5), more suitable for further analysis.

4.2 Mean-field model

4.2.1 Steady-state concentrations

From definition (4.6) it follows that we can derive the mixed moments of random processes $X = X(t)$ and $Y = Y(t)$ using the image of $p(x, y, t)$. It is done by differ-

entiating (4.6) and setting $(\phi, \psi) = (0, 0)$:

$$\mathbb{E}(X^k Y^m) = (-1)^{k+m} \frac{\partial^{k+m} P}{\partial \phi^k \partial \psi^m}(0, 0). \tag{4.8}$$

In particular, applying this to (4.7), we obtain the system of moment equations:

$$\begin{aligned} \dot{\mathbb{E}}(X) &= \alpha\beta - \delta\, \mathbb{E}(XY) \\ \dot{\mathbb{E}}(Y) &= \alpha\beta - \mathbb{E}(Y). \end{aligned} \tag{4.9}$$

As is typical for kinetics systems with bi-molecular reactions, the moment equations are not closed [16], i.e. higher-order moments appear in the equations for the means. This difficulty can be eliminated provided that the noise in the concentrations of x and y is low; then we can remove the expectation operators from (4.9), obtaining the mean-field model:

$$\begin{aligned} \dot{x} &= \alpha\beta - \delta xy \\ \dot{y} &= \alpha\beta - y. \end{aligned} \tag{4.10}$$

We take a look at steady state mean concentrations of x and y; this leads us to the solutions:

$$x = \frac{1}{\delta}, \quad y = \alpha\beta. \tag{4.11}$$

Interpreting transcription as input, which is characterized by production rate $\alpha\beta$, and the concentration X as the ultimate output of the system, the first equality in (4.11) indicates that the model is perfectly adaptating: the steady state of X does not depend on the rate of transcription. Thus, the model in the small noise regime successfully balances out any disturbance of the input. This statement is reinforced by deriving time-dependent solution in the next section, which is drawn as a red curve on Figure 4.2.

4.2.2 Time-dependent solutions

Let us solve the system (4.10), provided that the initial conditions are given by:

$$x(0) = x_0, \quad y(0) = y_0,$$

where x_0 and y_0 are positive constants. The solution for $y(t)$ is derived directly from the second equation in (4.10) and the initial condition above:

$$y(t) = \alpha\beta \left(\left(\frac{y_0}{\alpha\beta} - 1 \right) e^{-t} + 1 \right).$$

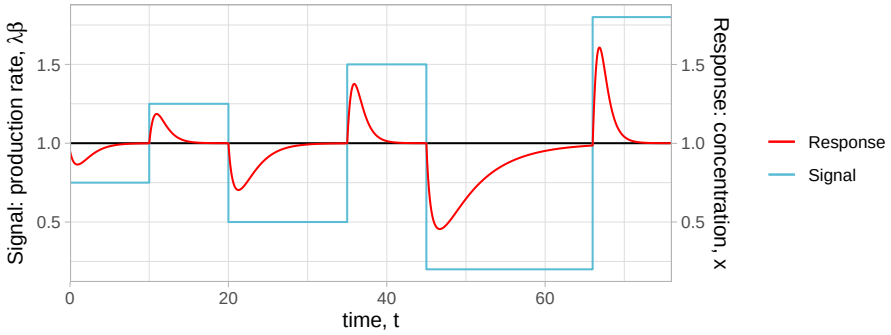


Figure 4.2: Response of FFL to changes in the input signal. A positive shift of the production rate leads to a significant short-term increase in the concentration of the species X. Otherwise, a negative shift leads to a short-term decrease in X. Note that the lower the production rate is, the longer time is needed to bring X to the steady level.

We substitute it into the first equation in (4.10) and obtain a following ODE with respect to $x(t)$:

$$\dot{x} + \rho x(Ke^{-t} + 1) = \frac{\rho}{\delta}, \tag{4.12}$$

where new constants are $K = \frac{y_0}{\alpha\beta} - 1$ and $\rho = \alpha\beta\delta$ (the effective production rate). Now we can find an integration factor $\mu(t)$; in this equation, it is given by:

$$\mu(t) = e^{\int \rho(Ke^{-t}+1)\mathrm{d}t} = e^{\rho t - \rho K e^{-t}}.$$

After multiplication of equation (4.12) by $\mu(t)$ and collapsing the left-hand side by the product rule, we find

$$(x\mu)' = \frac{\rho}{\delta}\mu,$$

i.e.

$$x(t) = \frac{\rho}{\delta}\frac{1}{\mu(t)}\left(\int_0^t \mu(\tau)\,\mathrm{d}\tau + C\right), \tag{4.13}$$

where C is an arbitrary constant. Proceeding with the integral term in (4.13), we integrate it by parts and preform a substitution $u(\tau) = \rho K e^{-\tau}$, obtaining:

$$\int_0^t \mu(\tau)\,\mathrm{d}\tau = \frac{1}{\rho}\left(e^{\rho(t-Ke^{-t})} - e^{-\rho K}\right) + \frac{(\rho K)^\rho}{\rho}\int_{u(0)}^{u(t)} u^{-\rho}e^{-u}\,\mathrm{d}u\,.$$

In case that $K > 0$ (i.e. the initial level of Y is greater then the production rate $\alpha\beta$), the concentration of X evolves in time as follows:

$$\begin{aligned} x(t) = {} & \frac{1}{\delta}\left(1 + (\delta x_0 - 1)e^{-\rho(t+K-Ke^{-t})}\right) \\ & + \frac{(\rho K)^\rho}{\delta}\frac{\gamma(1-\rho, \rho K e^{-t}) - \gamma(1-\rho, \rho K)}{e^{\rho(t-Ke^{-t})}}. \end{aligned} \tag{4.14}$$

By $\gamma(\cdot,\cdot)$ is denoted the incomplete gamma function (details on this special function

are provided in Section 2.1.1). The solution (4.14) is valid for $\rho \in (0,1)$ due to the requirement for positivity of the parameter s. However, by performing $\lceil \rho \rceil$ times integration by parts in (4.13) and subsequent following the approach above, the solution $x(t)$ can be expressed in terms of the special functions for any $\rho \in (k, k+1],\ k \in \mathbb{N}$.

In case that $K < 0$, we preform integration by a substitution $u(\tau) = -\rho K e^{-\tau}$, and the solution is

$$x(t) = \frac{1}{\delta}(1 + (\delta x_0 - 1)e^{-\rho(t+K-Ke^{-t})}) - \frac{(-\rho K)^\rho}{\delta e^{\rho(t-Ke^{-t})}} \int_{u(0)}^{u(t)} u^{-\rho} e^{-u} \,\mathrm{d}u\,, \tag{4.15}$$

where the arbitrary constant remains the same. Both of the solutions above have the same form at $K = 0$:

$$x(t) = \frac{1}{\delta}(1 - e^{-\rho t}) + x_0 e^{-\rho t},$$

which can be also seen directly from (4.12).

4.3 Stationary moments

The smallness of noise is guaranteed only in the asymptotic regime of small but frequent bursts, i.e. provided that $\beta \to 0$, $\alpha \to \infty$, while assuming that the production rate $\alpha\beta = O(1)$. However, these assumptions are restrictive and we are interested in general large-time behavior of $p(x, y, t)$. After the steady state is reached, $\frac{\partial}{\partial t} p(x, y, t) = 0$; the probability density function $p(x, y)$ as well as its image $P(\phi, \psi)$ do not depend on time:

$$\delta\phi \frac{\partial^2 P}{\partial\phi\partial\psi} - \psi \frac{\partial P}{\partial\psi} - P \frac{\alpha\beta(\phi+\psi)}{\beta(\phi+\psi)+1} = 0. \tag{4.16}$$

Notice that according to (4.6), the image $P(\phi, \psi)$ at $\phi = 0$ is equal to the single Laplace transform of the marginal distribution of the species Y, which is $p_Y(y) = \int_0^\infty p(x, y)\,\mathrm{d}x$. From this follows that we can derive $p_Y(y)$ directly by equating $\phi = 0$ in (4.16). The solution of derived equation provides an image of $p_Y(y)$:

$$P(0, \psi) = (\beta\psi + 1)^{-\alpha}. \tag{4.17}$$

The obtained function is the Laplace image of the probability density function of unregulated gene expression [85]. This is not surprising since within our model the synthesis of Y is not affected by any regulatory mechanisms.

Adhering to this logic, the marginal distribution of X can in principle be obtained

from $P(\phi, 0)$, yet it does not seem possible to obtain an analytical solution as setting $\psi = 0$ in (4.16) does not yield a closed equation. Nevertheless, it is possible to derive explicit expressions for $\mathbb{E}(X)$, cf. (4.8), as well as higher-order moments of X:

$$\mathbb{E}(X^k) = (-1)^k \frac{\partial^k P}{\partial \phi^k}(0, 0). \tag{4.18}$$

Let us start with differentiation of (4.16) with respect to ϕ, then equating ϕ to zero:

$$(\delta - \psi)\frac{\partial}{\partial \psi}\Big(\frac{\partial P}{\partial \phi}(0, \psi)\Big) - \frac{\alpha\beta\psi}{\beta\psi + 1}\frac{\partial P}{\partial \phi}(0, \psi) - \frac{\alpha\beta}{(\beta\psi + 1)^2}P(0, \psi) = 0. \tag{4.19}$$

We define the function $f(\psi) = \partial_\phi P(0, \psi)$, its expanded form is following:

$$f(\psi) = -\int_0^\infty \int_0^\infty xp(x, y)e^{-\psi y}\,\mathrm{d}y\,\mathrm{d}x\,. \tag{4.20}$$

Afterwards, using substitution (4.20) and replacement (4.17) for $P(0, \psi)$, we reduce the initial PDE (4.19) to the following ODE:

$$(\delta - \psi)f'(\psi) - f(\psi)\frac{\alpha\beta\psi}{\beta\psi + 1} - \frac{\alpha\beta}{(\beta\psi + 1)^{\alpha+2}} = 0, \tag{4.21}$$

the general solution of which is:

$$f(\psi) = C(\beta\psi + 1)^{-q}(\delta - \psi)^{-\delta\beta q} - \frac{\beta\psi + \alpha\beta\delta + 1}{\delta(\delta\beta(\alpha + 1) + 1)}\frac{1}{(\beta\psi + 1)^{\alpha+1}}, \tag{4.22}$$

where $q = \frac{\alpha}{1+\beta\delta}$.

However, we do not have initial conditions for equation (4.21) to determine the integration constant C. Instead, we can see from definition (4.20) the solution $f(\psi)$ as $\psi = 0$ is the negative value of the first moment of X, i.e. $f(0) = -\mathbb{E}(X)$; therefore, there are indirect conditions on (4.22) due to the studied model. First, concentration of the protein must be non-negative value; next, the model is placed in a cell, whose capacity is bounded, so $\mathbb{E}(X)$ must be a finite value. Using the limit comparison theorem for improper integrals [92], the formulated restrictions can also be applied to the solution (4.22):

$$-\infty < f(\psi) \leq 0. \tag{4.23}$$

Any nonzero C leads to a singularity at a point $\psi = \delta$ and a contradiction with conditions (4.23); then the only satisfying value is $C = 0$, and the solution of (4.21) is

$$f(\psi) = -\frac{\beta\psi + \alpha\beta\delta + 1}{\delta(\delta\beta(\alpha + 1) + 1)}\frac{1}{(\beta\psi + 1)^{\alpha+1}}. \tag{4.24}$$

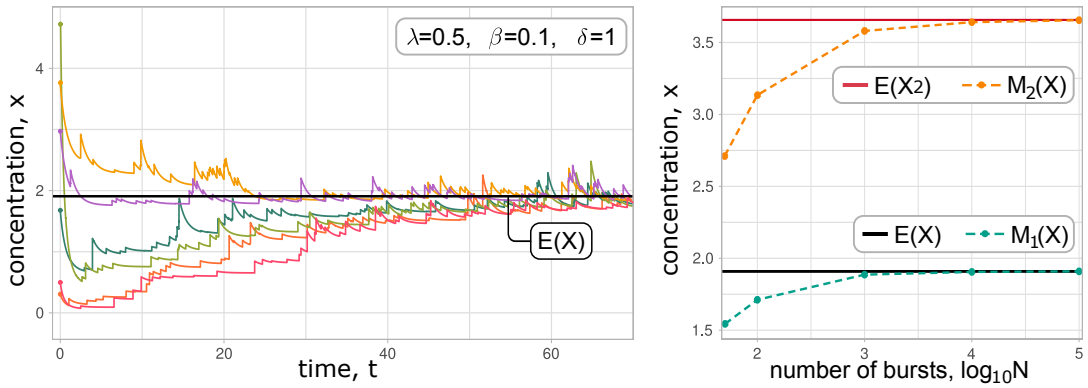


Figure 4.3: *(Left)* Sample trajectories of the mRNA concentrations given that initial conditions are drawn from the uniform distribution on the interval $[0, 2\,\mathbb{E}(X)]$. In the long-term observation, the initial conditions are not influencing the dynamics of the process. *(Right)* Convergence of the numerically computed M_1 and M_2 (dashed lines) to the corresponding analytical values of $\mathbb{E}(X)$ and $\mathbb{E}(X^2)$ (solid lines), respectively. Parameters of the simulation are following: $\delta = 1$, $\alpha = 2$, $\beta = 2$.

As was mentioned before, the definition of $f(\psi)$ makes it possible to use the Laplace image property (4.18) to obtain the steady-state mean value of X by setting $\psi = 0$ in (4.24); this yields

$$\mathbb{E}(X) = \frac{1}{\delta} - \frac{\beta}{\delta\beta(\alpha+1)+1}, \tag{4.25}$$

a graph of which is shown in Fig. 4.4 (left panel). Clearly, in a high frequency mode, the mean value in steady-state only depends on the hazard rate of interaction between the species:

$$\lim_{\alpha\to\infty} \mathbb{E}(X) = \frac{1}{\delta}, \tag{4.26}$$

which is consistent with the mean-field result (4.11).

Let us take the second derivative of (4.16) as $\phi = 0$ and repeat the same approach. The the second moment of x is

$$\begin{aligned}\mathbb{E}(X^2) =& \frac{(\xi-\delta\beta)^2}{\delta^2\xi(\xi+\delta\beta)} + \frac{2\beta(\xi-\delta\beta)}{\delta(\alpha+2)(2\xi-1)(\xi+\delta\beta)} \\ &+ \frac{\alpha\beta}{\delta(\alpha+2)}, \qquad \text{where } \xi = \beta\delta(\alpha+1)+1.\end{aligned} \tag{4.27}$$

Although we do not provide a definite proof, we expect that an explicit formula can be derived for any k-th moment.

The verification of the obtained results (4.25) and (4.27) was performed using a stochastic simulation approach. We constructed an algorithm, in which burst events are simulated using the inversion sampling method and between two consecutive bursts, the trajectories of species X and Y are determined by the solutions of (4.1). Let us denote the generated trajectory of X on a time interval $[0, T]$ as the function

$x_T(t)$. By the mean value theorem for integrals we calculate the i-th sample moment $M_i(X)$ as follows:

$$M_i(X) = \frac{1}{T}\int_0^T x_T^i(t)\,\mathrm{d}t\,. \tag{4.28}$$

The simulations show that the first moment $M_1(X)$ (sample mean) and the second moment $M_2(X)$ consistently converge to the obtained values of $\mathbb{E}(X)$ and $\mathbb{E}(X^2)$ (Fig. 4.3, right panel). Deviation of the computational results are due to the influence of the initial conditions, thus it is crucial that the simulation duration is sufficiently long and the process is stabilized (Fig. 4.3, left panel). Further details concerning simulations of a piecewise continuous trajectory $x(t)$ and the computations of $M_i(X)$ are discussed in Appendix A1.

4.4 Imperfect adaptation

In this chapter sections, we constructed a stochastic model of gene expression that implements the feedforward loop motif. This allowed us to obtain analytical values of mean concentration in steady-state. According to the obtained results, the IFFL exhibits a nearly perfect adaptation in the low noise regime, i.e if the mean burst size β is relatively small. This version of feedforward loop is well suited to support homeostasis in cell processes. Nevertheless, it is not obvious how robust it is in presence of high noise.

One way to evaluate is by comparison with another control mechanism. In Section 3, we have studied gene expression, where the production of gene product X is regulated by applying NFB loop to its transcriptional frequency. Derived steady-state distribution also allowed us to obtain an explicit formula for the steady-state mean concentration of X (3.11).

Despite the different mechanics of regulation and parameter sets, comparing equations (3.14) and (4.26) allows us to match both models in a high-frequency mode, then we set

$$\delta = \beta^{-1}e^{K/\beta}E_1(-K/\beta)$$

to ensure that both IFFL and NFB approach the same value (Fig 4.4, right panel).

As also shown in the right panel of Fig. 4.4, in the presence of high noise the feedforward loop becomes insensitive to changes in production rate much earlier, than the negative feedback. Although the system's parameters have greater impact as the mean burst size grows, the FFL still provides a non-zero concentration even for vanishingly small burst frequencies, unlike the NFB. This can be explained by noting that the main source of decreasing mRNA concentration X in FFL is miRNA Y (which is unstable), but not natural degradation.

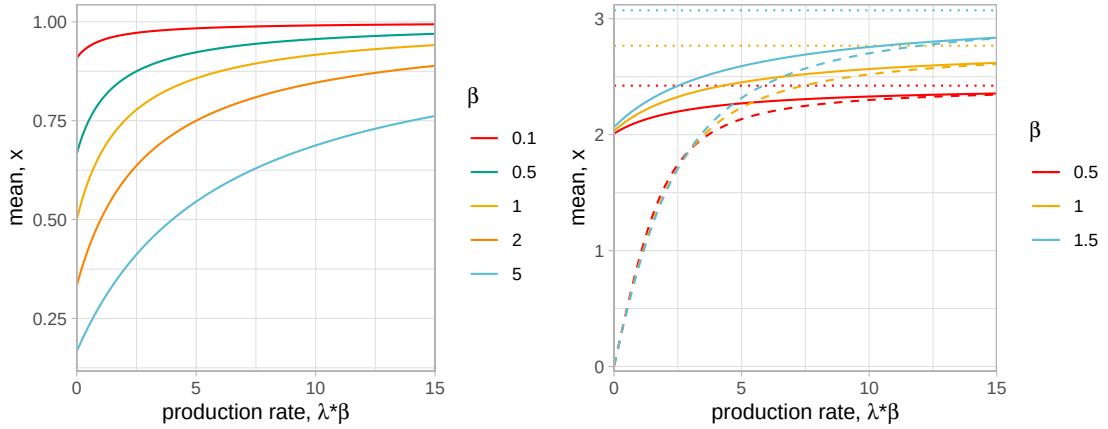


Figure 4.4: *(Left)* influence of production rate on the steady-state mean concentration of X as $\delta = 1$; *(Right)* influence of IFFL (bold lines) and negative feedback (dashed lines) on $\mathbb{E}(x)$; both models were tuned so that in the high frequency mode $\mathbb{E}(x)$ approaches to the same limit (dotted lines). We let dissociation constant $K = 2$.

Based on these observations, we can conclude that if fine tuning of factors is possible in a system, FFL provides a much narrower interval of possible steady-state concentrations; it can be an efficient way of maintaining homeostasic expression of a gene.

4.5 Model expanded to the natural degradation of mRNA

Our main goal was to study the minimal model with isolated incoherent feed forward loop (IFFL), which allows us to obtain results in the interpretable parameter space. Thus, we assumed that mRNA is stable and thus does not degrade. In this section, we study a more general case, where mRNA is unstable and its degradation cannot be neglected.

The reaction set corresponding to the new model (compared to the one in Fig. 4.1a):

$$R_1 : \quad \emptyset \xrightarrow{\alpha} \mathrm{B} \times (\mathrm{X} + \mathrm{Y}), \qquad R_3 : \mathrm{Y} \xrightarrow{1} \emptyset,$$
$$R_2 : \mathrm{X} + \mathrm{Y} \xrightarrow{\delta} \mathrm{Y}, \qquad R_4 : \mathrm{X} \xrightarrow{\gamma} \emptyset.$$

has additional reaction channel R_4, which captures mRNA degradation at a constant rate γ. R_4 will change only the deterministic dynamics of mRNA level and will not affect miRNA:

$$\dot{x} = -\delta xy - \gamma x, \quad \dot{y} = -y,$$

where γ is the natural degradation rate of mRNA. The sample trajectories of this model compared to one with $\gamma = 0$ are shown in Fig. 4.5 (left panel).

The evolution of the probability density function $p(x, y, t)$ is given by the following equation:

$$\frac{\partial p(x,y,t)}{\partial t} = \frac{\partial}{\partial x}\left((\delta xy + \gamma x)p(x,y,t)\right) + \frac{\partial}{\partial y}\left(yp(x,y,t)\right) - \alpha(x,y,t)p(x,y,t) + J_{\text{in}}(x,y,t), \quad (4.29)$$

where J_{in} is identical to (4.4), since stochastic dynamics are not influenced by the γ. The subsequent approach remains the same to one in Section 4.3. It allows us to convert (4.29) into ODE with respect to the function $f(\psi)$ (defined in (4.20)) with the property $f(0) = -\mathbb{E}(X)$:

$$(\psi - \delta)f'(\psi) + f(\psi)\frac{(\alpha\beta + \gamma\beta)\psi + \gamma}{\beta\psi + 1} = -\frac{\alpha\beta}{(\beta\psi + 1)^{\alpha+2}} = 0. \quad (4.30)$$

Since for any non-homogeneous linear ODE

$$f'(\psi) + f(\psi)f_0(\psi) = g(\psi)$$

a general solution is given by

$$f(\psi) = e^{-F(\psi)}\left(\int_0^{\psi} e^{F(x)}g(x)\,\mathrm{d}x + C\right), \quad F(\psi) = \int f_0(\psi)\,\mathrm{d}\psi\,, \quad (4.31)$$

we obtain the homogeneous solution of (4.30), which has a similar form as one in (4.22):

$$f_H(\psi) = C\,(\beta\psi + 1)^{-q}\,(\psi - \delta)^{-\gamma - \delta\beta q}\,.$$

The significant difference appears in the particular solution $f_p(\psi)$, which is given by the integral:

$$\begin{aligned} f_p(\psi) &= (\beta\psi + 1)^{-P-\alpha-2}(\psi - \delta)^{-Q-1}\int_0^{\psi}(\beta x + 1)^P(x - \delta)^Q\,\mathrm{d}x \\ &= (\beta\psi + 1)^{-P-\alpha-2}(\psi - \delta)^{-Q-1}I(\psi), \end{aligned}$$

where $P = -\frac{\alpha\beta\delta}{\delta\beta+1} - 2$ and $Q = \frac{\alpha\beta\delta}{\delta\beta+1} + \gamma - 1$.

The hypergeometric function ${}_2F_1$ has the suitable representation as an integral with variable upper bound (See Section 2.1.3). We use substitution $y = x - \delta$ to bring the integral $I(\psi)$ to the required form (2.19). It leads to $a = -P$, $b = Q + 1$, and $(\kappa u + 1) = \frac{\beta\psi+1}{\delta\beta+1}$, which satisfy conditions in (2.19). Then we evaluate the

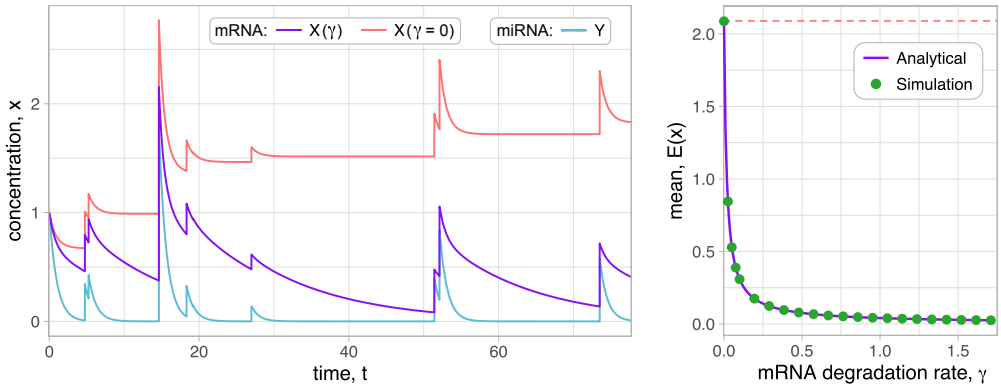


Figure 4.5: *(Left)* Sample trajectories of naturally degrading mRNA as $\gamma = \delta/2$ (violet line) compared to the stable mRNA as $\gamma = 0$ (red line); miRNA concentration is not affected by and its dynamics remains the same (blue line). *(Right)* Violet line is (4.32), red dashed line correspond to (4.25). Green dots are values obtained by simulation, which is constructed and tuned according to Appendix A. Parameters are following: $\delta = 0.4,\ \alpha = 0.1,\ \beta = 0.5$.

integral:

$$I(\psi) = (1+\delta\beta)^P \frac{(\psi-\delta)^{Q+1}}{Q+1}\, {}_2F_1\left(-P, Q+1, Q+2; \frac{\beta(\delta-\psi)}{\delta\beta+1}\right) + C_1,$$

where appearance of a constant C_1 is caused by the substitution $y = x - \delta$, which changed the interval of integration to $(-\delta, \psi - \delta)$. Then $I(\psi)$ was split into two integrals with middle boundary equal to zero: the first one was rewritten as the hypergeometric function as per (2.19); the second one is independent of ψ and denoted by C_1, afterwards C_1 is absorbed by the arbitrary constant C in (4.31). We obtain:

$$\begin{aligned} f_p(\psi) = &- \frac{\alpha\beta(\beta\psi+1)^{-(\alpha+1)}}{\gamma+\beta\delta(\alpha+\gamma)} \left(\frac{\beta\psi+1}{\delta\beta+1}\right)^{\xi+1} \\ &\times {}_2F_1\left(\xi+\gamma, \xi+2, \xi+\gamma+1; \frac{\beta(\delta-\psi)}{\beta\delta+1}\right), \qquad \xi = \frac{\alpha\beta\delta}{1+\beta\delta}. \end{aligned}$$

After using an argument transformation from Property 2.2(b), the particular solution takes form:

$$f_p(\psi) = -\frac{\alpha\beta(\beta\psi+1)^{-(\alpha+1)}}{\gamma+\beta\delta(\alpha+\gamma)}\, {}_2F_1\left(1, \gamma-1, \xi+\gamma+1; \frac{\beta(\delta-\psi)}{\beta\delta+1}\right).$$

Finally, we set $\psi = 0$ and use (4.18) to obtain the mean protein concentration:

$$\mathbb{E}(X) = \frac{\alpha\beta}{\gamma+\beta\delta(\alpha+\gamma)}\, {}_2F_1\left(1, \gamma-1, \xi+\gamma+1; \frac{\xi}{\alpha}\right). \tag{4.32}$$

In particular, if $\gamma = 1$, i.e., if miRNA and mRNA have the same degradation rate,

then the hypergeometric function is equal to one and mean mRNA concentration is:

$$\mathbb{E}(X) = \frac{\alpha\beta}{1 + \beta\delta(\alpha + 1)}.$$

Under the assumption about stable mRNA ($\gamma = 0$), the hypergeometric function becomes a polynomial:

$${}_2F_1\left(1, -1, \xi + 1; \frac{\xi}{\alpha}\right) = \frac{\alpha\beta\delta + 1}{\alpha\beta\delta},$$

and (4.32) reduces to (4.25).

Conclusions

We described the mRNA-miRNA interaction by a Markov process with stochastic jumps and deterministic drift which incorporates the main biochemical features of the regulatory motifs. In contrast with previous uses of the framework, the current model is two-dimensional, and requires, in particular, the use of the double Laplace transform. Another distinction is the appearance of the convolution about axis, which is the nonlocal term of the master equation. This appears because of a perfect correlation between mRNA and miRNA bursts, i.e., both species are produced simultaneously and in equal proportions, which notably affected the solution approach.

We have studied the steady-state moments of the species in feed forward loop, with the main focus on their analytical expressions. Since the model has nonlinear kinetics, the moments equations are not closed. Nevertheless, the Laplace transform still provided a convenient means to derive the explicit moments. Thus, using the current framework a variety of interactions between species forming an IFFL [93], e.g. pair holin-antiholin [94], can be also studied.

It is expected that the current approach can be extended to general burst size distributions or further regulation of the underlying processes. Furthermore, we believe that the simple model can help in understanding expanded descriptions that reflect realistic cell scenarios, especially ones involving cell growth and cell division [95], [96].

Chapter 5

Positive feedback on dilution

In this chapter, we study the positive feedback on protein dilution that causes differences in protein statistics between single-cell and population perspectives, compared to the one in Chapter 3. This effect can arise from various causes, such as protein burden or cell resource exhaustion. As a result, the cell growth slows down due to high protein levels and results in a slower dilution rate.

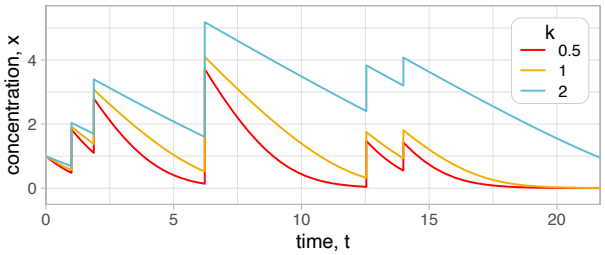
The chapter consists of two parts: in the first one (Sections 5.1–5.2), we study univariate models (single cell and population), which track only the protein concentration. We implicitly assume that the cell division occurs randomly, but the high protein level decreases not only the growth rate but also the rate of cell proliferation. First, we derive the stationary protein distribution in the single-cell model with standard assumptions (Section 2.6). Then, we relax the restriction on the burst size distribution from being solely exponential and assume any non-negative distribution with a known mean instead; we derive expressions for the stationary moments in this generalized model. Second, we derive the PDF of the protein concentration in the cell population, find the condition for population prosperity, and determine the rate of its growth.

Next, we analyze the weak side of the univariate model: the probability that a cell becomes infinitely large, which contradicts biological principles. In the second part of this chapter, we eliminate this problem by studying bivariate models, which capture both protein and cell volume dynamics. Now, the cell divides when it reaches a critical volume. We derive PDFs of the protein concentration and cell volume; we discover that despite a different division mechanism, the protein distribution remains the same as in univariate models. Finally, we compare the effect of the positive feedback with the unregulated model of gene expression.

This chapter is based in part on the articles previously published in International Conference on Computational Methods in Systems Biology [86] and American Control Conference [97], the contents of which have been adapted and expanded.

$$R_1 : \emptyset \xrightarrow{\alpha} \mathrm{B} \times \mathrm{X}$$
$$R_2 : \mathrm{X} \xrightarrow{\gamma(X)} \emptyset$$

(a)



(b)

Figure 5.1: (a) Description of the process as a chemical reaction system. (b) Sample trajectories of the X concentration for different values of the parameter k, while the size and occurrence time of bursts remain unchanged; parameters of the simulation are following: $\alpha = 1/5$, $\beta = 2$.

5.1 Single cell model

The model of a single cell studied in this section is an extension of the basic (unregulated) model described in Section 2.4. The protein concentration $x(t)$ evolves in time stochastically according to the following rules. Protein synthesis occurs in random discrete events with stochastic rate α. Each burst event creates a jump in the protein concentration, its size B is drawn from exponential distribution with mean β. In between successive bursts, the protein concentration is diluted due to the cell growth. In some cases, production of the cell puts greater burden on a cell and causes its slower growth (biological details are provided in Chapter 1). To describe the feedback in dilution, we define the dilution rate as a Hill function of protein concentration. Finally, we assume long-living protein and neglect its natural degradation rate. Then $x(t)$ decays deterministically according to the differential equation

$$\dot{x} = -x\gamma(x), \quad \gamma(x) = \frac{1}{1+kx}. \tag{5.1}$$

The constant $k > 0$ is known as the feedback strength, which characterizes how steeply the dilution rate reacts to increasing protein level. Since half of the maximum dilution rate is achieved at $x = 1/k$, a stronger feedback implies that less protein is needed to halve the cell expansion rate. Increasing the amount of protein leads to slower dilution, concentration decays slower, and hence positive feedback. Without loss of generality, we assume that the maximum dilution rate is equal to one. This effect of increasing the feedback strength k on the protein concentration trajectories is exemplified in Figure 5.1b. The special case $k = 0$ corresponds to the absence of feedback, when the cell grows exponentially and independently of the protein level. This is an unregulated expression, in which the steady-state protein distribution $p(x)$ is known to follow a gamma distribution (Section 2.6).

The model can be thought of as a piecewise deterministic formulation of a simple chemical reaction network with bursty production reaction and deterministic

decay reaction (Figure 5.1a). The time evolution of the probability density $p(x,t)$ is described by the Chapman-Kolmogorov equation (2.24):

$$\frac{\partial p(x,t)}{\partial t} = \frac{\partial}{\partial x}\left(x\gamma(x)p(x,t)\right) - \alpha p(x,t) + \frac{\alpha}{\beta}\int_0^x e^{-(x-y)/\beta}p(y)\,\mathrm{d}y\,. \tag{5.2}$$

We assume that $t = 0$ is the initial time at which the concentration of protein $x(t)|_{t=0} = x_0$ is fixed and known. Therefore, the pdf is initially equal to the delta function located at x_0:

$$p(x, t = 0) = \delta(x - x_0).$$

In order to find the stationary distribution, we write it as a probability conservation equation. It is done by gathering the last two terms of (5.2) using the Leibniz integral rule (2.38), then we set $\frac{\partial p}{\partial t} = 0$ and integrate the equation, which yields an integral equation:

$$\frac{x}{1+kx}p(x) = \alpha\int_0^x e^{-(x-y)/\beta}p(y)\,\mathrm{d}y\,. \tag{5.3}$$

We use the Laplace transform approach to solve equation (5.3), we obtain the steady-state distribution and derive formulae for protein moments in case of an unknown burst size distribution. First, we move the denominator $1 + kx$ on the right-hand side of the equation, after we perform rearrangement under the integral sign in the following way:

$$xp(x) = \alpha\int_0^x (1 + k(x-y) + ky)e^{-(x-y)/\beta}p(y)\,\mathrm{d}y\,, \tag{5.4}$$

which allows us to represent the right side of equation as the sum of three distinct convolutions (see Definition 2.13):

$$xp(x) = \alpha\left((B * p)(x) + k(xB * p)(x) + k(B * xp)(x)\right), \tag{5.5}$$

where by $B(\cdot)$ we denoted the exponential function, i.e., $B(x) = e^{-x/\beta}$, the Laplace transform of which is known (Property 2.4(a)).

Let us define an image $P(s)$ as a function of a Laplace variable s:

$$P(s) = \int_0^\infty p(x)e^{-sx}\,\mathrm{d}x\,. \tag{5.6}$$

Applying the Laplace transform to (5.5), we find:

$$\frac{\mathrm{d}P(s)}{P(s)} = \left(\frac{1}{s + 1/\beta} - \frac{\alpha + 1}{s + 1/\beta - \alpha k}\right). \tag{5.7}$$

The general solution of (5.7) is given by:

$$P(s) = C\frac{s + 1/\beta}{(s + 1/\beta - \alpha k)^{\alpha+1}}, \tag{5.8}$$

where C is an arbitrary constant. Note that the right hand side is a power function of the Laplace variable s, which is shifted by value $\eta = 1/\beta - \alpha k$. In order to return to the original function $p(x)$, we apply the inverse Laplace transform to the general solution (5.8) (see Property 2.4(b)):

$$p(x) = Ce^{-\eta x}x^{\alpha-1}\frac{1 + kx}{\Gamma(\alpha)}.$$

We set C so that the integral of the right side is equal to one, i.e. that $p(x)$ is probability density function on the non-negative domain; the steady state probability density function of protein concentration is given by:

$$p(x) = \frac{\eta^2\beta}{\Gamma(\alpha)}(\eta x)^{\alpha-1}e^{-\eta x}(1 + kx), \quad \eta = 1/\beta - \alpha k > 0. \tag{5.9}$$

The imposed condition $\eta > 0$ is necessary for the existence of $p(x)$. In terms of the model parameters, the condition read

$$\alpha\beta < \frac{1}{k}, \tag{5.10}$$

which can be interpreted from a biological point of view as the requirement that the average production flow $\lambda\beta$ be less than the effective degradation flow (given by $\lim_{x\to\infty} \mathrm{d}x \,/\, \mathrm{d}t = 1/k$). Otherwise, the protein dilution is not fast enough to compensate for the protein production rate, and the mean level unboundedly increases over time; thus, the stationary distribution does not exist.

In the absence of the degradation regulation, i.e $k = 0$, $p(x)$ becomes probability density function of the unregulated gene expression (Section 2.6):

$$p(x) = \frac{1}{\beta\Gamma(\alpha)}\left(\frac{x}{\beta}\right)^{\alpha-1}e^{-x/\beta}.$$

Multiplying (5.9) by $x^n,\ n \in \mathbb{N}$, and integration it on the interval $[0, \infty)$ give the value of n-th raw moment:

$$\mathbb{E}(x^n) = (1 + nk\beta)\frac{(\alpha)_n}{\eta^n}, \tag{5.11}$$

where $(\alpha)_n = \Gamma(\alpha+n)/\Gamma(\alpha)$ is the Pochhammer symbol (2.2). As the value production rate approaches the maximum degradation rate, i.e., $\alpha\beta \approx 1/k$, the coefficient

of variation is always finite and is less than $\frac{1}{1+\alpha}$. This is because the existence conditions provide a counterbalancing relationship between the feedback strength and the intrinsic noise (stochastic production), thus the noise is finite.

5.1.1 Generalisation to any burst size distribution

Let us return to the initial mater equation of the process, but now we suppose that the random variable B of the burst size has any appropriate distribution, i.e., the probability density function of B is $b(x) : [0, \infty) \to [0, \infty)$ with the expected value is equal to β. The Chapman-Kolmogorov equation in the steady state takes form:

$$\frac{\mathrm{d}}{\mathrm{d}x}\left(x\gamma(x)p(x)\right) = \alpha\left(p(x) - \int_0^x b(x-y)p(y)\,\mathrm{d}y\right). \tag{5.12}$$

We collapse the last two terms in (5.12) as per (2.38) and subsequently integrate it; afterwards, the equation of the concentration distribution $p(x)$ is given by

$$\frac{x}{1+kx}p(x) = \alpha\int_0^x \bar{\mathrm{B}}(x-y)p(y)\,\mathrm{d}y\,.$$

where $\bar{\mathrm{B}}(x)$ is CCDF corresponding to $b(x)$. Performing of the same rearrangement as in (5.4) and application of the Laplace transform to the result yields ODE

$$P'(s) = P(s)\frac{B(s) - kB'(s)}{kB(s) - \frac{1}{\alpha}}, \tag{5.13}$$

where, by analogy with $P(s)$ in (5.6), we denote by $B(s)$ a Laplace image of $\bar{\mathrm{B}}(x)$

$$B(s) = \int_0^\infty \bar{\mathrm{B}}(x)e^{-sx}\,\mathrm{d}x\,. \tag{5.14}$$

The advantages of using the complementary cumulative distribution function (CCDF) and its Laplace image, in the case when the burst size distribution is unknown, is following. First, the mean value of a non-negative random variable, such as the burst size, can be expressed through CCDF as its integral over the $\mathbb{R}^+$, which is related to $B(s)$ as

$$B(0)\big|_{s=0} = \beta.$$

Next, we express the k-th derivative of $B(s)$ in a form containing the raw moments:

$$B^{(k)}(s) = \frac{(-1)^k}{k+1}\left(\left[\bar{\mathrm{B}}(x)e^{-sx}x^{k+1}\right]_0^\infty + \int_0^\infty x^{k+1}(b(x) + s\,\bar{\mathrm{B}}(x))e^{-sx}\,\mathrm{d}x\right),$$

where the first term in brackets is always zero. Then we set $s = 0$ and obtain:

$$B^{(k)}(0) = (-1)^k \frac{\mathbb{E}(B^{k+1})}{k+1}, \tag{5.15}$$

which allows us to obtain exact moments of B even without knowing explicitly the burst size distribution.

In order to obtain statistics of the protein concentration, we use the cumulant generating function

$$K(s) = \ln \mathbb{E}(e^{sx}) \equiv \ln \int_0^\infty p(x) e^{sx} \, \mathrm{d}x \,, \tag{5.16}$$

and the corresponding cumulants of n-th order:

$$\kappa_n = \frac{\mathrm{d}^n K(s)}{\mathrm{d}s^n}\Big|_{s=0}, \quad n \in \mathbb{N},$$

which by definition is equal to mean, the variance, and the third central moment of the random variable X for $n \in \{1, 2, 3\}$, respectively; the trivial case of $n = 0$ leads to $\kappa_0 = 0$. Combining (5.6) and (5.16) yields $K(-s) = \ln P(s)$, so that the cumulants of X are

$$(-1)^n \kappa_n = \frac{\mathrm{d}^n \ln P(s)}{\mathrm{d}s^n}\Big|_{s=0},$$

whereas the natural logarithm of the Laplace image $P(s)$ we obtain directly from ODE (5.13); the cumulants of interest are

$$\begin{aligned}
-\kappa_1 &= \frac{B'(0) - B(0)/k}{B(0) - 1/\alpha k}, \\
\kappa_2 &= B'(0) \frac{B'(0) - 1/(\alpha k^2)}{(B(0) - 1/(\alpha k))^2} - B''(0) \frac{1}{B(0) - 1/(\alpha k)}, \\
-\kappa_3 &= 2(B'(0))^2 \frac{B'(0) - 1/(\alpha k^2)}{(B(0) - 1/(\alpha k))^3} - B''(0) \frac{3B'(0) - 1/(\alpha k^2)}{(B(0) - 1/(\alpha k))^2} + B'''(0) \frac{1}{B(0) - 1/(\alpha k)}.
\end{aligned}$$

Using (5.15), we obtain:

$$\begin{aligned}
\mathbb{E}(x) &= \frac{\mathbb{E}(B^2) + 2\beta/k}{2\hat{\eta}}, \\
\mathrm{Var}(x) &= \nu \frac{\mathbb{E}(B^2)}{2\hat{\eta}^2} + \frac{\mathbb{E}(B^3)}{3\hat{\eta}}, \\
\mu_3(x) &= \nu \frac{\mathbb{E}^2(B^2)}{2\hat{\eta}^3} + \frac{\mathbb{E}(B^3)}{3\hat{\eta}^2}(\mathbb{E}(B^2) + \nu) + \frac{\mathbb{E}(B^4)}{4\hat{\eta}},
\end{aligned}$$

where $\mu_3(x)$ is the third central moment and $\nu = \mathbb{E}(B^2)/2 - 1/(\alpha k^2)$; the constant $\hat{\eta} = 1/\alpha k - \beta > 0$ here represents the existence condition with the same interpretation as for (5.10).

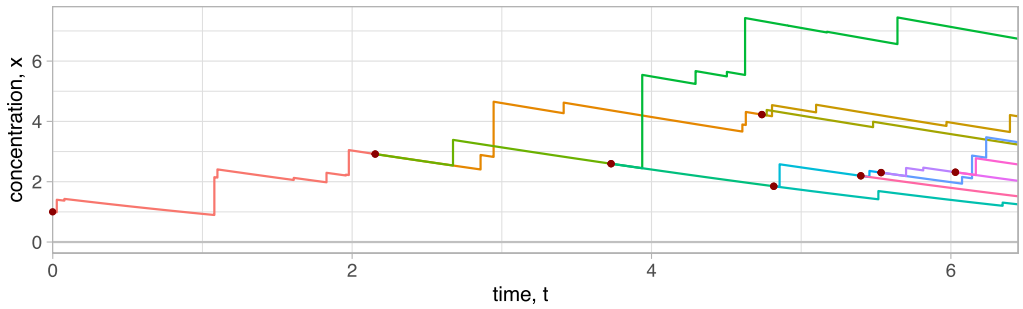


Figure 5.2: The time evolution of a sample population from a single cell to eight of them. The red dots mean time and concentration, at which given mother cell was divided into two daughter cells. An unique colour was assigned to each cell, so to show how the protein concentration changes while it exists. Parameters of the simulation are following: $\alpha = 1.9,\ \beta = 0.5,\ k = 1\ x_0 = 1$.

5.2 Population model

To extend the single-cell framework to a population one, we elaborate on the dynamics of cell proliferation. Firstly, during cell division, a mother cell splits into two identical daughter cells, each inheriting identical protein level to those of the mother. During the cell cycle, the protein dynamics remains identical to those defined in Section 5.1. Finally, the probability that a cell divides depends on its protein content and it follows the same rate as dilution. Therefore, division events are modelled by a non-homogeneous Poisson process with rate $\gamma(x) = 1/(1+kx)$ and the net population growth rate $g(x,t)$ is also equal to $\gamma(x)$. The sample trajectories of cells in such population are shown in Fig. 5.2.

The expected population density $h(x,t)$ of cells with concentration x at time t satisfies the population balance equation (2.34):

$$\begin{aligned}\frac{\partial h(x,t)}{\partial t} = \frac{\partial}{\partial x}\left(\frac{x}{1+kx}h(x,t)\right) + \frac{h(x,t)}{1+kx} \\ + \alpha\int_0^x b(x-y)h(y,t)\,\mathrm{d}y - \alpha h(x,t),\end{aligned} \tag{5.17}$$

where $b(\cdot)$ is the probability density function of the random burst size, which was described in details in (5.12). The population balance equation differs from the Chapman-Kolmogorov equation by the inclusion of a growth term, which is equal to the product of the cell count $h(x,t)$ and the dilution rate $1/(1+kx)$.

For large enough t, it is possible to represent the population function $h(x,t)$ in a separable form, i.e.,

$$h(x,t) = e^{\lambda t}p(x), \tag{5.18}$$

then (5.17) becomes

$$\frac{p(x)}{1+kx} - \lambda p(x) + \frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{xp(x)}{1+kx}\right) - \alpha\frac{\mathrm{d}}{\mathrm{d}x}\int_0^x \bar{\mathrm{B}}(x-y)p(y)\,\mathrm{d}y = 0, \tag{5.19}$$

where the last term and the function $\bar{\mathrm{B}}(\cdot)$ were derived as per (2.38). Among the possible eigenvalues λ, we will look for the one that has the largest real part – the principal eigenvalue. In the large time limit, the principal eigenvalue gives the effective growth rate, and the principal eigenvector $p(x)$ gives the population distribution of the protein distribution. Below we find explicit results (5.28) and (5.31) for both.

We define an auxiliary function,

$$q(x) = \frac{p(x)}{1+kx}, \tag{5.20}$$

and substitute it into (5.19); afterwards we apply the Laplace transform, which yields

$$Q(s) - \lambda P(s) - sQ'(s) - \alpha sB(s)P(s) = 0, \tag{5.21}$$

where $P(s)$ and $Q(s)$ are the Laplace images of functions $p(x)$ and $q(x)$, respectively, defined by (5.14). Applying the Laplace transform directly to the function $q(x)$ (5.20), one can obtain:

$$P(s) = Q(s) - kQ'(s),$$

which is used to transform (5.21) into an ODE for $Q(s)$:

$$\frac{\mathrm{d}Q(s)}{Q(s)} = \frac{\lambda + \alpha sB(s) - 1}{\lambda k + \alpha ksB(s) - s}\,\mathrm{d}s\,. \tag{5.22}$$

Despite the separable form of this equation, additional complexity is brought by the generalisation of the burst size distribution. However, we assume that $\bar{\mathrm{B}}(x)$ is CCDF of the exponential distribution with mean β, the Laplace image of which is known; then (5.22) simplifies to

$$\frac{\mathrm{d}Q(s)}{Q(s)} = \frac{\frac{1-\lambda}{\beta} + s(1-\rho)}{-\frac{\lambda k}{\beta} - s(k\rho - \frac{1}{\beta}) + s^2}\,\mathrm{d}s\,, \quad \rho = \alpha + \lambda. \tag{5.23}$$

The solution approach requires the partial fraction decomposition of the right-hand side of (5.23). The quadratic in the denominator has two real roots

$$s_{1,2} = \frac{1}{2}\left(k\rho - \frac{1}{\beta} \pm \sqrt{D}\right), \quad D = (k\rho + \frac{1}{\beta})^2 - \frac{4\alpha k}{\beta}, \tag{5.24}$$

where s_1 is strictly positive and s_2 is strictly negative for any positive values of

parameters α, β, λ and k. The decomposition of (5.23) takes the form:

$$\frac{\mathrm{d}Q(s)}{Q(s)} = \frac{A_1}{s - s_1} + \frac{A_2}{s - s_2}, \tag{5.25}$$

where A_1 and A_2 are defined by

$$A_{1,2} = \frac{1-\rho}{2} \pm \frac{1 - \rho + 2\alpha + \beta k \rho(1-\rho)}{2\beta\sqrt{D}}. \tag{5.26}$$

The solution of the ODE (5.25) is

$$Q(s) = C(s - s_1)^{A_1}(s - s_2)^{A_2}. \tag{5.27}$$

The Laplace transform (5.27) has to be analytic in the complex half-plane $\mathrm{Re}(s) > 0$. Therefore, $A_1 \in \{0, 1, 2, \ldots\}$. In particular, the principal eigenvalue is obtained by setting $A_1 = 0$, which implies

$$\lambda = 1 - \frac{\alpha k \beta}{k\beta + 1}, \quad \lambda > 0. \tag{5.28}$$

The imposed condition on the eigenvalue λ, which represents the exponential growth rate of cell numbers in a population, specifies that λ must be non-negative for the population to thrive. We show that $A_1 = 0$ indeed provides the principal eigenvalue and study further eigenvalues for $A_1 \in \mathbb{N}$ in Section 5.2.1.

We substitute (5.28) into (5.24) and (5.26) and simplify; the resulting values are strictly negative, and we introduce additional symbols σ, ξ for their opposite values, which are strictly positive:

$$\begin{aligned} s_2 &= \frac{\alpha k}{k\beta + 1} - \frac{1}{\beta}, & \sigma &= -s_2 > 0, \\ A_2 &= -\frac{\alpha}{k\beta + 1}, & \xi &= -A_2 > 0. \end{aligned} \tag{5.29}$$

Inserting $s = 0$ into (5.21) and using the normalisation condition $P(0) = 1$ yield

$$Q(0) = \lambda, \tag{5.30}$$

which is used to find the value of C in (5.27).

Applying the inverse Laplace transform to (5.27), we obtain

$$q(x) = \frac{\beta\sigma^2}{\Gamma(\xi)} e^{-\sigma x} (\sigma x)^{\xi - 1},$$

i.e., by relation (5.20),

$$p(x) = (1 + kx)\frac{\beta\sigma^2}{\Gamma(\xi)}e^{-\sigma x}(\sigma x)^{\xi-1}, \tag{5.31}$$

where the constants σ and ξ are defined by (5.29). Note that without regulation, as $k = 0$, the PDF of the protein concentration in the population is identical to the unregulated one in the single cell (2.40). We use the definition $\mathbb{E}(x^n) = \int_0^\infty x^n p(x)\,\mathrm{d}x$ to obtain the n-th raw moment expressions for any given n:

$$\mathbb{E}(x^n) = (1 + nk\beta)\frac{(\xi)_n}{\sigma^n},$$

where $(\xi)_n$ is the Pochhammer symbol (2.2), and the integral is convergent only is $\sigma > 0$. The mean concentration of the protein is

$$\mathbb{E}(x) = \frac{\alpha\beta(1 + k\beta)}{1 + k\beta - \alpha\beta k}.$$

Leaving technical reasons aside, the necessary condition for a cell to survive is finitness of its mean value; on the other side, as it was mentioned, the population exists only if $\lambda > 0$. Both produce an identical existence condition:

$$\alpha - 1 < \frac{1}{k\beta}. \tag{5.32}$$

Similarly to the single-cell model, the population exists, if certain balance between the production and dilution rates is maintained. Otherwise, the protein dilution is not fast enough to compensate for the protein production rate, and the mean level unboundedly increases over time; thus, the stationary distribution does not exist. Subsequently, the cell with high protein concentration is overburdened and has low probability to proliferate, and the population growth terminates (this statement is further elaborated in Section 5.2.2). The population always exists in the low frequency regime ($\alpha < 1$), but otherwise the population distribution exists for values of λ that satisfy (5.32). This is a weaker condition than (5.10) that was found necessary for the existence of the stationary distribution in a single cell setting: a population distribution may exist even if the single cell distribution does not.

In Appendix B, we further study the statistics of the single-cell and population frameworks. Since cells with higher protein levels proliferate more slowly, there is an over-representation of cells with low protein concentrations. This results in a lower mean concentration in the population framework compared to the single-cell framework. After analyzing protein noise using the coefficient of variation, we find that protein concentration at the population level is relatively noisier than that in

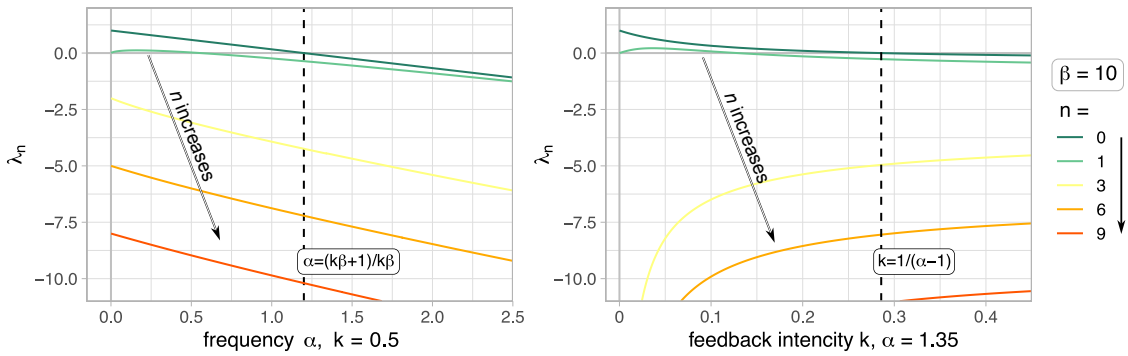


Figure 5.3: The eigenvalues λ_n^* are roots of (5.34) with the greatest real part. *(Left)* λ_n^* as function of the burst frequency α, other parameters are fixed: $k = 0.5$, $\beta = 10$; *(Right)* λ_n^* as the function of the effective degradation rate $k\beta$ ($\alpha = 1.35$). Vertical dashed lines represent the existence condition $\sigma > 0$ in both parameter spaces.

the single-cell approach. This result can also explain why, for the same parameters, the mean protein concentration in the population is lower than that at the single-cell level. Finally, while the single-cell distribution is more heavy-tailed (compared to the gamma distribution), the population distribution becomes more light-tailed due to a larger proportion of cells with low protein concentration.

5.2.1 Other eigenvalues

Let us study $A_1(\rho)$, where the variable ρ is the shifted by burst frequency eigenvalue λ, i.e. $\rho = \alpha + \lambda$:

$$A_1(\rho) = \frac{1-\rho}{2} + \frac{2\alpha + (k\rho+1)(1-\rho)}{2\sqrt{(k\rho+1)^2 - 4\alpha k}}, \quad k \equiv k\beta.$$

In addition to $A_1 = 0$, permissible values of the function A_1, according to (5.27), are natural numbers; we find further eigenvalues by setting $A_1(\rho) = n$:

$$(2n - 1 + \rho)\sqrt{(k\rho+1)^2 - 4\alpha k} = 2\alpha + (k\rho+1)(1-\rho), \quad n \in \mathbb{N}. \tag{5.33}$$

Then the eigenvalues are the roots of a following cubic polynomial:

$$\begin{aligned} p_n(\rho) =& nk^2\rho^3 + k\big(n(n-1)k + 2n\big)\rho^2 \\ &+ \big(2kn(n-1) + n + \alpha(k + 1 - 4kn)\big)\rho \\ &+ n(n-1)(4\alpha kn + 1) - \alpha(k + 1 + \alpha). \end{aligned} \tag{5.34}$$

Although we do not provide analytical proof, numerical computations provide the following properties of (5.34). The discriminant Δ_n of the polynomial $p_n(\rho)$ is strictly negative for any permissible set of values (α, β, k) and $n > 0$. It implicates

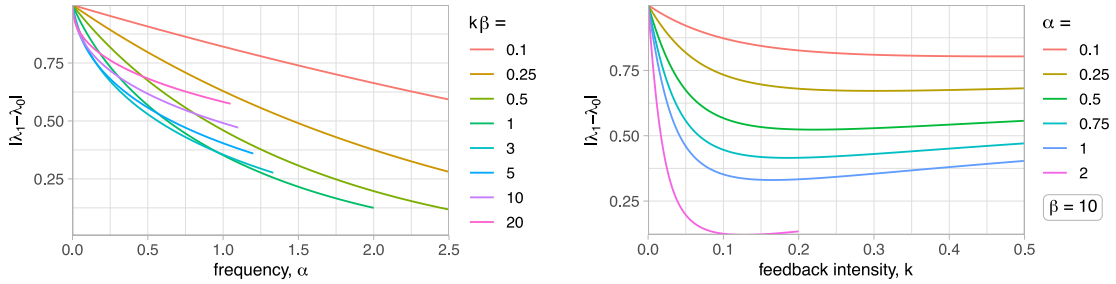


Figure 5.4: Spectral gap as the function of the burst frequency (left panel) and the feedback intensity (right panel). Break of the lines occurs, when the existence condition (5.32) is reached.

that (5.34) has a single real root and a pair of complex conjugate roots. Moreover, the greatest real part has the real root, which we denote by λ_n^*, for which hold the following inequality:

$$\lambda_n^* < \lambda_{n+1}^*, \quad n \in \mathbb{N}_0.$$

In Figure 5.3 are shown eigenvalues λ_n^* as functions of the burst frequency α (left panel) and the effective degradation rate $k\beta$ (right panel). Note, that $p_0(n)$ indeed provides the dominant eigenvalue.

Let us take a closer look at $p_1(\rho)$:

$$p_1(\rho) = k^2\rho^3 + 2k\rho^2 + (1 - 3\alpha k + \alpha)\rho - \alpha(k + 1 + \alpha). \tag{5.35}$$

The discriminant of (5.35) is:

$$\Delta_1 = -\alpha(1 + \alpha - k)^2 k^2 (4 + 27\alpha k^2),$$

which is, as all involved constants are positive, strictly negative. It indicates presence of one real root and two complex conjugate roots.

Firstly, to find λ_1, we solve (5.35) using the `R` package `polyroot` and subsequently choosing the real root (as is is mentioned in Section 5.2.1 it has the greatest real part). In Figure 5.4 are shown graphs of spectral gap $|\lambda_1 - \lambda_0|$. Firstly, we treat it as a function of the burst frequency α, then for each fixed value of $k\beta$, it is strictly decaying on an interval of permissible α (left panel). Next, for given set of values for α, we study Λ as a function of the production rate $k\beta$, which reaches maximum around zero. Smaller spectral gap then corresponds to such values of the parameters, which are close to the population existence condition (5.32).

5.2.2 Consequences for the cell volume

Here we explore the implicit consequences for the cell volume. During the cell cycle of a cell that begins at $t = t_b$ and ends at $t = t_e$, the volume increases exponentially

with rate that is equal to the dilution rate:

$$\frac{dv}{dt} = \gamma(x(t))v, \quad t_b < t < t_e.$$

The cell volume $v(t)$ is a continuous function for $t_b < t < t_e$. It has a discontinuous derivative at protein burst time points because of the discontinuity in the protein concentration $x(t)$.

Let $v_b = v(t_b)$ and $v_e = v(t_e)$ be the initial and final volume of the cell, respectively. The initial volume v_b is assumed to be known. The final volume v_e is a random variable because it depends on t_e, which is random. The propensity to end the cell cycle (2.30) can be rewritten as

$$Prob[v_e \in (v, v + dv)|v_e > v] = \frac{dv}{v} + o(dv), \quad v > v_b.$$

This implies that the complementary cumulative distribution function of v_e satisfies

$$Prob[v_e > v] = \exp\left(-\int_{v_b}^{v} \frac{d\tilde{v}}{\tilde{v}}\right) = \frac{v_b}{v}, \quad v > v_b,$$

which can be expressed in terms of probability density function of v_e as

$$f_{v_e}(v) = -\frac{d}{dv} Prob[v_e > v] = \frac{v_b}{v^2}, \quad v > v_b.$$

We see that the distribution of the final volume is heavy tailed. The mean value of the final volume is infinite. To address this, let us include the volume explicitly in the model and do a different volume sensing mechanism.

5.3 Bivariate model with explicit cell volume $v(t)$

Here we introduce bivariate model to the dilution feedback problem. We preserve all dynamics of protein concentration in the single cell, which were described in Section 5.1. During cell division, a mother cell splits into two identical daughter cells, each inheriting half of the mother's volume and protein amount; thus, the protein level in the daughter cells is equal to the mother's. The volume $v(t)$ remains strictly increasing during the cell cycle. In particular, between successive bursts, the cell volume increases with a concentration dependent rate and the concentration is diluted according to differential equations

$$\frac{\dot{v}}{v} = -\frac{\dot{x}}{x} = \gamma(x) = \frac{1}{1 + kx}. \tag{5.36}$$

The rule (5.37), which triggers division event, is based on a biological concept of a Sizer [98]. When the cell volume reaches a given threshold $2v^*$, a cell division event occurs, and the volume is immediately halved

$$v = 2v^* \to v = v^*. \tag{5.37}$$

Typical protein and volume trajectories of a single cell line are shown in Figure 5.5a.

Bivariate Chapman–Kolmogorov equation

In the single-cell model, we are looking at a single cell line. We do not follow the other daughter cell that is created in a cell division. The model is then a piecewise deterministic bivariate Markov process $\mathbf{x} = (x(t), v(t)) \in R^2_+$, which was introduced in Section 2.4. In this case, the deterministic motion of $\mathbf{x}$ (2.21) is given by (5.36). The random jumps of $x(t)$ remain exponentially distributed; the volume $v(t)$ has a purely deterministic trajectory with discontinues due to the cell division rule (5.37) and nonsmooth points caused by bursts.

Let $f(x, v, t)$ denote the probability density function (PDF) of the random vector $\mathbf{x} = (x(t), v(t))$ at time t. We have the initial condition

$$f(x, v, t = 0) = \delta(x - x_0)\delta(v - v_0),$$

where $x_0 > 0$ and $v_0 > 0$ are known initial concentration and volume. We thereby assume that the initial volume is less that the critical volume at which a cell division event is triggered: $v_0 < 2v^*$.

For $t > 0$, PDF $f(x, v, t)$ satisfies a bivariate Chapman–Kolmogorov equation (2.24), which is fitted for the dilution problem in Section 5.3.1. We assume that the protein concentration and cell volume are independent random variables at steady state. Thus a unique normalised stationary solution to the Chapman–Kolmogorov equation (the stationary distribution) can be found in a separable form:

$$f(x, v) = p(x)g(v), \quad x > 0, \quad v^* < v < 2v^*,$$

where $p(x)$ and $g(w)$ are PDFs of the protein concentration and the cell volume respectively.

The stationary distribution existence condition (2.29) requires the process to be irreducible and aperiodic. However, in the absence of feedback ($k = 0$), the cell cycle length is equal to the doubling time $T = \ln 2$ of the exponential function, and the cell volume $v(t)$ is a T-periodic function (Figure 5.5b). Because of the periodicity, the time-dependent pdf $f(x, v, t)$ does not converge to the stationary distribution

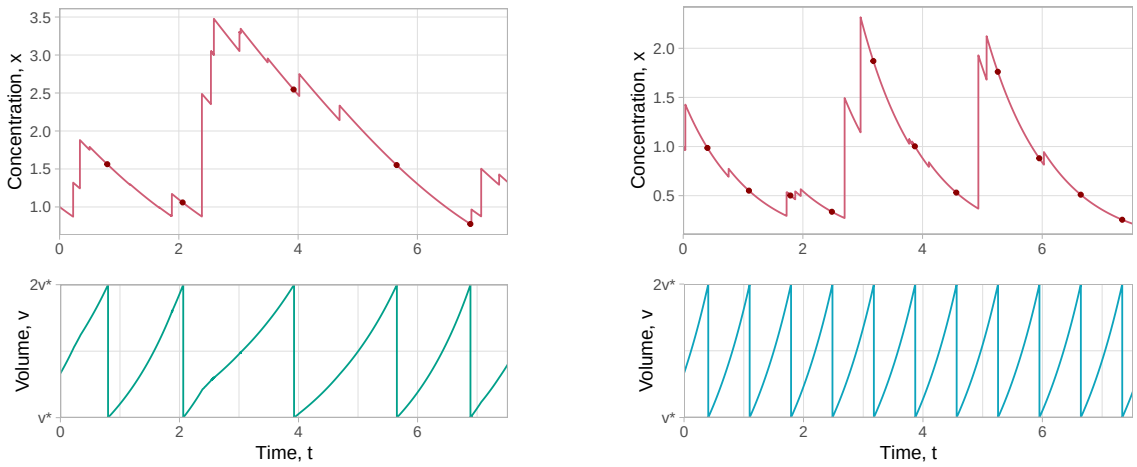


Figure 5.5: Sample trajectories of the concentration $x(t)$ and the volume $v(t)$ in a single cell affected by (a) the strong feedback as $k = 2$, (b) the absence of the feedback, i.e., $k = 0$; other parameters of simulations are $\alpha = 0.9$, $\beta = 5$.

$f(x, v)$ if $k = 0$. The inclusion of feedback breaks the periodicity (Figure 5.5a). Therefore, for $k > 0$,

$$f(x, v, t) \sim f(x, v), \quad t \to \infty,$$

i.e. the time-dependent distribution converges to the stationary distribution in the large-time limit (the ergodic property).

5.3.1 Single cell with explicit volume dynamics

The probability for the cell to be of volume $v^* \leq v \leq 2v^*$ and to have the protein concentration $x > 0$ at time $t > 0$ is given by the joint probability density function is $f(x, v, t)$; its time evolution is described by Chapman-Kolmogorov equation:

$$\begin{aligned}\frac{\partial f(x,v,t)}{\partial t} = \frac{\partial}{\partial x}\left(x\gamma(x)f(x,v,t)\right) - \frac{\partial}{\partial v}\left(v\gamma(x)f(x,v,t)\right) \\ + \frac{\alpha}{\beta}\int_0^x e^{-\frac{x-y}{\beta}} f(y,v,t)\,\mathrm{d}y - \alpha f(x,v,t).\end{aligned} \tag{5.38}$$

The initial and boundary conditions are:

$$f(x, v, 0) = \delta(x - x_0)\delta(v - v_0), \tag{5.39a}$$

$$f(x, 2v^*, t) = \frac{1}{2} f(x, v^*, t), \tag{5.39b}$$

where x_0 is initial protein concentration in the cell, v_0 is its initial volume. The Chapman–Kolmogorov equation (5.38) is a partial integro-differential equation. The differential operator on the right hand side of (5.38) drives the drift of the probability mass due to the deterministic flow (5.36). The integral operator in (5.38)

provides the transfer of probability mass due to instantaneous protein bursts. The boundary condition (5.39b) captures the halving of cell volume in cell division. Integrating (5.38) over the state space $(x, v) \in (0, \infty) \times [v^*, 2v^*]$ confirms that the total probability $\int_0^\infty \int_{v^*}^{2v^*} f(x, v, t)\,\mathrm{d}x\,\mathrm{d}v$ remains constant over time for solutions $f(x, v, t)$ to (5.38). The boundary fluxes thereby cancel thanks to the boundary condition (5.39b).

Since our aim is to find a stationary distribution $f(x, v)$, we set $\partial f/\partial t = 0$ and subsequently apply the Leibniz integral rule (2.38) to the last two terms, which yields an integral equation:

$$\frac{\partial}{\partial x}\left(x\gamma(x)f(x,v)\right) - \frac{\partial}{\partial v}\left(v\gamma(x)f(x,v)\right) - \alpha\frac{\partial}{\partial x}\left(\int_0^x e^{-\frac{x-y}{\beta}} f(y,v)\,\mathrm{d}y\right) = 0, \quad (5.40)$$

the boundary condition for which is the same as (5.39b).

We use the Fourier method of separation variables, i.e. we assume that $f(x, v)$ can be represented as a separable function

$$f(x,v) = \frac{q(x)g(v)}{\gamma(x)},$$

The marginal protein stationary distribution is then $p(x) = q(x)/\gamma(x)$ and that of the cell volume is $g(x)$. We substitute the new representation of $f(x, v)$ into (5.40) and rearrange the result so that on the left-hand side all terms depend only on the variable x, and on the right-hand side only on v:

$$\frac{(xq(x))'}{q(x)} - \frac{\alpha}{q(x)}\frac{\mathrm{d}}{\mathrm{d}x}\int_0^x e^{-\frac{x-y}{\beta}}\frac{q(y)}{\gamma(y)}\,\mathrm{d}y = \frac{(vg(v))'}{g(v)} = \xi. \quad (5.41)$$

Since both sides are independent from one another, each side must be equal to a constant, which we denoted as ξ. We rewrite (5.41) and add the boundary condition on $g(v)$:

$$(vg(v))' + \xi g(v) = 0, \quad g(2v^*) = \frac{1}{2}g(v^*) \quad (5.42a)$$

$$(xq(x))' - \alpha\frac{\mathrm{d}}{\mathrm{d}x}\left(\int_0^x e^{-\frac{x-y}{\beta}}\frac{q(y)}{\gamma(y)}\,\mathrm{d}y\right) - \xi q(x) = 0. \quad (5.42b)$$

First, we solve equation (5.42a), which is an ordinary differential equation of the first order. The general solution is

$$g(v) = Cv^{-\xi-1}; \quad (5.43)$$

upon applying the boundary condition, we find that $\xi = 0$. Since $g(v)$ is a probability distribution, the integral of $g(v)$ over the interval $(v^*, 2v^*)$ must be equal to one.

The stationary distribution of the cell volume is thus

$$g(v) = \frac{1}{\ln(2)v}. \tag{5.44}$$

The mean volume of the cell $\mathbb{E}(V) = \frac{v^*}{\ln(2)}$.

We proceed with the second equation (5.42b); after we substitute $\xi = 0$, it reduces to the Volterra integral equation:

$$xq(x) = \alpha \int_0^x e^{-\frac{x-y}{\beta}}(1+ky)q(y)\,\mathrm{d}y\,,$$

the solution of which is known (see Section 5.1):

$$p(x) = \frac{\eta^2\beta}{\Gamma(\alpha)}(1+kx)(\eta x)^{\alpha-1}e^{-\eta x}, \quad \eta = 1/\beta - \alpha k > 0. \tag{5.45}$$

Bivariate branching process

We elaborate the population framework described in Section 2.5. At the end of the cell cycle triggers a branching (division) event, when the current process (the mother cell) is terminated and replaced with two new bivariate processes (daughter cells):

$$(x_i(t), v_i(t)), \quad t_b^i \le t < t_e^i, \quad i = 1, 2, \ldots,$$

where t_b^i and t_e^i denote the beginning and the end of the cell cycle of the ith cell, and $x_i(t)$ gives the protein concentration at time t of the ith cell and $v_i(t)$ gives its volume. The time t_e in now the moment, when the mother cell volume reaches a given threshold $2v^*$.

General description and properties of the bivariate branching process remains the same as it was described for univariate one in Section 2.5. The composition of the population at time t can be represented by the empirical population density (or empirical measure)

$$m(x, v, t) = \sum_{i:t_b^i<t<t_e^i} \delta(x - x_i(t))\delta(v - v_i(t)).$$

The measure is defined on the two-dimensional state space $(0, \infty) \times [v^*, 2v^*]$. The empirical measure is not normalised:

$$n(t) = \int_{v^*}^{2v^*} \int_0^\infty m(x, v, t)\,\mathrm{d}x\,\mathrm{d}v = \#\{i : t_b^i < t < t_e^i\},$$

gives the total number of cells at time t.

Let us consider the expected value of the empirical population density

$$h(x, v, t) = \mathbb{E}\, m(x, v, t).$$

At the initial time $t = 0$, we have a single cell with non-random initial concentration x_0 and initial volume v_0, so that

$$h(x, v, 0) = m(x, v, 0) = \delta(x - x_0)\delta(v - v_0).$$

For $t > 0$, the expected population density $h(x, v, t)$ satisfies a population balance equation. This is formulated in Section 5.3.2. The large-time behaviour of the population balance equation is characterised by its principal eigenvalue λ and the associated eigenfunction $f(x, v)$ and the adjoint eigenfunction $w(x, v)$. Spectral decomposition implies that the expected (non-random) population density satisfies

$$h(x, v, t) \sim w(x_0, v_0)e^{\lambda t}p(x)g(v), \quad t \to \infty, \tag{5.46}$$

The adjoint eigenfunction $w(x_0, v_0) > 0$ characterises the influence of initial condition. We do not determine its functional form. Similarly to what was said in the single cell model, (5.46) holds only in the aperiodic case ($k \neq 0$).

By the theory of supercritical branching processes [82], [83], the (random) empirical population density satisfies

$$m(x, v, t) \sim W(x_0, v_0)e^{\lambda t}p(x)g(v), \quad t \to \infty, \tag{5.47}$$

where $W(x_0, v_0) > 0$ is a random variable dependent on initial data such that $\mathbb{E}\, W(x_0, v_0) = w(x_0, v_0)$. Equivalently,

$$n(t) \sim W(x_0, v_0)e^{\lambda t}, \quad \frac{m(x, v, t)}{n(t)} \sim p(x)g(v), \quad t \to \infty. \tag{5.48}$$

The total population $n(t)$ increases exponentially. The influence of the initial condition and the initial low population noise is encompassed in the random pre-exponential factor $W(x_0, v_0)$. The protein concentration and cell volume are independent of each other in a large population. The normalised protein concentration distribution is the same as in the univariate population model.

5.3.2 Population with explicit volume dynamics

The expected population density $h(x, v, t)$ of cells with concentration $x > 0$ and volume $v^* \leq v \leq 2v^*$ at a particular point of time $t \geq 0$ satisfies a population balance equation. In the interior of the state space ($x > 0$, $v^* < v < 2v^*$), the population

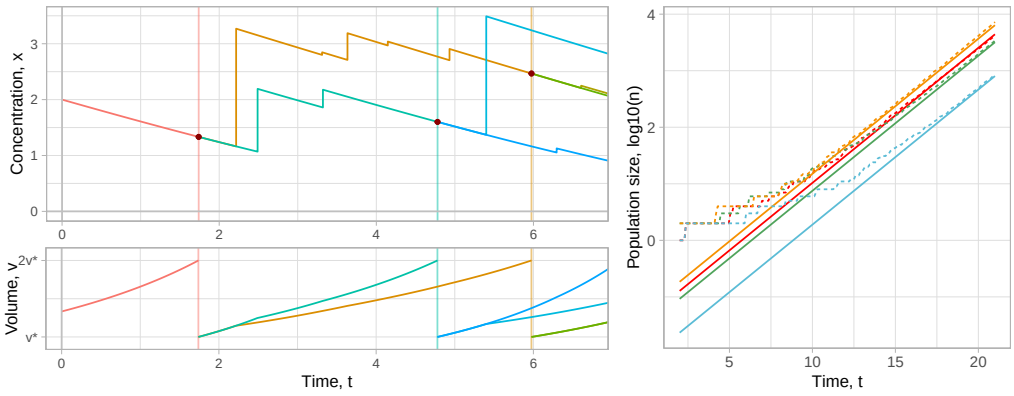


Figure 5.6: *(Left)* the time evolution of a sample cell population, where an individual colour is assigned to each cell. We plot the protein concentration in the cell and its volume, and mark the division time (vertical lines). *(Right)* The time evolution of the log-scaled sample population size in four different simulation runs with the same initial conditions and parameters values $\alpha = 0.9$, $\beta = 0.5$, $k = 2$, and $v^* = 1.5$ (dashed lines); the solid lines correspond to the exponential growth with the rate constant (5.28), shifted by the random factor $\log_{10} W(x_0, v_0)$, whose specific values were estimated for each simulation using linear regression. Each simulation has random fluctuations at the beginning and afterwards converges to the expected exponential growth.

balance equation coincides with the Chapman–Kolmogorov equation (5.38):

$$\begin{aligned}\frac{\partial h(x,v,t)}{\partial t} = \frac{\partial}{\partial x}\left(x\gamma(x)h(x,v,t)\right) - \frac{\partial}{\partial v}\left(v\gamma(x)h(x,v,t)\right) \\ + \frac{\alpha}{\beta}\int_0^x e^{-\frac{x-y}{\beta}} h(y,v,t)\,\mathrm{d}y - \alpha h(x,v,t).\end{aligned} \tag{5.49}$$

The initial and boundary conditions are:

$$h(x,v,0) = h_0(x,v), \tag{5.50a}$$

$$h(x,2v^*,t) = \frac{1}{4}h(x,v^*,t), \tag{5.50b}$$

where $h_0(x,v)$ in (5.50a) is an initial population density. The extra factor of two in the boundary condition (5.50b) compared to the previous boundary condition (5.39b) reflects the cell doubling at the end of a cell cycle in the population scenario.

The population grows in time, and we assume that $h(x,v,t)$ has the following separable form:

$$h(x,v,t) = e^{\lambda t}\frac{q(x)g(v)}{\gamma(x)},$$

which we substitute into (5.49). Doing so yields

$$(vg(v))' + \xi g(v) = 0, \qquad g(2v_0) = \frac{1}{4} g(v_0), \tag{5.51a}$$

$$\left(\frac{\lambda}{\gamma(x)} - \xi\right) q(x) - (xq(x))' + \alpha \frac{\mathrm{d}}{\mathrm{d}x} \int_0^x \frac{q(y)}{\gamma(y)} e^{-\frac{x-y}{\beta}} \,\mathrm{d}y = 0, \tag{5.51b}$$

where the boundary condition for the volume dependence is derived from (5.50b). As we can see, equation (5.51a) is equivalent to (5.42a) except the boundary condition; hence the general solution remains to be (5.43), and the boundary condition yields $\xi = -1$; then the distribution of the cell volume and its mean value are:

$$g(v) = \frac{2v^*}{v^2}, \quad \mathbb{E}(v) = 2v^* \ln(2). \tag{5.52}$$

Finally, we substitute the obtained value of $\xi = -1$ into (5.51b), leading to the integro-differential equation:

$$q(x) - \lambda(1 + kx)q(x) + (xq(x))' - \alpha \frac{\mathrm{d}}{\mathrm{d}x} \int_0^x (1 + ky) q(y) e^{-\frac{x-y}{\beta}} \,\mathrm{d}y = 0,$$

This is the same equation as (5.19), wherein we have already replaced $p(x)$ with $q(x)$ and chose the exponential burst kernel for $B(x)$. The solution is following:

$$p(x) = (1 + kx) \frac{\beta \sigma^2}{\Gamma(\xi)} e^{-\sigma x} (\sigma x)^{\xi - 1}, \tag{5.53}$$

where the constants σ and ξ are given by (5.29).

The obtained distributions (5.52)–(5.53) were verified using the stochastic simulations. The recursive algorithm mimicking population extension, the algorithm for bivariate modelling, and required formulas are provided in Appendix A2.

Conclusions

The main result is the large-time distribution of cell volume (5.44) and protein concentration (5.45) in the single cell framework and in the population framework, (5.52) and (5.53) respectively. Interestingly, the two are independent in the large-time limit (but interdependent transiently in the presence of the dilution feedback). We expect that the large-time independence carries over to more complex models of cell division than the reset rule (5.37). However, additional coupling between protein and cell size (e.g. volume-dependent production, partitioning noise) may introduce a dependence between the two variables [99].

The single-cell stationary distributions (5.44)–(5.45) exist if the product $\alpha\beta$

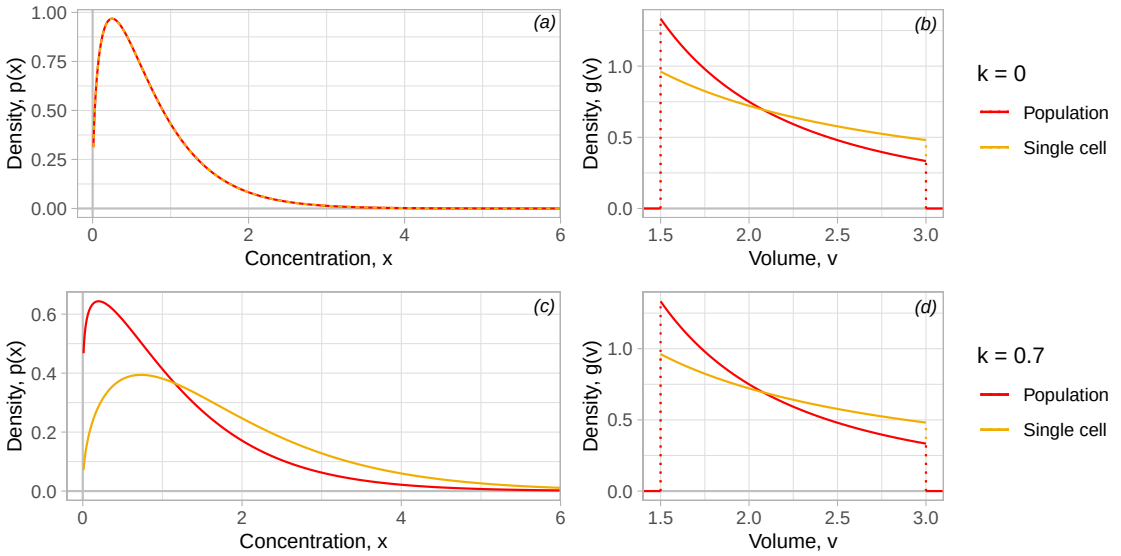


Figure 5.7: The effect of absence (the first row) and presence (the second row) of feedback in dilution on the large-time distributions of protein concentration x and cell volume v. Parameters are as follow: $\alpha = 1.5$, $\beta = 0.5$, $k = 0.6$, $v^* = 1.5$.

of burst frequency and burst size is less than the maximal dilution rate $1/k = \lim_{x\to\infty} x/(1+kx)$. Clearly, in the alternative case ($\alpha\beta > 1/k$), the build up of protein prevents stationarity [100]. In the population scenario, the large-time distribution (5.52)–(5.53) exists if the large-time population growth rate constant λ (5.28) is positive. In the alternative scenario ($(\alpha-1)\beta > 1/k$), the build up of protein overburdens the cells and stalls the population growth. We note that this cannot happen in the low burst frequency (high noise) scenario $\alpha < 1$.

In the absence of dilution feedback, the volume process is periodic and hence does not converge to its stationary distribution (Figure 5.5b). Periodicity also appears for more complex cell division mechanisms than (5.37) as long as (i) the volume grows exponentially and (ii) the cell divides into two equal halves [101]. Feedback in dilution makes the growth protein-dependent and hence non-exponential. As a consequence, we get rid of the periodicity (Figure 5.5a) and obtain ergodicity i.e. convergence of the large-time distribution (Figure 5.6).

Figure 5.7 visualises the effects of dilution feedback and population framework on the protein and volume distribution. Inclusion of feedback tilts the concentration distribution to the right (Figures 5.7a and 5.7c). Inclusion of feedback does not affect the volume distribution (Figures 5.7b and 5.7d). Population volume distribution is tilted to the left compared with the single cell distribution (Figures 5.7b and 5.7d). Without feedback, the concentration distribution is the same gamma distribution in a population and for a single cell (Figure 5.7a). With feedback, the population distribution is tilted to the left compared to the single cell distribution (Figure 5.7c). Consequently, the fraction of cells above a concentration threshold in

the population is smaller than the fraction of time a single cell has concentration above the threshold. This has important consequences for drug-tolerant persisters in microbial and cancer cells that will be rarer than as predicted by classical simulation if the feedback is present [102], [103].

Chapter 6

Negative feedback on dilution

This chapter presents ongoing research and preliminary findings, reflecting the state of our current studies of negative feedback loop on the protein dilution.

The essential core of the single cell model in this case of feedback remains identical to one in Section 5.2. The difference lays in the response function $\gamma(x)$, which remains of the Hill type, but now reflects negative feedback:

$$\gamma(x) = \frac{kx}{1+kx}. \tag{6.1}$$

The Chapman-Kolmogorov equation of the probability density function $p(x,t)$ is identical to (5.2). Then it simplifies to an integral equation:

$$\frac{kx^2}{1+kx}p(x) = \alpha \int_0^x e^{-(x-y)/\beta} p(y)\,\mathrm{d}y\,. \tag{6.2}$$

A preparatory step is similar to one in (5.20), we introduce an auxiliary function $q(x) = \frac{p(x)}{1+kx}$ and its Laplace transform $Q(s) = \mathcal{L}[q(x)](s)$. The substitution of $q(x)$ into (6.2) leads to a following equation:

$$kx^2 q(x) = \alpha \int_0^x e^{-(x-y)/\beta} q(y)(1+ky)\,\mathrm{d}y\,.$$

Subsequently we apply the Laplace transform and obtain a linear second order differential equation:

$$\left(s + \frac{1}{\beta}\right) Q''(s) + \alpha Q'(s) - \frac{\alpha}{k} Q(s) = 0, \tag{6.3}$$

which is can be converted to the (regular) Bessel differential equation [104], then the solution is following:

$$Q(s) = \left(s+\frac{1}{\beta}\right)^{\frac{1-\alpha}{2}} \left(C_1 J_{1-\alpha}\left(2i\sqrt{\frac{\alpha}{k}}\sqrt{s+\frac{1}{\beta}}\right) + C_2 Y_{1-\alpha}\left(2i\sqrt{\frac{\alpha}{k}}\sqrt{s+\frac{1}{\beta}}\right)\right),$$

where $J_{1-\alpha}(\cdot)$ and $Y_{1-\alpha}(\cdot)$ are the Bessel functions of the first kind (2.12) and the second kind (2.13), respectively. To make the range of the solution real, we use their modified versions [65]:

$$Q(s) = \left(s + \frac{1}{\beta}\right)^{\frac{1-\alpha}{2}} \left(\tilde{C}_1 I_{1-\alpha}\left(2\sqrt{\frac{\alpha}{k}}\sqrt{s+\frac{1}{\beta}}\right) + \tilde{C}_2 K_{1-\alpha}\left(2\sqrt{\frac{\alpha}{k}}\sqrt{s+\frac{1}{\beta}}\right)\right),$$

where $I_{1-\alpha}(\cdot)$ and $K_{1-\alpha}(\cdot)$ are the modified Bessel functions of the first and the second kinds (2.15) and (2.16), respectively. Considering that the Laplace transform of must be bounded function, but $I_{1-\alpha}$ is an increasing function on $\mathbb{R}^+$, we set $\tilde{C}_1 = 0$. The used properties of Bessel functions and modified Bessel functions are provided in Section 2.1.

To express the final form of $Q(s)$, we use the following integral form of the modified Bessel function of the second kind (2.17):

$$K_{1-\alpha}(x) = \frac{1}{2}\left(\frac{x}{2}\right)^{1-\alpha} \int_0^\infty \exp\left\{-t - \frac{x^2}{4t}\right\} \frac{1}{t^{2-\alpha}}\,\mathrm{d}t\,,$$

then

$$Q(s) = \left(s + \frac{1}{\beta}\right)^{1-\alpha} \int_0^\infty \exp\left\{-t - \frac{\alpha(s+\beta^{-1})}{kt}\right\} \frac{1}{t^{2-\alpha}}\,\mathrm{d}t\,. \tag{6.4}$$

Finally, substitution $x\left(s+\frac{1}{\beta}\right) = t$ shows that

$$Q(s) = C\int_0^\infty e^{-sx}\exp\left\{-\frac{x}{\beta} - \frac{\alpha}{kx}\right\} x^{\alpha-2}\,\mathrm{d}x\,. \tag{6.5}$$

is by definition the Laplace image, the original function is following:

$$q(x) = Cx^{\alpha-2}\exp\left\{-\frac{x}{\beta} - \frac{\alpha}{kx}\right\}.$$

The steady state probability density function of the protein concentration regulated by negative feedback loop is given by

$$p(x) = C(1+kx)x^{\alpha-2}\exp\left\{-\frac{x}{\beta} - \frac{\alpha}{kx}\right\}, \tag{6.6}$$

where the value of the constant C is chosen so that $p(x)$ becomes a valid density function, i.e., its integral over $\mathbb{R}_0^+$ is equal to one:

$$C = \frac{(\beta\xi)^{1-\alpha}}{2^{2-\alpha}\left(K_{1-\alpha}(\xi) + \sqrt{\alpha\beta k}K_{-\alpha}(\xi)\right)}, \qquad \xi = \sqrt{\frac{4\alpha}{k\beta}}.$$

In Fig. 6.1 (left panel) is shown $p(x)$ for different parameter values. In Ap-

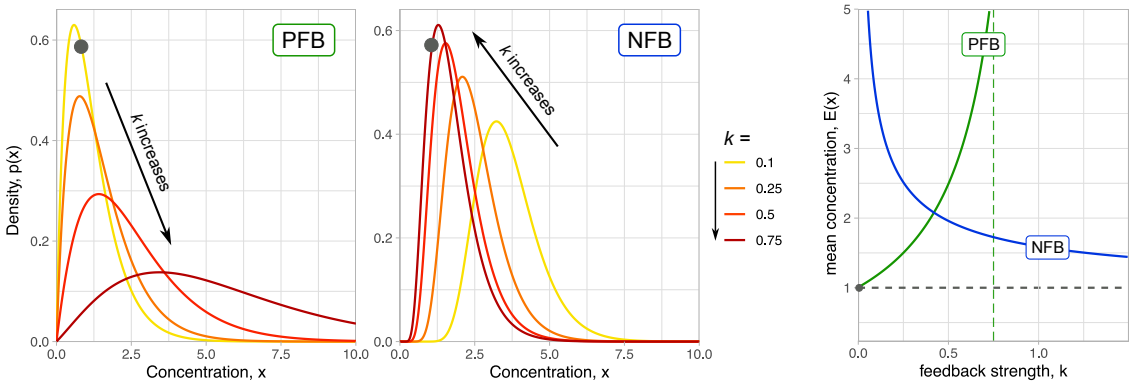


Figure 6.1: *(Left)* Influence of the negative feedback on PDF $p(x)$ compared to positive feedback. By grey dots are marked distributions that are the closest to unregulated case. *(Right)* The mean protein concentration as the function of the feedback strength k. Parameters values are $\alpha = 2,\ \beta = 0.5$.

pendix A3 we use stochastic simulations to prove the obtained distribution (6.6).

The mean protein concentration satisfies:

$$\mathbb{E}(x) = \frac{\beta\xi}{2}\frac{K_{-\alpha}(\xi) + \sqrt{\alpha\beta k}K_{-\alpha-1}(\xi)}{K_{-\alpha+1}(\xi) + \sqrt{\alpha\beta k}K_{-\alpha}(\xi)}.$$

It follows from the form of the pdf (6.6) that the k-th moment can be obtained for any given $k \in \mathbb{N}$. We recall that the mean concentration in PFB (5.11) is equal to unregulated case (2.41) as $k = 0$; as k reaches $\eta = 0^-$ it diverges. As it is shown in Figure 6.1, in NFB, the mean concentration gradually converges to value in unregulated case (2.41) (dashed horizontal line). In addition, a single cell exists for any set of parameter (α, β, k).

Concluding Remarks and Prospects

We conclude this work by summarising our key findings, discussing the limitations of our approach, and introducing potential future research directions. We have developed analytical frameworks for both single-cell analysis and cell populations, providing robust tools to explore the dynamic behaviour of gene expression through various control circuits. Our investigation focused on the regulation of burst frequency, protein dilution, and interactions via auxiliary species.

Specifically, we examined how negative feedback on burst frequency occurs when a protein inhibits the accessibility of its promoter. We obtained the protein distribution for the limiting Hill-type response function and demonstrated that the protein distributions in a single cell and a cell population are identical for any general form of the response function.

The feed-forward type of regulation emerges when mRNA and its antagonist – miRNA – are co-produced from a common coding sequence. We show that in the low-noise regime, it perfectly adapts to disturbances in production parameters and maintains the mean mRNA level constant. If noise is moderate, then this property is partially lost. Yet, the mean concentration is proven to be less volatile than in the case of negative feedback on frequency.

In cases where protein concentration affects cell growth, we observe feedback on dilution. We primarily consider scenarios where an excessive amount of protein imposes a burden on cellular mechanisms and slows down cell expansion. Below, we summarise the results:

- The protein distribution is unaffected by cell division: both randomly triggered and deterministically conditioned cell division mechanisms yielded identical protein distributions.
- In the absence of feedback, both the single cell and population exhibit a protein distribution revert to the gamma distribution of the unregulated gene expression.
- The inclusion of feedback has an opposite effect on PDFs at the single-cell and population levels. In a single cell, the strengthening of feedback means the cell spends more time with a higher protein concentration, so the distribution flattens. In the cell population, the difference is caused by cells with

low protein concentration: they proliferate faster, produce descendants with also low protein concentration, and this effect accumulates. Thus the protein distribution becomes narrower and more right-skewed.

- At both levels, there is a similar condition for the distribution existence, which prevents overproduction of protein.

The negative feedback on dilution occurs when a specific protein is essential for cell growth, and higher concentrations accelerate this growth. Although this is still a work in progress, we demonstrate that it exhibits the expected effects, opposite to those of positive feedback. If the feedback is low, then the protein accumulates within the cell and the mean concentration diverges; sufficiently high feedback, on the other hand, has the same characteristics as unregulated expression.

In reflecting on the limitations of our work, several challenges become apparent, particularly when dealing with more complex regulatory circuits. Firstly, the more realistic regulatory functions may lead to the models both difficult to solve. Additionally, a larger parametric space is harder to analyse, limiting their practical interpretation. Furthermore, our approach is suitable for a low number of species that can be incorporated, which may oversimplify real biological systems, where multiple interactions are the norm. Additionally, we employ a simplification in the production rate by combining the rates for transcription and translation into a single rate for protein synthesis. While this approach helps in a model construction, it potentially overlooks nuances between these stages that could be critical in understanding gene expression dynamics fully. These limitations highlight areas for potential refinement and suggest caution when applying our findings to more complex biological scenarios.

Our near-future work plans and ideas are:

- Proceed with the study of negative feedback from the population perspective. We already know that PBE in this case becomes the Heun equation, which we expect to solve numerically.
- Continue study of the division mechanisms. Possible improvements are incorporation of partition noise (i.e., descendants inherit half of mother protein molecules on average) and stochastic division conditions (e.g. the cell division is randomly triggered, but after reaching a threshold volume).
- Study a more realistic positive feedback function, where a basal level of natural protein degradation is included.
- Optimise simulations and transfer the simulation programs, especially those for cell populations, to more computationally efficient programming languages (e.g., Julia).

Bibliography

[1] M. Smith, K. R. Ghusinga, and A. Singh, "Comparison of feedback strategies for noise suppression in protein level," in *2019 American Control Conference (ACC)*, IEEE, 2019, pp. 1513–1518.

[2] U. Alon, *An introduction to systems biology: design principles of biological circuits*. Chapman and Hall/CRC, 2006.

[3] X. Dai, H. Liu, S. Shen, *et al.*, "Yap activates the hippo pathway in a negative feedback loop," *Cell research*, vol. 25, no. 10, pp. 1175–1178, 2015.

[4] S. Krishna, A. M. Andersson, S. Semsey, and K. Sneppen, "Structure and function of negative feedback loops at the interface of genetic and metabolic networks," *Nucleic acids research*, vol. 34, no. 8, pp. 2455–2462, 2006.

[5] G. Harmange, R. A. R. Hueros, D. L. Schaff, *et al.*, "Disrupting cellular memory to overcome drug resistance," *Nature communications*, vol. 14, no. 1, p. 7130, 2023.

[6] P. Shivakumar and M. Krauthammer, "Structural similarity assessment for drug sensitivity prediction in cancer," in *BMC bioinformatics*, Springer, vol. 10, 2009, pp. 1–7.

[7] A. Torkamani and N. J. Schork, "Background gene expression networks significantly enhance drug response prediction by transcriptional profiling," *The Pharmacogenomics Journal*, 2011. DOI: 10.1038/tpj.2011.35.

[8] P. Thomas, "Intrinsic and extrinsic noise of gene expression in lineage trees," *Scientific reports*, vol. 9, no. 1, p. 474, 2019.

[9] N. Sahu, J.-P. Stéphan, D. D. Cruz, *et al.*, "Functional screening implicates mir-371-3p and peroxiredoxin 6 in reversible tolerance to cancer drugs," *Nature Communications*, 2016. DOI: 10.1038/ncomms12351.

[10] C. Y. Hu, J. D. Varner, and J. B. Lucks, "Generating effective models and parameters for rna genetic circuits," *ACS synthetic biology*, vol. 4, no. 8, pp. 914–926, 2015.

[11] E. Yeung, J. Kim, Y. Yuan, J. Gonçalves, and R. M. Murray, "Data-driven network models for genetic circuits from time-series data with incomplete measurements," *Journal of the Royal Society Interface*, vol. 18, no. 182, p. 20 210 413, 2021.

[12] S. Wu, Z.-P. Liu, X. Qiu, and H. Wu, "Modeling genome-wide dynamic regulatory network in mouse lungs with influenza infection using high-dimensional ordinary differential equations," *PloS one*, vol. 9, no. 5, e95276, 2014.

[13] D. J. Wilkinson, *Stochastic modelling for systems biology*. Chapman and Hall/CRC, 2018.

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

[15] P. Thomas, N. Popović, and R. Grima, "Phenotypic switching in gene regulatory networks," *Proceedings of the National Academy of Sciences*, 2014. DOI: 10.1073/pnas.1400049111.

[16] A. Singh and J. P. Hespanha, "Approximate moment dynamics for chemically reacting systems," *IEEE Transactions on Automatic Control*, vol. 56, no. 2, pp. 414–418, 2010.

[17] A. Borri, P. Palumbo, and A. Singh, "Metabolic noise reduction for enzymatic reactions: The role of a negative feedback," in *2015 54th IEEE Conference on Decision and Control (CDC)*, IEEE, 2015, pp. 2537–2542.

[18] N. Friedman, L. Cai, and X. S. Xie, "Linking stochastic dynamics to population distribution: An analytical framework of gene expression," *Physical review letters*, vol. 97, no. 16, p. 168 302, 2006. DOI: 10.1103/PhysRevLett.97.168302.

[19] J. J. Tyson, K. C. Chen, and B. Novak, "Sniffers, buzzers, toggles and blinkers: Dynamics of regulatory and signaling pathways in the cell," *Current opinion in cell biology*, vol. 15, no. 2, pp. 221–231, 2003.

[20] R. Rudnicki and A. Tomski, "On a stochastic gene expression with pre-mrna, mrna and protein contribution," *Journal of Theoretical Biology*, vol. 387, pp. 54–67, 2015. DOI: https://doi.org/10.1016/j.jtbi.2015.09.012. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0022519315004555.

[21] P. Bokes and A. Singh, "Protein copy number distributions for a self-regulating gene in the presence of decoy binding sites," *PloS one*, vol. 10, no. 3, e0120555, 2015. DOI: 10.1371/journal.pone.0120555.

[22] T. Friedlander and N. Brenner, "Cellular properties and population asymptotics in the population balance equation," *Phys. Rev. Lett.*, vol. 101, p. 018 104, 1 2008. DOI: 10.1103/PhysRevLett.101.018104.

[23] G. Subramanian, D. Ramkrishna, A. Fredrickson, and H. Tsuchiya, "On the mass distribution model for microbial cell populations," *The Bulletin of mathematical biophysics*, vol. 32, pp. 521–537, 1970.

[24] C.-C. Shu, A. Chatterjee, W.-S. Hu, and D. Ramkrishna, "Modeling of gene regulatory processes by population-mediated signaling: New applications of population balances," *Chemical engineering science*, vol. 70, pp. 188–199, 2012.

[25] S. E. Eilbeck Karen Lewis, C. J. Mungall, M. Yandell, L. Stein, R. Durbin, and M. Ashburner, "The sequence ontology: A tool for the unification of genome annotations," version 2.5.3, *Genome Biology*, vol. 6, no. 5, R44, Apr. 2005.

[26] L. Rowen, "Gene structure and organization," in *eLS*. American Cancer Society, 2005.

[27] K. Struhl, "Fundamentally different logic of gene regulation in eukaryotes and prokaryotes," *Cell*, vol. 98, no. 1, pp. 1–4, Jul. 1999.

[28] T. Phillips, "Regulation of transcription and gene expression in eukaryotes," *Nature Education*, vol. 1, no. 1, p. 199, 2008.

[29] A. Köhler and E. Hurt, "Exporting rna from the nucleus to the cytoplasm," *Nature reviews Molecular cell biology*, vol. 8, no. 10, pp. 761–773, 2007.

[30] A. Grönlund, P. Lötstedt, and J. Elf, "Transcription factor binding kinetics constrain noise suppression via negative feedback," *Nature communications*, vol. 4, no. 1, pp. 1–5, 2013.

[31] A. Guarino, D. Fiore, and M. Di Bernardo, "In-silico feedback control of a mimo synthetic toggle switch via pulse-width modulation," in *2019 18th European Control Conference (ECC)*, IEEE, 2019, pp. 680–685.

[32] P. Bokes and A. Singh, "Gene expression noise is affected differentially by feedback in burst frequency and burst size," *Journal of mathematical biology*, vol. 74, no. 6, pp. 1483–1509, 2017.

[33] D. M. Suter, N. Molina, D. Gatfield, K. Schneider, U. Schibler, and F. Naef, "Mammalian genes are transcribed with widely different bursting kinetics," *science*, vol. 332, no. 6028, pp. 472–474, 2011.

[34] D. G. Spiller, C. D. Wood, D. A. Rand, and M. R. White, "Measurement of single-cell dynamics," *Nature*, vol. 465, no. 7299, pp. 736–745, 2010.

[35] A. G. Matera, R. M. Terns, and M. P. Terns, "Non-coding rnas: Lessons from the small nuclear and small nucleolar rnas," *Nature reviews Molecular cell biology*, vol. 8, no. 3, pp. 209–220, 2007.

[36] B. Sollner-Webb and E. B. Mougey, "News from the nucleolus: Rrna gene expression," *Trends in biochemical sciences*, vol. 16, pp. 58–62, 1991.

[37] Z. He, M. Kokkinaki, D. Pant, G. I. Gallicano, and M. Dym, "Small rna molecules in the regulation of spermatogenesis," *Reproduction*, vol. 137, no. 6, pp. 901–911, 2009.

[38] E. Ferro, C. E. Bena, S. Grigolon, and C. Bosia, "Microrna-mediated noise processing in cells: A fight or a game?" *Computational and structural biotechnology journal*, vol. 18, pp. 642–649, 2020.

[39] B. John, A. J. Enright, A. Aravin, *et al.*, "Human microrna targets," *PLoS biology*, vol. 2, no. 11, e363, 2004.

[40] A. Rodriguez, S. Griffiths-Jones, J. L. Ashurst, and A. Bradley, "Identification of mammalian microrna host genes and transcription units," *Genome research*, vol. 14, no. 10a, pp. 1902–1910, 2004.

[41] M. R. Fabian, N. Sonenberg, and W. Filipowicz, "Regulation of mrna translation and stability by micrornas," *Annual review of biochemistry*, vol. 79, pp. 351–379, 2010.

[42] Y. Lee, K. Jeon, J.-T. Lee, S. Kim, and V. N. Kim, "Microrna maturation: Stepwise processing and subcellular localization," *The EMBO journal*, vol. 21, no. 17, pp. 4663–4670, 2002.

[43] V. P. Zhdanov, "Kinetic models of gene expression including non-coding rnas," *Physics Reports*, vol. 500, no. 1, pp. 1–42, 2011.

[44] S.-L. Lin, J. D. Miller, S.-Y. Ying, *et al.*, "Intronic microrna (mirna)," *BioMed Research International*, vol. 2006, 2006.

[45] L. C. G. Hinske, P. A. Galante, W. P. Kuo, and L. Ohno-Machado, "A potential role for intragenic mirnas on their hosts' interactome," *BMC genomics*, vol. 11, no. 1, pp. 1–13, 2010.

[46] C. Bosia, M. Osella, M. E. Baroudi, D. Corà, and M. Caselle, "Gene autoregulation via intronic micrornas and its functions," *BMC systems biology*, vol. 6, no. 1, pp. 1–16, 2012.

[47] G. T. Reeves, "The engineering principles of combining a transcriptional incoherent feedforward loop with negative feedback," *Journal of biological engineering*, vol. 13, no. 1, pp. 1–11, 2019.

[48] J. Tsang, J. Zhu, and A. van Oudenaarden, "Microrna-mediated feedback and feedforward loops are recurrent network motifs in mammals," *Molecular cell*, vol. 26, no. 5, pp. 753–767, 2007.

[49] T. I. Lee, N. J. Rinaldi, F. Robert, *et al.*, "Transcriptional regulatory networks in saccharomyces cerevisiae," *science*, vol. 298, no. 5594, pp. 799–804, 2002.

[50] S. S. Shen-Orr, R. Milo, S. Mangan, and U. Alon, "Network motifs in the transcriptional regulation network of escherichia coli," *Nature genetics*, vol. 31, no. 1, pp. 64–68, 2002.

[51] M. H. Khammash, "Perfect adaptation in biology," *Cell Systems*, vol. 12, no. 6, pp. 509–521, 2021.

[52] N. Barkai and S. Leibler, "Robustness in simple biochemical networks," *Nature*, vol. 387, no. 6636, pp. 913–917, 1997.

[53] C. Marques, W. Guo, P. Pereira, *et al.*, "The triage of damaged proteins: Degradation by the ubiquitin-proteasome pathway or repair by molecular chaperones," *The FASEB journal: official publication of the Federation of American Societies for Experimental Biology*, vol. 20, no. 6, p. 741, 2006.

[54] M. H. Glickman and A. Ciechanover, "The ubiquitin-proteasome proteolytic pathway: Destruction for the sake of construction," *Physiological reviews*, 2002.

[55] B. Levine and G. Kroemer, "Biological functions of autophagy genes: A disease perspective," *Cell*, vol. 176, no. 1, pp. 11–42, 2019.

[56] N. Kavčič, K. Pegan, and B. Turk, "Lysosomes in programmed cell death pathways: From initiators to amplifiers," *Biological Chemistry*, 2016. DOI: 10.1515/hsz-2016-0252.

[57] J. Pei, G. Wang, L. Feng, *et al.*, "Targeting lysosomal degradation pathways: New strategies and techniques for drug discovery," *Journal of Medicinal Chemistry*, 2021. DOI: 10.1021/acs.jmedchem.0c01689.

[58] D. Molenaar, R. Van Berlo, D. De Ridder, and B. Teusink, "Shifts in growth strategies reflect tradeoffs in cellular economics," *Molecular systems biology*, vol. 5, no. 1, p. 323, 2009. DOI: 10.1038/msb.2009.82.

[59] E. Dekel and U. Alon, "Optimality and evolutionary tuning of the expression level of a protein," *Nature*, vol. 436, no. 7050, pp. 588–592, 2005. DOI: 10.1038/nature03842.

[60] O. Patange, C. Schwall, M. Jones, *et al.*, "Escherichia coli can survive stress by noisy growth modulation," *Nature communications*, vol. 9, no. 1, p. 5333, 2018. DOI: 10.1038/s41467-018-07702-z.

[61] S. Vadia and P. A. Levin, "Growth rate and cell size: A re-examination of the growth law," *Current opinion in microbiology*, vol. 24, pp. 96–103, 2015. DOI: 10.1016/j.mib.2015.01.011.

[62] L. Duso and C. Zechner, "Stochastic reaction networks in dynamic compartment populations," *Proceedings of the National Academy of Sciences*, vol. 117, no. 37, pp. 22 674–22 683, 2020. DOI: 10.1073/pnas.2003734117.

[63] A. Genthon, "Analytical cell size distribution: Lineage-population bias and parameter inference," *Journal of the Royal Society Interface*, vol. 19, no. 196, p. 20 220 405, 2022. DOI: 10.1098/rsif.2022.0405.

[64] B. Turpin, E. Y. Bijman, H.-M. Kaltenbach, and J. Stelling, "Population design for synthetic gene circuits," in *Computational Methods in Systems Biology: 19th International Conference, CMSB 2021, Bordeaux, France, September 22–24, 2021, Proceedings 19*, Springer, 2021, pp. 181–197. DOI: 10.1007/978-3-030-85633-5_11.

[65] M. Abramowitz, I. A. Stegun, and R. H. Romer, *Handbook of mathematical functions with formulas, graphs, and mathematical tables*, 1988.

[66] F. Bowman, *Introduction to Bessel functions*. Courier Corporation, 2012.

[67] G. N. Watson, *A treatise on the theory of Bessel functions*. The University Press, 1922, vol. 3, pp. 181–182.

[68] L. Glasser, K. T. Kohl, C. Koutschan, V. H. Moll, and A. Straub, "The integrals in gradshteyn and ryzhik. part 22: Bessel-k functions," *Scientia. Series A. Mathematical Sciences. New Series*, vol. 22, pp. 129–151, 2012.

[69] I. S. Gradshteyn and I. M. Ryzhik, *Table of integrals, series, and products*. Academic press, 2014.

[70] M. Fila, "Integrálne transformácie a špeciálne funkcie. (Slovak) [integral transformations and special functions]," Lectures notes, 2018.

[71] G. A. Korn and T. M. Korn, *Mathematical handbook for scientists and engineers: definitions, theorems, and formulas for reference and review*. Courier Corporation, 2000, pp. 222–231.

[72] M. R. Spiegel, S. Lipschutz, and J. Liu, *Schaum's Outlines: Mathematical Handbook of Formulas and Tables*, 3rd ed. McGraw-Hill, 2009, ch. 33.

[73] L. Debnath, "The double laplace transforms and their properties with applications to functional, integral and partial differential equations," *International Journal of Applied and Computational Mathematics*, vol. 2, no. 2, pp. 223–241, 2016.

[74] R. Dhunde, N. Bhondge, and P. Dhongle, "Some remarks on the properties of double laplace transforms," *International Journal of Applied Physics and Mathematics*, vol. 3, no. 4, pp. 293–295, 2013.

[75] G. A. Coon and D. L. Bernstein, "Some properties of the double laplace transformation," *Transactions of the American Mathematical Society*, vol. 74, no. 1, pp. 135–176, 1953.

[76] Z. Schuss, *Theory and applications of stochastic processes: an analytical approach*. Springer Science & Business Media, 2009, vol. 170.

[77] C. W. Gardiner *et al.*, *Handbook of stochastic methods*. Springer Berlin, 1985, vol. 3.

[78] D. A. Dawson, B. Maisonneuve, J. Spencer, and D. Dawson, *Measure-valued Markov processes*. Springer, 1993.

[79] M. Doumic and M. Hoffmann, "Individual and population approaches for calibrating division rates in population dynamics: Application to the bacterial cell cycle," *arXiv preprint arXiv:2108.13155*, 2021. DOI: 10.48550/arXiv.2108.13155.

[80] D. Ramkrishna, *Population balances: Theory and applications to particulate systems in engineering*. Elsevier, 2000.

[81] F. Sporleder, Z. Borka, J. Solsvik, and H. A. Jakobsen, "On the population balance equation," *Reviews in Chemical Engineering*, vol. 28, no. 2-3, pp. 149–169, 2012.

[82] K. B. Athreya, P. E. Ney, and P. Ney, *Branching processes*. Courier Corporation, 2004.

[83] S. Janson, "Functional limit theorems for multitype branching processes and generalized Pólya urns," *Stochastic Processes and their Applications*, vol. 110, no. 2, pp. 177–245, 2004. DOI: 10.1016/j.spa.2003.12.002.

[84] A.-M. Wazwaz, *Linear and nonlinear integral equations*. Springer, 2011, vol. 639.

[85] I. Zabaikina, P. Bokes, and A. Singh, "Optimal bang–bang feedback for bursty gene expression," in *2020 European Control Conference (ECC)*, IEEE, 2020, pp. 277–282.

[86] I. Zabaikina, P. Bokes, and A. Singh, "Joint distribution of protein concentration and cell volume coupled by feedback in dilution," in *Computational Methods in Systems Biology*, J. Pang and J. Niehren, Eds., Springer Nature Switzerland, 2023.

[87] N. G. Van Kampen, *Stochastic processes in physics and chemistry*, 3rd ed. Elsevier, 1992, vol. 1.

[88] P. Bokes and A. Singh, "Noise induced bimodality in genetic circuits with monostable positive feedback," *bioRxiv*, p. 464 297, Jan. 2018.

[89] U. Alon, *An introduction to systems biology: design principles of biological circuits*. Chapman and Hall/CRC, 2006.

[90] J. N. Weiss, "The hill equation revisited: Uses and misuses," *The FASEB Journal*, vol. 11, no. 11, pp. 835–841, 1997.

[91] A. I. Khuri, "Applications of dirac's delta function in statistics," *International Journal of Mathematical Education in Science and Technology*, vol. 35, no. 2, pp. 185–195, 2004.

[92] R. B. Schinazi, *From calculus to analysis*. Springer Science & Business Media, 2011.

[93] A. Re, D. Corá, D. Taverna, and M. Caselle, "Genome-wide survey of microrna–transcription factor feed-forward regulatory circuits in human," *Molecular BioSystems*, vol. 5, no. 8, pp. 854–867, 2009.

[94] S. Dey, S. Kannoly, P. Bokes, J. J. Dennehy, and A. Singh, "Feedforward genetic circuits regulate the precision of event timing," in *2021 European Control Conference (ECC)*, IEEE, 2021, pp. 2127–2132.

[95] C. Jia, A. Singh, and R. Grima, "Concentration fluctuations in growing and dividing cells: Insights into the emergence of concentration homeostasis," *PLoS Computational Biology*, vol. 18, no. 10, e1010574, 2022. DOI: 10.1371/journal.pcbi.1010574.

[96] D. R. Schoenberg and L. E. Maquat, "Regulation of cytoplasmic mrna decay," *Nature Reviews Genetics*, vol. 13, no. 4, pp. 246–259, 2012.

[97] I. Zabaikina, Z. Zhang, C. Nieto, P. Bokes, and A. Singh, "Amplification of noisy gene expression by protein burden: An analytical approach," in *2023 American Control Conference (ACC)*, IEEE, 2023, pp. 2861–2866.

[98] G. Facchetti, F. Chang, and M. Howard, "Controlling cell size through sizer mechanisms," *Current Opinion in Systems Biology*, vol. 5, pp. 86–92, 2017. DOI: 10.1016/j.coisb.2017.08.010.

[99] C. Nieto-Acuña, J. C. Arias-Castro, C. Vargas-García, C. Sánchez, and J. M. Pedraza, "Correlation between protein concentration and bacterial cell size can reveal mechanisms of gene expression," *Physical Biology*, vol. 17, no. 4, p. 045 002, 2020. DOI: 10.1088/1478-3975/ab891c.

[100] I. Zabaikina, Z. Zhang, C. Nieto, P. Bokes, and A. Singh, "Quantifying noise modulation from coupling of stochastic expression to cellular growth: An analytical approach," *bioRxiv*, pp. 2022–10, 2022. DOI: 10.1101/2022.10.03.510723.

[101] P. Bokes and A. Singh, "Cell volume distributions in exponentially growing populations," in *Computational Methods in Systems Biology: 17th International Conference, CMSB 2019, Trieste, Italy, September 18–20, 2019, Proceedings 17*, Springer, 2019, pp. 140–154. DOI: 10.1007/978-3-030-31304-3_8.

[102] E. Rotem, A. Loinger, I. Ronin, *et al.*, "Regulation of phenotypic variability by a threshold-based mechanism underlies bacterial persistence," *Proceedings of the National Academy of Sciences*, vol. 107, no. 28, pp. 12 541–12 546, 2010. DOI: 10.1073/pnas.1004333107.

[103] S. M. Shaffer, M. C. Dunagin, S. R. Torborg, *et al.*, "Rare cell variability and drug-induced reprogramming as a mode of cancer drug resistance," *Nature*, vol. 546, no. 7658, pp. 431–435, 2017. DOI: 10.1038/nature22794.

[104] A. Polyanin and V. Zaitsev, *Handbook of Exact Solutions for Ordinary Differential Equations*, Second Edition. Chapman & Hall, 2002, pp. 519–531.

[105] M. Geller and E. W. Ng, "A table of integrals of the exponential integral," *Journal of Research of the National Bureau of Standards*, vol. 73, no. 3, pp. 191–210, 1969.

[106] S. M. Ross, *Simulation*. Academic Press, 2013, ch. 5.

[107] S. M. Ross, in *Introduction to Probability and Statistics for Engineers and Scientists (Sixth Edition)*, S. M. Ross, Ed., Sixth Edition, Academic Press, 2021, ch. Simulation, bootstrap statistical methods, and permutation tests, pp. 619–647.

[108] R. M. Corless, G. H. Gonnet, D. E. Hare, D. J. Jeffrey, and D. E. Knuth, "On the lambert w function," *Advances in Computational mathematics*, vol. 5, pp. 329–359, 1996. DOI: doi.org/10.1007/BF02124750.

Appendix A

Simulation algorithms

A1. Feed-forward loop

This section is aimed at verification of theoretical results obtained in Section 4.3. This is done by constructing Algorithm 1, which simulates the process and calculates the first two moments of mRNA and miRNA concentrations.

Referring to Section 4, a trajectory of X concentration in time interval $[0, T]$ has a finite number of discontinuity points, which are due to the random producing process $B(t)$. Between bursts X decays deterministically as per dynamical system (4.1).

Let us consider the process step by step. We define Δt_i as a waiting time after the i-th burst b_i until the next one; thereby the time of the next burst occurrence is $t_{i+1} = t_i + \Delta t_i$. The quantity Δt_i is drawn from the exponential distribution with mean $1/\lambda$; moreover, the size of the burst itself is also drawn from the exponential distribution with mean β. Then, according to the theorem of average value, the mean value of the piecewise continuous trajectory $x(t)$ is given by

$$M_1(X) = \frac{1}{T}\sum_{i=0}^{N} \bar{X}_i, \quad \bar{X}_i = \int_{t_i}^{t_{i+1}} x(t; b_i)\,\mathrm{d}t\,, \tag{A.1}$$

where N is number of observed bursts, T is the total time of observation and equal to $\sum_{i=0}^{N} \Delta t_i$. Also, despite that above we mainly appeal to X, the same is applied to Y.

The functions of deterministic decay $x(t)$ and $y(t)$ are obtained by solving the dynamical system (4.1). First, we obtain the solution of $y(t)$ from the second equation (since it is separated and does not involve $x(t)$), next we substitute it into the equation of $x(t)$. Due to discontinuities caused by bursts, we denote by $x_i(t)$ and

Algorithm 1 Simulation of the mRNA-miRNA interaction

Require: $\lambda, \beta, \delta; x_0, y_0; N$
Ensure: $M_1(X), M_2(X)$

Draw U from the standard uniform distribution
$b_0 \leftarrow 0$ ▷ no burst at the $t = 0$
$\Delta t_0 \leftarrow -\ln U/\lambda$
Draw vectors U_1, U_2 of N independent random values from the standard uniform distribution
for $i \leftarrow 1, N$ **do**
    $b_i \leftarrow -\beta \ln U_{1,i}$
    $\Delta t_i \leftarrow -\frac{1}{\lambda} \ln U_{2,i}$
end for
for $i \leftarrow 1, N$ **do**
    $y_i \leftarrow b_i + y_{i-1} e^{-\Delta t_{i-1}}$
    $x_i \leftarrow b_i + x_{i-1} \exp\{-\delta y_{i-1}(1 - e^{-\Delta t_{i-1}})\}$
    $\bar{X}_i, \bar{Y}_i \leftarrow$ as per (A.4) and (A.5)
    $\bar{X}_i^2, \bar{Y}_i^2 \leftarrow$ as per (A.6)
end for
$M_1(X) \leftarrow \sum \bar{X}_i / \sum \Delta t_i$
$M_2(X) \leftarrow \sum \bar{X}_i^2 / \sum \Delta t_i$

$y_i(t)$ the separate solutions on the each interval $[t_i, t_{i+1})$, which are the following:

$$x_i(t) = \hat{x}_i \exp\left\{-\delta \hat{y}_i (1 - e^{-(t-t_i)})\right\}, \qquad \hat{x}_i = x_{i-1}(t_i) + b_i, \tag{A.2}$$

$$y_i(t) = \hat{y}_i \, e^{-(t-t_i)}, \qquad \hat{y}_i = y_{i-1}(t_i) + b_i, \tag{A.3}$$

where $\hat{x}_i$ and $\hat{y}_i$ are initial conditions for the solution on each interval Δt_i; clearly, $\hat{x}_0$ and $\hat{y}_0$ are the pre-set concentrations at the beginning of observation at time $t = 0$.

Firstly, let us find $\bar{X}_i$ terms by integrating (A.2) on Δt_i:

$$\bar{X}_i = \hat{x}_i e^{-\delta \hat{y}_i} \int_{t_i}^{t_{i+1}} e^{ae^{-t}} \, \mathrm{d}t \,, \quad a = \delta \hat{y}_i e^{t_i}.$$

After performing a substitution $u(t) = ae^{-t}$, the integral takes form of special function $\mathrm{Ei}(x)$ [105]. Its evaluating on the interval $(u(t_i); u(t_{i+1}))$ gives the terms $\bar{X}_i$ in (A.1):

$$\bar{X}_i = \hat{x}_i \exp\{-\delta \hat{y}_i\}\big(\mathrm{Ei}(\delta \hat{y}_i) - \mathrm{Ei}(\delta \hat{y}_i e^{-\Delta t_i})\big). \tag{A.4}$$

Subsequently, integrating (A.3) on each interval Δt_i yields the terms $\bar{Y}_i$ of partial sums for $M_1(Y)$:

$$\bar{Y}_i = \hat{y}_i \, (1 - e^{-\Delta t_i}). \tag{A.5}$$

Analogically, we define the second moments of $x(t)$ and $y(t)$ by

$$M_2(X) = \frac{1}{T}\sum_{i=0}^{N} \bar{X}_i^2, \quad \bar{X}_i^2 = \int_{t_i}^{t_{i+1}} (x(t;b_i))^2 \,\mathrm{d}t\,.$$

After repeating the same approach, we obtain the terms $\bar{X}_i^2$ and $\bar{Y}_i^2$ in form

$$\begin{aligned} \bar{X}_i^2 &= \hat{x}_i^2 \exp\{-2\delta\hat{y}_i\}\big(\mathrm{Ei}(2\delta\hat{y}_i) - \mathrm{Ei}(2\delta\hat{y}_i e^{-\Delta t_i})\big), \\ \bar{Y}_i^2 &= \frac{\hat{y}_i^2}{2}\,(1 - e^{-2\Delta t_i}). \end{aligned} \tag{A.6}$$

Given that we know the piecewise deterministic behaviour of the trajectories, the last missing ingredient is to find an appropriate method for generation random values of Δt_i and b_i. Since they both are drawn from the exponential distribution, we choose the inverse transform method, which is simple, precise, and computationally efficient [106].

Finally, the first and the second moments of X, obtained by executing the simulation algorithm, converge to values of corresponding analytical expressions (4.25) and (4.27). As shown in Figure 4.3, a satisfactory number of bursts to observe is $N \approx 10^5$, after which the computational time significantly increases without a notable gain in the accuracy of result.

A2. Positive feedback on dilution. Single cell and population

Let us start with the univariate model, where we simulate the time evolution of the protein concentration in a single cell, in which the cell volume is not taken into account. The corresponding mathematical model is described in Section 5.1. Since the burst frequency and the mean size are from exponential distributions, we generate them using the inverse transform technique [107]. It is done by sampling u_α and u_β from the standard uniform distribution $U(0,1)$ and using them as a value of the corresponding exponential CDF. Then the random burst size b and the waiting time until next burst Δt_α are following:

$$b = -\beta \ln u_\beta, \tag{A.7a}$$

$$\Delta t_\alpha = -\frac{1}{\alpha} \ln u_\alpha. \tag{A.7b}$$

During a period of time after the burst at time t_0 and until the next one, the protein concentration decays from the level x_0 as per (5.1), the solution of which is

$$x(t) = \frac{1}{k} W_0\left(k x_0 e^{k x_0} e^{t_0 - t}\right), \tag{A.8}$$

Algorithm 2 Simulation of the cell population

global $\mathbf{c} = (\alpha, \beta, k)$
require $X = x_0$, $T = 0$, T_{stop}, ***population***
    ▷ T_{stop} is a simulation endpoint, ***population*** is an empty dataframe
function TRAJECTORY($\mathbf{c}, x_0, t_0, t$)
    return intermediate X values during deterministic decay as per (A.8)
end function
function SCELL(***population***, X, T)
    create dataframe ***cell***
    while $T < T_{stop}$ **do**
        Generate u_α, u_e
        Compute Δt_α, Δt_e ▷ as per (A.7b) and (A.10)
        if $\min(\Delta t_\alpha, \Delta t_e) = \Delta t_\alpha$ **then**
            bind TRAJECTORY($X,\ T,\ T + \Delta t_\alpha$) to ***cell***
            $X \leftarrow x(t_0 = T,\ t = T + \Delta t_\alpha)$ ▷ as per (A.8)
            $X \leftarrow X + b$ ▷ as per (A.7a)
            $T \leftarrow T + \Delta t_\alpha$
        else
            bind TRAJECTORY($X,\ T,\ T + \Delta t_e$) to ***cell***
            $X \leftarrow x(t_0 = T,\ t = T + \Delta t_e)$
            $T \leftarrow T + \Delta t_e$
            bind ***cell*** to ***population***
            population $\leftarrow$ SCELL(***population***, $X,\ T$)
            population $\leftarrow$ SCELL(***population***, $X,\ T$) **break**
        end if
    end while
    end while
    return ***population***
end function

where $W_0(\cdot)$ is the principal branch a special function called the Lambert W function [108], which is by definition the function satisfying $W(z)e^{W(z)} = z$ as $z > -1/e$.

We proceed to a simulation of the cell population, in which each individual performs independently yet identically to the single cell case above. Additional complexity is brought by the division mechanism described in Section 2.5.

Firstly, we address an issue with defining the remaining existence time of the cell until its division Δt_e. As follows from (2.30), the division hazard rate $\gamma(x(t))$ is time-dependent, which means that each event at time t_0 (either birth or burst) affects the probability of the division during the period $(t_0, t_0 + \Delta t)$ as follows:

$$Prob[t_e < t_0 + \Delta t_e | t_e > t_0] = 1 - \exp\left\{ - \int_{t_0}^{t_0 + \Delta t_e} \gamma(x(t_e))\, \mathrm{d}t_e \right\}, \tag{A.9}$$

where $x(t)$ is given by (A.8). To sample from this non-homogeneous exponential distribution, we again use inverse transform technique: we generate a value u_e from $U(0, 1)$ and use it as a value of the probability (A.9). After performing integration,

Algorithm 3 Simulation of the bivariate single-cell model

global c $= (\alpha, \beta, k, v^*)$
require $X = x_0$, $V = v_0$, $T = 0$, T_{stop}, ***cell***
function TRAJECTORY(v_0, x_0, t_0, t)
    return intermediate values of X and V on the interval (t_0, t) as per (A.11)
end function
while $T < T_{stop}$ **do**
    Compute $\Delta t_\alpha(u_\alpha)$, Δt_e ▷ as per (A.7b) and (A.12)
    if $\min(\Delta t_\alpha, \Delta t_e) = \Delta t_\alpha$ **then**
        bind TRAJECTORY$(V, X, T, T + \Delta t_\alpha)$ to ***cell***
        $V \leftarrow v(t_0 = T, t = T + \Delta t_\alpha)$
        $X \leftarrow x(t_0 = T, t = T + \Delta t_\alpha) + b$
        $T \leftarrow T + \Delta t_\alpha$
    else
        bind TRAJECTORY$(V, X, T, T + \Delta t_e)$ to ***cell***
        $V \leftarrow v^*$
        $X \leftarrow x(t_0 = T, t = T + \Delta t_\alpha)$
        $T \leftarrow T + \Delta t_e$
    end if
end while
end while
return ***cell***

we obtain:

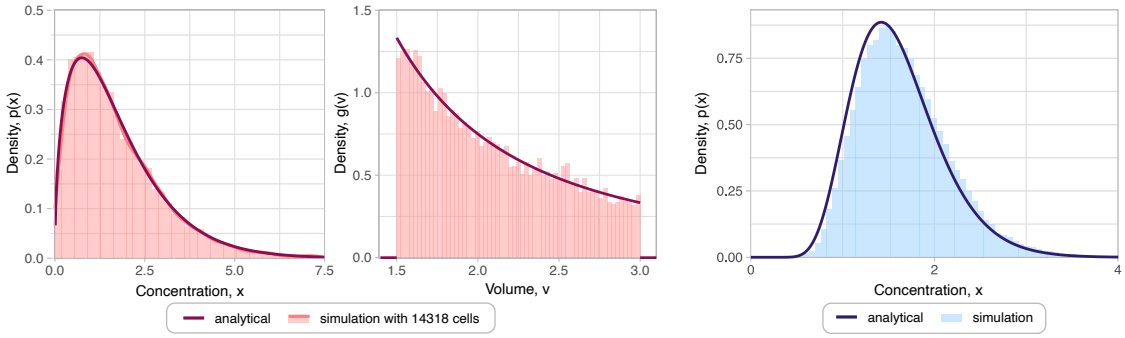
$$u_e = \frac{1}{kx_0} W(kx_0 e^{kx_0} e^{-\Delta t_e}),$$

which leads to the explicit expression of the waiting time until division:

$$\Delta t_e = kx_0(1 - u_e) - \ln u_e. \tag{A.10}$$

Secondly, to imitate the population growth, we develop a recursive algorithm. Its core function generates the time evolution of the protein concentration in a single cell from birth until death. Specifically, after the birth and every next burst (if it occurs), we generate candidate waiting periods for the division and the burst (Δt_α and Δt_e). The minimal value in this pair defines which event comes next. If $\min(\Delta t_\alpha, \Delta t_e)$ is Δt_α, then the burst occurs next, and the whole cycle repeats; otherwise, the cell divides, and the core function calls itself twice, which corresponds to the mother cell division into two daughter cells (see details in Algorithm 2).

We proceed to the bivariate model of a single cell described in Section 5.3. Its simulation requires an explicit time-dependent function of the cell volume. We find it using (5.36), integration of which leads to the fact that the cell volume is inversely dependent on the protein concentration, i.e., $v(t) = C/x(t)$. Given that at the time



(a) Population with PFB

(a) Single cell with NFB

Figure A.1: (a) Comparison of the analytical distributions (5.52)–(5.53) to the results of a large-time kinetic Monte Carlo simulation with parameters $\alpha = 2$, $\beta = 0.45$, $k = 0.7$, $v^* = 1.5$, $T = 20$, and initial conditions $x_0 = v_0 = 2$. (b) Comparison of the analytical distribution (6.6) and one obtained using simulations with parameters $\alpha = 5, \beta = 0.2, k = 1$, a simulation endpoint $T = 22.5$, an initial condition $x_0 = 1$.

of the last event is t_0, the concentration is x_0 in the cell of volume v_0, we obtain:

$$v(t) = \frac{k x_0 \, v_0}{W_0 \left(k x_0 e^{k x_0} e^{t_0 - t} \right)}. \tag{A.11}$$

Since the division occurs at the moment, when the cell reaches the critical volume $2v^*$, then the waiting time until the division is:

$$\Delta t_e = k x_0 \left(1 - \frac{v_0}{2v^*} \right) - \ln \frac{v_0}{2v^*}, \tag{A.12}$$

which is now the deterministic value in comparison with a univariate case (A.10), where it is randomly drawn. The whole simulation approach is given in Algorithm 3.

The simulation algorithm of the bivariate population is essentially the same as Algorithm 2. Yet it requires two changes: the first one is generating Δt_e using (A.12) instead of (A.10); the second one is storing volume trajectories V, which is done by using lines 3-4, 9, 14 from Algorithm 3. We use this algorithm to prove our results for stationary the protein and cell volume distributions (5.52)–(5.53). As it is shown in Fig. A.1a, the results obtained from simulation are consistent with the analytical formula.

A3. Negative feedback on dilution. Single cell and population

In the previous section, we constructed step by step the simulation algorithm for the model with positive feedback on dilution. Modeling the negative feedback instead of positive one does not alter the behavior and principal assumptions of the model. Thus, in this chapter, we only derive functions of the protein trajectory $x(t)$ and

the waiting time until division Δt_e for Algorithm 2 and Algorithm 3, so that they simulate dynamics of the negative feedback loop.

The deterministic decay of the protein is governed by ODE:

$$\dot{x} = x\gamma(x),$$

where $\gamma(x)$ is given by (6.1); the solution follows:

$$x(t) = \left(kW_0 \left(\frac{1}{kx_0} e^{1/kx_0} e^{-\Delta t_e} \right) \right)^{-1}, \quad x_0 = x(t_0), \tag{A.13}$$

where t_0 is the time of the last burst (for the single cell) or the last event (for the population); $W_0(\cdot)$ is the principal branch of the Lambert W function. Here, the assumptions about burst size and frequency are identical to negative feedback model. Thus, the burst size and the waiting time until the next burst are generated using inverse transform sampling method (A.7a)–(A.7b).

The last thing we address is the waiting time until the next division, which is required for the population simulation. After evaluating the integral (A.9), we obtain:

$$u_e = \frac{1}{kx_0 W_0 \left(\frac{1}{kx_0} e^{1/kx_0} e^{-\Delta t_e} \right)},$$

from which follows:

$$\Delta t_e = \frac{1 - u_e}{kx_0 u_e} - \ln u_e. \tag{A.14}$$

To support the result of the analytical approach (6.6), we use (A.13) instead of (A.8) and ignore the volume component in Algorithm (3). Figure A.1b shows that the resulting empirical density is in agreement with the analytical one.

Appendix B

Moments analysis of the effects of the feedback strength

In the main text, we focus on the steady-state distribution of the protein in single-cell and population frameworks. In this appendix, we study the statistics (mean concentration, protein noise, and skewness) as functions of feedback strength k, with parameters α and β. In this appendix we also use subscripts to differentiate the single-cell (SC) and population (Pop) statistics.

In the single cell, the protein distribution $p_{SC}(x)$ (5.9) has the existence condition $\eta > 0$, from which it follows that k must be within the interval $K_{SC} = [0, 1/\alpha\beta)$, for a given set of parameters (α, β, γ). We use (5.11) to derive the statistics of interest:

$$\mathbb{E}_{SC}(x) = \frac{\alpha}{\eta}(1 + k\beta), \tag{B.1}$$

$$(\mathrm{CV}_x^2)_{SC} = \frac{1}{\alpha}\left(1 - \frac{1+\alpha}{\left(1+\frac{1}{k\beta}\right)^2}\right), \tag{B.2}$$

$$\mathrm{Skew}_{SC}(x) = 2\frac{(\alpha + 1 - (\eta\beta)^3)}{(\alpha + 1 - (\eta\beta)^2)^{3/2}}. \tag{B.3}$$

It is clear that as k approaches $1/\alpha\beta$, in the expression (B.1), the mean value $\mathbb{E}_{SC}(x)$ diverges due to a singularity in the denominator (Fig. B.1B, green line). To explore the behavior of the noise level (B.2) and skewness (B.3) on K_{SC}, we use the first derivative test. For the noise level, we find that $\partial(\mathrm{CV}_x^2)_{SC}/\partial k < 0$ meaning that $(\mathrm{CV}_x^2)_{SC}$ is decreasing function of k within K_{SC} (Fig. B.1D, green line). We also use this approach for the skewness; its derivative is given by:

$$\frac{\partial(\mathrm{Skew}_x)_{SC}}{\partial k} = \frac{3\alpha\beta^2\eta(\beta\eta - 1)}{(1 + \alpha - (\beta\eta)^2)^{5/2}}, \tag{B.4}$$

where the quadratic function in denominator is always positive (it is concave and its two zeroes are not in K_{SC}), then $\partial(\mathrm{Skew}_x)_{SC}/\partial k < 0$ for any $k \in K_{SC}$. We obtain

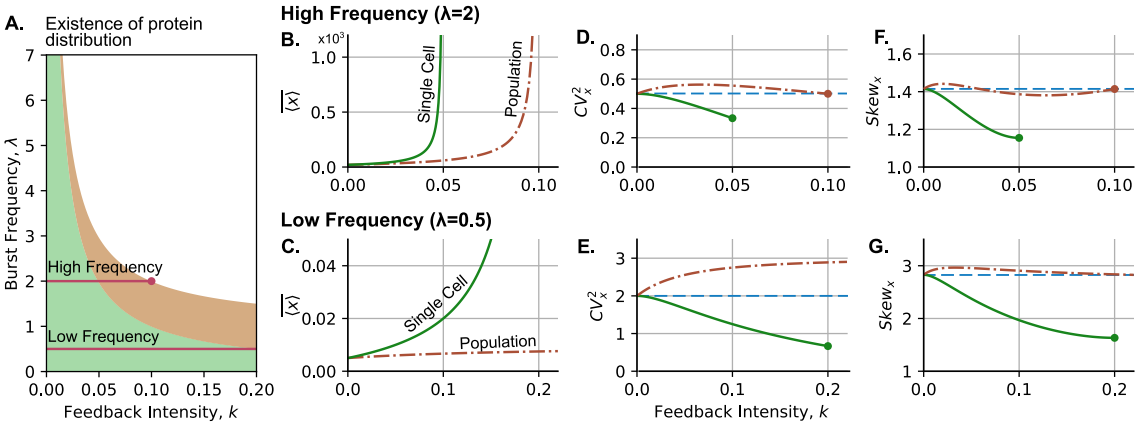


Figure B.1: Cell population is always noisier and more right-skewed than a single cell, despite feedback intensity. (**A.**) Phase diagram of distribution existence. Red line represents increase of k keeping α=2 and the low frequency α=0.5. (**B.**) Mean protein for high frequency and (**C.**) for low frequency (**D.**). Protein noise for high frequency and (**E.**) for low frequency. (**F.**) Protein Asymmetry for high frequency and (**G.**) for low frequency. On (**B.**)–(**G.**) are shown statistics of single cell (green solid line) and population (brown dash-dotted line) compared to unregulated case (blue horizontal dashed lines). Parameters: $\beta = 10$, $\gamma = 1$.

that both $(\mathrm{CV}_x^2)_{SC}$ and $(\mathrm{Skew}_x)_{SC}$ are monotonically decreasing functions with local maxima and minima are left and right endpoints of the interval K_{SC}, respectively. In particular:

$$\lim_{k\to 1/\langle x\rangle}(\mathrm{CV}_x^2)_{SC} = \frac{1}{\alpha+1},$$
$$\lim_{k\to 1/\langle x\rangle}(\mathrm{Skew}_x)_{SC} = \frac{2}{\sqrt{\alpha+1}}.$$

In conclusion, for a given production flow $\alpha\beta$, the feedback of any strength reduces protein noise at the single cell level (green lines in Figs. B.1D–B.1E) and makes the distribution less skewed (green lines in Figs. B.1F–B.1G) compared to the unregulated expression (blue horizontal lines in corresponding figures).

We perform similar approach for the statistics of the population framework:

$$\mathbb{E}_{Pop}(x) = \alpha\beta\frac{1+k\beta}{1+k\beta-\alpha\beta k}, \tag{B.5}$$
$$\left(\mathrm{CV}_x^2\right)_{Pop} = \frac{1}{\alpha}\left(1+\tilde{k}-\alpha\tilde{k}^2\right), \tag{B.6}$$
$$\mathrm{Skew}_{Pop}(x) = 2\sqrt{\frac{1}{\alpha}}\frac{1+2\tilde{k}-3\alpha\tilde{k}^2+\alpha^2\tilde{k}^3}{\left(1+\tilde{k}-\alpha\tilde{k}^2\right)^{3/2}}, \tag{B.7}$$

where $\tilde{k} = k\beta/(1+k\beta)$ is an auxiliary constant. Now, the permissible interval of k is $K_{Pop} = \big[0, 1/\beta(\alpha-1)\big)$, for $\alpha > 1$ (the high frequency limit), and $K_{Pop} = [0,\infty)$,

for $\alpha < 1$ (the low frequency limit). The behavior of mean, noise level, and skewness differs in these two cases. In the low-frequency mode ($\alpha < 1$), we obtain:

$$\lim_{k\to\infty} \mathbb{E}_{Pop}(x) = \frac{\overline{\langle x \rangle}}{1-\alpha}, \tag{B.8}$$

$$\lim_{k\to\infty} (\mathrm{CV}_x^2)_{Pop} = \frac{\beta}{\overline{\langle x \rangle}}(2-\alpha), \tag{B.9}$$

$$\lim_{k\to\infty} \mathrm{Skew}_{Pop}(x) = 2\sqrt{\frac{1}{\alpha}}\frac{\alpha^2 - 3\alpha + 3}{(2-\alpha)^{3/2}}. \tag{B.10}$$

Then the protein distribution in the population with the low transcriptional frequency ($\alpha < 1$) has higher, but always bounded statistics compared to unregulated case. These behaviour is shown in the second row of Fig. B.1.

In high frequency mode ($\alpha > 1$) as k reaches the right endpoint of K_{Pop}, the mean $\mathbb{E}_{Pop}(x)$ diverges, but noise and skewness become identical to the unregulated case:

$$\lim_{k\to 1/\beta(\alpha-1))} (\mathrm{CV}_x^2)_{Pop} = \mathrm{CV}_x^2, \qquad \lim_{k\to 1/\beta(\alpha-1)} \mathrm{Skew}_{Pop}(x) = \mathrm{Skew}(x) \tag{B.11}$$

which is shown in the first row of Fig. B.1.

The first derivative of the squared coefficient of variation,

$$\frac{\partial(\mathrm{CV}_x^2)_{Pop}}{\partial k} = \frac{\beta}{\overline{\langle x \rangle}}\frac{(1-2\alpha)k\beta + 1}{1+k\beta},$$

has a single root at $1/(2\alpha-1)\beta$, indicating that over the interval K_{Pop}, $(\mathrm{CV}_x^2)_{Pop}$ is monotonically increasing if $2\alpha < 1$; otherwise, it is concave (Figs. B.1D–B.1E). The first derivative test for skewness involves analysis of cubic equation, which was done numerically. We conclude that the low frequency leads to $\mathrm{Skew}_{Pop}(x) > \mathrm{Skew}(x)$ on the whole interval K_{Pop}, maximum is reached within K_{SC} (Fig. B.1G). The high frequency leads to minor fluctuations of $\mathrm{Skew}_{Pop}(x)$ around $\mathrm{Skew}(x)$ with single intersection within K_{Pop} (Fig. B.1F).

Overall, we use the statistics of the unregulated case as a critical points for comparison both perspectives. We conclude that for given production rate $\alpha\beta$ and admissible values of k protein distribution in the population is always noisier and more right skewed compared to the single-cell one.