THÈSE DE DOCTORAT
DE L'UNIVERSITÉ DE LILLE I

Spécialité :

Laser, Molécules et Rayonnement atmosphérique

présentée par

Jingkui WANG

Pour obtenir le titre de

Docteur de L'UNIVERSITÉ DE LILLE I

Sujet de la thèse

Dynamical Effects of Delay, Fluctuation and Transcriptional Pausing on Genetic Networks

Soutenue le 20 Septembre 2012
devant le jury composé de :

Thomas ERNEUX	**Rapporteur**
Andrea PARMEGGIANI	**Rapporteur**
Marc LEFRANC	**Directeur de thèse**
Quentin THOMMEN	**Co-Encadrant de thèse**
Ralf BLOSSEY	**Examinateur**
M. Carmen ROMANO	**Examinatrice**

Acknowledgements

This thesis you are about to read is not the work of myself alone. Many people have contributed to it by their help, support, guidance and efforts. I would like to thank all those people.

First and foremost, I would like to gratefully and sincerely thank my supervisor Prof. Marc Lefranc for his support and patience. During my PhD study, he taught me to be a good scientist by sharing his career experiences, giving me intellectual freedom in my work, supporting my attendance at different conferences, encouraging me to develop new ideas and demanding a high quality of work. In particular, he made lot of efforts to find time to correct this manuscript carefully towards the end of my study despite his busy schedule.

I would also like to express my sincere gratitude to my co-supervisor Dr. Quentin Thommen for his enthusiasm, patient guidance, encouragement and generous help. I am always amazed by his quick responses and efficiency for solving the scientific problems. I gained a lot from his vast and diverse knowledge in theoretical physics and biology and also in programming which I was not good at. Especially, he made lots of efforts to valid our results and to write the paper to be published for the subject of chapter 3 in this thesis.

I am most grateful to my colleague Dr. Benjamin Pfeuty who joined our group during the first year of my PhD. I deeply appreciate his constructive advice, frequent discussions we had and his friendship, which helped me and encouraged me many times. To work with him was a real pleasure to me.

I am also sincerely grateful to our collaborator in University of Aberdeen, Dr. Mamen Romano for accepting my visit and for our precious discussion which stimulated the subject of chapter 4 in this thesis.

In addition, I would like to thank my colleague Dr. Constant Vandermoere and Dr. Pierre-Emmanuel Morant who helped me in programming and personal life.

I would also like to express my deep appreciation to all thesis committee members: Prof. Thomas Erneux, Dr. Andrea Parmeggiani, Dr. Ralf Blossey and Dr. Mamen Romano. I am especially thankful to Prof. Thomas Erneux and Dr. Andrea Parmeggiani for their spending precious time reading this manuscript and also for their invaluable comments, advice and encouragement.

I am always deeply grateful to my dear wife Shen Li, who is always there to support me and encourage me. Finally, and most importantly, thank God Father, as "by the grace of God, I am what I am".

Contents

General introduction

Life is a miracle, and many of its mysteries fascinate scientists, such as: what is life? where does it come from? why is like we know it? Each of these questions stimulates our deep desire of understanding nature and also ourselves. Based on the discovery of the structure of DNA by James Watson and Francis Crick in the 1950s and subsequent fundamental discoveries in molecular cell biology, life sciences have experienced two revolutions. The first revolution essentially revealed how the genetic information carried in DNA molecules is expressed into various proteins that exert functions at different levels of cellular organization, which led to the development of molecular biology and of such techniques as recombinant DNA technology. The second revolution occurred following the explosion of data from genomics and proteomics by the mid-1980s.

Recently many engineers, mathematicians and physicists have begun to collaborate with biologists in powerful new ways. The advent of interdisciplinary research in biology has stimulated many discoveries and innovations, bringing new experimental technologies for measuring and manipulating molecular and cellular events with high resolution and powerful computational tools for data analysis. This led to great opportunities to extend our view of molecular cell biology. This is why Susan Hockfield, the sixteenth president of the Massachusetts Institute of Technology (MIT) predicted the coming of the new biological revolution during her speech in 2009 [96].

Quite interestingly, experimental studies revealed that macromolecules in cells like DNAs, RNAs and proteins are extremely dynamical, and that they do not operate separately. In fact, by interacting with each other, macromolecules form complex networks called genetic networks. The structure of genetic networks has been shown to be the key factor explaining how living cells change in space and time to respond to environmental variations and stimuli, make the necessary decisions to stay alive, grow and reproduce, differentiate and perpetuate the species. Therefore, dynamical properties and physiological functions of living cells are decided by genetic networks. We will focus particularly on one typical dynamical behavior of genetic networks which has been highlighted by intensive theoretical and experimental investigations: oscillations. Biological oscillations are involved in many biological functions, for example, circadian rhythms [53, 75, 160] or cell division [129]. Other representative examples are immune response [66], cell growth/death [93], embryo development [2, 60], calcium dynamics [58], etc.

Two main fundamental questions about oscillatory behavior in cellular mechanisms are how (what are the underlying molecular mechanisms) and why (what are the physiological functions of oscillating molecular signals). However, due to the complex topology and huge size of genetic networks, both questions cannot be comprehended by sheer intuition alone. One may view genetic networks similar to electronic circuits. Instead of resistors, capacitors and transistors linked together by wires, genes, RNAs, proteins and other macromolecules

are connected by biochemical reactions. Thus, the question can be raised of whether genetic networks can be usefully described by mathematical modeling, in the same way as electronic circuits are. The answer is yes. Accordingly, many powerful modeling and analysis approaches from physics and mathematics can be applied.

In order to study the design principles of biological oscillations, deterministic mathematical equations are commonly used to describe the time evolution of average quantities of components of genetic networks. Some key ingredients of oscillations have been uncovered. A negative feedback is a necessary condition, but it should be sufficiently delayed by biological processes or intermediate steps that take a certain amount of time; nonlinear degradation mechanisms of RNAs or proteins can also trigger oscillations. It has been suggested that nonlinear degradation is an important source of delay. Morant *et al.* [138] have analytically demonstrated how these key ingredients of oscillations interact and can be traded off against each other in a simple system, a self-repressing gene where gene activity responds slowly to variations in protein level [85].

Due to small copy numbers of reacting molecules, the genetic networks that regulate cellular dynamics are subject to noisy stimuli and large intrinsic fluctuations. These fluctuations have been proposed as useful mechanisms responsible for phenotypic heterogeneity [157, 120], coordinated expression of a large set of genes and probabilistic differentiation strategies [195, 35, 5, 77]. Increasing experimental and theoretical investigations suggest that cellular fluctuations play crucial roles in the design principle of the livings [195, 35, 5, 77]. In view of this, deterministic equations which assume biochemical reactions with an infinite number of participating molecules are not appropriate. There is a growing need for mathematical modeling approaches which takes into account fluctuations. One such approach is the master equation, however there are few cases where it can be solved analytically.

Fluctuations in genetic networks arise not only from elementary biochemical reactions, but also from these highly regulated and multiple-steps processes such as transcription [165, 44, 104, 142, 143, 121], in which macromolecular enzyme RNA Polymerases (RNAPs) synthesizes the mRNA according to the gene sequence. A specific dynamical behavior of RNAP which has attracted much interest is the stochastic pausing where the RNAP is halted at a nucleotide along the genetic sequence and which is observed in both prokaryotes and eukaryotes. Pausing can severely affect transcription dynamics, probably contributing to the transcriptional bursting widely observed in most of the genes of prokaryotes [85, 20, 71], yeast [198, 181, 19] and eukaryotes including mammalians [156, 155, 22, 42, 44, 176].

Motivated by important roles and the intriguing dynamical behaviors of genetic networks, we will theoretically elucidate dynamical effects of time delay, fluctuation and transcriptional pausing on genetic networks. This thesis work mainly includes three studies.

Combination of time delays and biological oscillations

The first study is about delays, one of the key ingredients of biological oscillations. Various sources of delay exist in cellular processes, such as transcription and translation which feature many intermediate steps and take a minimum amount of time, molecule transport, phosphorylation and etc. Understanding how these various time delays combine is crucial to explain the robustness of experimentally observed oscillations [175], however this has been relatively little studied. A further question is whether delays from various sources play similar dynamical

roles. In order to address these questions, the action of a delay must be specified mathematically. In the literature, there are actually two different strategies for modeling delays. In some studies, the delay appears in an explicit manner as the time-delayed value of some dynamical variable [137, 127, 184] and is referred to as "explicit delay". In other studies, it is originated in a reaction step [123, 80] and called "reaction delay". Even though both of them are commonly used, their similarities and differences have not yet quantitatively studied. In this first study, we will investigate the combination principles of delays and compare the different influences of explicit and reaction delays on oscillatory behavior. To this aim, we will study a self-repressing gene circuit comprising delays respectively due to the dynamics of gene activity and to molecule transport.

Influence of fluctuations on the biological oscillation of genetic networks

In the second study, we investigate the influence of fluctuations on the oscillatory behavior of genetic networks. For this purpose, we will develop a cumulant expansion derived from the master equation to incorporate fluctuations in a deterministic description. We will apply this approach to the same self-repressing gene circuit we have studied before. In this simple circuit, fluctuations are due to the small copy numbers of mRNA and proteins, and more importantly to the binary nature of gene activity, which undergoes transitions between the "ON" and "OFF" states. We will show how fluctuations shift the steady state predicted by deterministic equations and how fluctuations can induce oscillations in this reduced system.

Stochasticity of transcriptional pausing and its dynamics effect

The aim of the third study is to study the dynamical effects induced by transcriptional pausing and its influence on transcription rate. To this aim, we will introduce a classical system in out-of-equilibrium statistical physics named the Total Asymmetric Simple Exclusion Process (TASEP). We will use it to model transcription and study how the dynamics of transcription is affected by pauses occurring in a stochastic fashion with weak dependence on DNA sequence. For simplicity, we will restrict ourselves to the case of TASEP model with periodic boundary condition. It will be found that mean-field approach works well only in the limit case where pause are short. We will then suggest a statistical approach to compute the transcription rate modified by pausing over the entire range of pause duration. An expression of the transcription rate agreeing well with numerical simulations will be given.

Plan

In Chapter 1, we first introduce the motivation of this thesis work. Then we present different mathematical approaches to describe the dynamics of genetic networks, such as deterministic equations which are helpful to reveal the key ingredients of biological oscillations, the cumulant expansion of the master equation which allows to take into account fluctuations, and the TASEP

model describing the transcription process. In Chapter 2, we introduce an extension of the model proposed by Morant *et al.* [138]. In this model there are two delays respectively due to gene response dynamics and protein transport, with the latter being modeled as a reaction or an explicit delay. We will discuss the combination of these two delays and compare the different influences of reaction and explicit delays on oscillatory behavior. In Chapter 3, we take the self-repressing gene circuit as an example to apply the cumulant expansion of the master equation. This approach allows us to investigate the influence of fluctuations on the steady state and oscillation regions predicted by deterministic models. In Chapter 4, we study how transcription is affected by RNAP pausing in using the TASEP model and predict the transcription rate for all pause durations.

Chapter 1

Dynamical modeling of gene networks

1.1 Biological motivation

1.1.1 Biological oscillations

Living cells may be viewed as dynamical systems which have to function and maintain themselves in highly changing environments. They receive information from inside and outside the cells, process this information to make decisions, and then trigger responses which are appropriate for survival and reproduction. Essential cellular functions rely on complex networks in which macromolecules, such as genes, RNAs and proteins, interact with each other via biochemical reactions so as to generate collective behaviors. These genetic regulatory networks are intrinsically dynamic and highly nonlinear. Understanding their dynamics is the key to describe the temporal and spatial changes that a cell undergoes to respond to stimuli, grow and reproduce, differentiate and do other important and necessary activities.

The development of new quantitative real-time measures of molecular dynamics in living cells, but also the capacity to synthesize genetic networks, has shed light on a striking dynamical behavior of genetic networks: oscillation [114, 92, 87, 95, 83, 84, 6, 144, 41]. Oscillatory behavior is characterized by periodically repeating variations of some measures or quantities and is observed in a great number of biological systems, with a wide range of timescale from seconds to days (Table. 1.1). Some representative examples of biological oscillations are found in the cell division cycle [129], circadian rhythms [53, 75, 160, 3, 32], immune response [66], embryo development [95], cell death and apoptosis [93, 52], and cellular calcium dynamics [58]. It has been suggested that oscillations are more efficient than a steady state to encode and transfer information both in time and space, and that they actively contribute to robust regulations of biological functions at different levels of cellular organization and as well as to flexible responses to environmental variations [21, 148]. However, oscillations are not easy to understand, because they do not result from the action of a single molecular actor but from the combined interaction of several of them. We review below a few biological systems that can display sustained oscillations, as well as the components involved in the oscillating network.

In vertebrate embryo, the somites [2, 60] are the basis of the segmental pattern of the body and give rise to the axial skeleton and the dermis of the back. Fig (1.1) illustrates the phenomenon of somite formation in Zebrafish [137], which has attracted a lot of interest. We see a spatial pattern which is not unlike a trace on a magnetic recording tape. It is actually a record of the temporal cyclic variation in expression levels of some genes in the cells of the

Function	Components	Period
Metabolism	Glucose, ATP, phospho-fructokinase	2 min
Somitogenesis	Her1, Her7, Notch	30-90 min
Signaling	Cyclic AMP, receptor, adenylate cyclase	5 min
Signaling	Ca^{2+}, lns(1,4,5)P_3	>1 s
Signaling	NF-kB, lkB, lKK,	$\sim$ 2 h
Signaling	p53, MDM2	5 h
Yeast endoreplication cycles	Cig2, Cdc10, Rum1	1-2 h
Frog egg cycles	Cycb, Wee1, Cdc25, Cdc20	30 min
Circadian rhythm	PER, TIM, CLOCK, CYC	24 h

Table 1.1: **Survey of biological systems displaying sustained oscillations**. Table taken from [145].

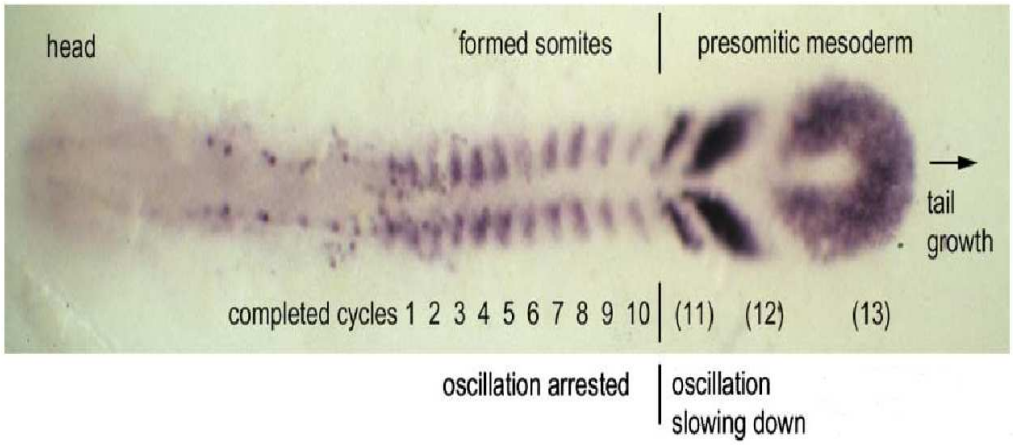

Figure 1.1: **A Zebrafish embryo at 10-somite stage**. This spatial pattern of segmentation is a record of temporal oscillation of the expression process of protein Her1 and Her7. Figure taken from [137].

presomitic mesoderm (PSM). Two genes, Her1 and Her7, and their associated proteins, which are responsible for the somite formation in Zebrafish, have been identified, since when these genes are knocked down by morpholino injections or chromosomal deletion, oscillations break down. The Her1/Her7 system has often served as a paradigmatic example of a simple genetic circuit comprising self-repressing genes which will be studied extensively in this thesis.

The second example of oscillations involves one of the most intensively studied proteins, the tumor suppressor p53 which is involved in preventing cancer in multicellular organisms. It functions by triggering DNA repair in case of damage, but is also a key protein involved in cell death and apoptosis pathways, which may be required if DNA cannot be repaired. In advanced and accurate single-cell experiments, the p53 protein was fused to fluorescent proteins so that the variation of its copy number over time could be recorded [93, 52]. After a stress-induced DNA damage, p53 is activated and its concentration increases. The increase of p53 protein levels in turn leads to the production of its degradation enzymes. p53 is then rapidly degraded, which after some time reduces the production of the degradation enzymes, and so on. This re-

sults in very dynamical variations of p53 levels in individual cells. It can be seen in Fig 1.2 that the interaction between p53 and its degradation enzymes gives rise to an oscillatory behavior. This causes particular interest to understand the mechanism underlying the appearance of these oscillations.

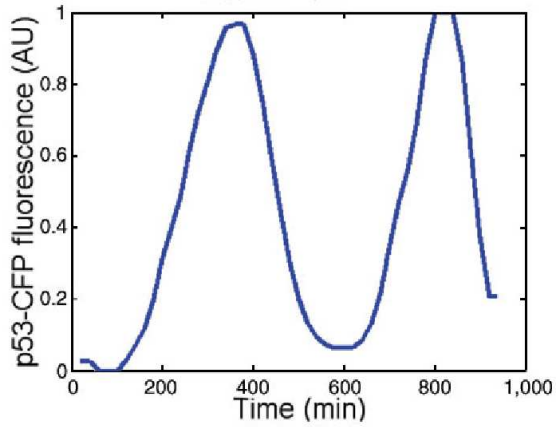

Figure 1.2: **Dynamics of tumor suppressor p53**. After DNA damage, the concentration of protein p53 fused by fluorescent proteins is measured in real time in individual cells and is found to display oscillatory behavior. Figure taken from [52].

Circadian rhythms provide another classical example of biological oscillations. Many organisms, including humans, have their life organized by the day/night cycle with period of 24 hours. During the day, for example, we stay awake and active, and at night we sleep and regain energy for the next day. People do not behave so because they are disciplined and follow a well organized life schedule but because their physiology is rythmically driven by a biological clock whose workings are gene regulatory networks oscillating inside our cells [53, 75, 160]. One piece of evidence for the existence of these so-called circadian clocks is the jet lag effect. After traveling over a long distance in a short amount of time, our circadian clock cannot adapt instantaneously and remains for some time phase-shifted with respect to the outside day/night cycle, which results in discomfort.

Quite remarkably, plants display also circadian rhythms. For instance, there is a plant called *Phaseolus coccineus* [40] whose leaves open during the day and close at night (Fig 1.3). Interestingly, if it is put in continuous light or darkness, the daily movements persist, which reveals that its endogenous circadian clock evolves independently of sunlight. The endogenous nature of circadian oscillations is further shown by the fact that in general the period of endogenous circadian oscillations is slightly different from 24 hours. Circadian clocks have been found in insects, fungi and cyanobacteria [3, 32]. Their physiological function is to help living organisms to adapt to the environmental changes and accordingly to regulate various physiological properties such as body temperature, feeding behavior, etc. The molecular components of the genetic networks generating circadian rhythms have been identified in different organisms, such as the proteins KaiA, KaiB and KaiC driving the circadian clock in *Synechococcus* or the protein frq involved in *Neurospora* clock [3].

Understanding the molecular basis of the genetic networks that are responsible for cellular

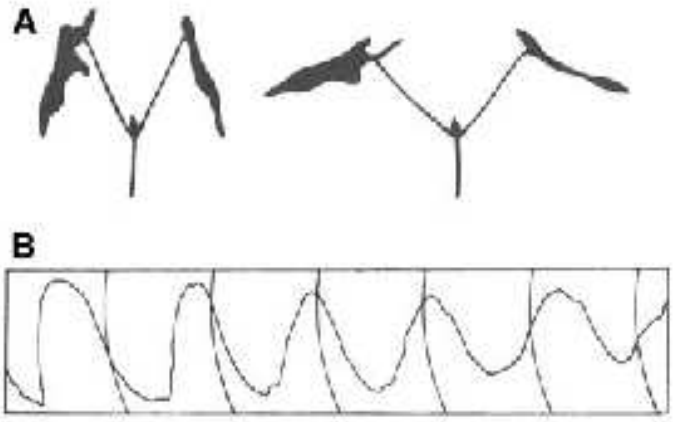

Figure 1.3: **Circadian rhythm in the plant *Phaseolus coccineus***. (A) Leaves of *P.coccineus* close at night (left) and open during the day (right). (B) Circadian rhythms of leaf movement in continuous light. The peaks of the curve represent the leaf position at night. The vertical lines delimitate time intervals of 24 hours. Therefore, the period of endogenous circadian oscillations in this plant is $\sim$ 27 hours. Figure taken from [40].

oscillations is crucial to unravel the dynamics of cells. To understand how oscillations emerge collectively from a molecular network requires first that we understand the molecular interactions that form the building blocks of this network. In cellular processes, the primarily involved molecules are DNA, mRNA and proteins, and the elemental biochemical processes in which they participate are synthesis (such as transcription or translation), degradation, complexation, covalent modifications, etc. In the next section, we will review the details of these elemental biochemical processes.

1.1.2 Elemental biochemical processes in cells

Transcription and translation We start with DNA molecules. DNA is a nucleic acid whose two strands are entangled as a double helix. It is composed of simple units, called nucleotides, and carries the genetic information that controls phenotype and functions of cells in all living organisms by directing the synthesis of proteins. The stability of this information is ensured by its discrete coding, and the segments of the DNA molecule which carry it are called genes. As for mRNA, it is a single-stranded nucleic acid made up of nucleotide components. As we describe below, mRNA results from the transcription of the genetic sequence and serves as a template for synthesizing proteins in a process called translation. Proteins are polymeric chains of amino acids typically folded into a specific form suitable to its function. Proteins are major components of cells, and play crucial roles in all cellular processes. For examples, they can participate in biochemical reactions and metabolism as vital enzymes but are also involved in cell signaling and other important cellular mechanisms.

According to the central dogma of molecular biology (Fig 1.4), the genetic information carried by genes is first copied into mRNA molecules during transcription, which is carried out by enzymatic macromolecules named RNA polymerase (RNAP). Then genetic information is transferred from mRNA to proteins by other macromolecules called ribosomes via the process

of translation. The whole process in which genetic information in genes is used to synthesize matured and functional proteins is referred as to gene expression. Now we examine the details of these two main processes of gene expression: transcription and translation.

Generally, the three main steps of transcription are initiation, elongation and termination. During initiation, RNAP and transcription factors (TFs), such as the σ factor involved in bacterial transcription, bind to a particular DNA region near the gene which is called the promoter. Then RNAP and TFs binding the promoter interact with each other and transform into an elongation machine which is ready to move along the gene and to transcribe genetic information. During elongation, RNAP, consuming energy brought by ATP molecules, moves along one strand of the DNA template and reads the genetic sequence nucleotide by nucleotide, whereby a freshly synthesized mRNA transcript comes out from the trail of RNAP. Elongation is a complex process. The dynamics of RNAPs is rather stochastic and they may have unusual behaviors which we will discuss at the end of Chapter 1. In addition, the elongation process involves a number of other enzymes which are also subject to stochastic dynamics. When RNAP reaches the last nucleotide of the gene, the final termination step occurs, and RNAP and TFs dissociate from DNA molecule. The mRNA transcript is released from RNAP, undergoes folding and phosphorylation, and then becomes mature mRNA which is ready to be translated for proteins.

Translation is quite similar to transcription. Compared to transcription, the template of translation is the mRNA sequence instead of the gene, and the enzymatic macromolecules reading the sequence are ribosomes. As translation begins, ribosome follows the template sequence nucleotide by nucleotide and binds transfer RNAs (tRNAs) which themselves fix the amino acids which are bricks of which proteins are made. When ribosomes reach the end site of mRNA, a polymer chain of amino acids is finished and eventually becomes a functional protein after undergoing posttranslational modifications.

Regulation As we have seen, gene expression consists of multiple combined steps. Each of these steps requires many macromolecules and factors so that it is highly regulated. Let us consider transcription regulation [90]. In fact, the transcription rate is highly dynamical. Even though the promoter is the region of DNA where RNAPs and TFs are meant to bind, it can actually be bound by other specific proteins. Their presence can either favor or inhibit the binding of RNAP and TFs, and thus control the transcription rate. The proteins preventing RNAP or TFs binding are called repressors. Likewise, those enhancing transcription are called activators. Repressors and activators are synthesized by their associated genes. Thus, genes interact with each other by driving the expression of each other via regulation.

Molecule transport After mature macromolecules like mRNAs and proteins are synthesized, they are transported to some organelles or other locations where they exert their specific functions. In particular, in eukaryotic cell, on one hand, mRNAs synthesized in nucleus needs be transported to the organelles of ribosomes in cytoplasm where translation occurs. On the other hand, those functional proteins which regulate transcription have to be relocated from cytoplasm to nucleus to control gene expression.

Molecule degradation Various mechanisms leading to the destruction of macromolecules play also an important role in biological oscillations [16]. Molecule degradation is actually a complex process [141] in which many specific enzymes are involved. For example, protein

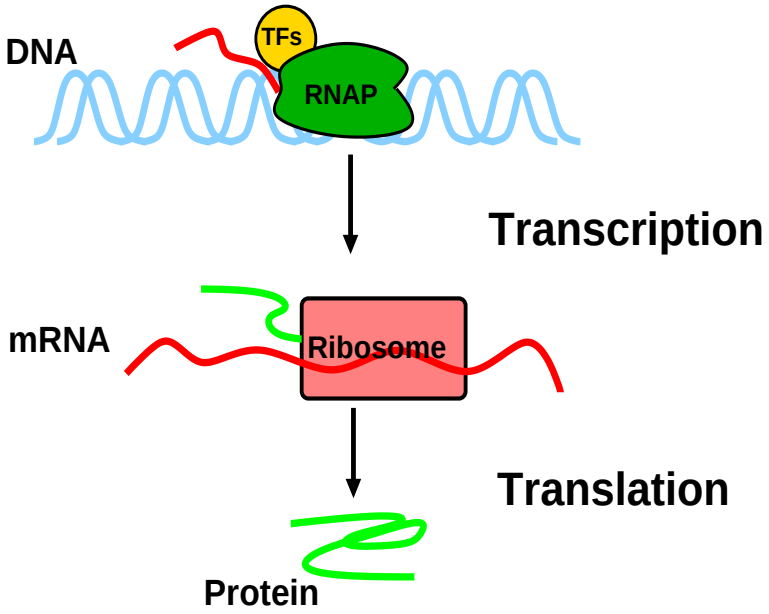

Figure 1.4: **Transcription and translation**. Gene expression mainly consists two important and complex processes: transcription and translation. During transcription, macromolecule RNA polymerase (RNAP) and transcription factors (TFs) bind DNA molecule and read through its sequence while the mRNA is synthesized behind them. Translation is similar to transcription, but is catalyzed by other macromolecules called ribosomes.

degradation is performed by enzymes called proteases, and mRNA degradation by ribonucleases. Thereafter we take as an example protein degradation. One of the most typical protein degradation mechanism is the so-called Michaelis-Menten degradation [141] in which a protein molecule P bound by protease E forms a complex C which in turn is converted into a product S ultimately destroyed and the protease E is released. Thereby, proteins are converted into smaller compounds but degradation enzymes, proteases, remain. The kinetic of biochemical reaction can be described by:

$$P + E \rightleftharpoons C \rightarrow S + E$$

If the enzymes are abundant, degradation is linear in the sense that the number of proteins degraded by unit time is proportional to the total number of proteins. In the case where their copy number is smaller than that of proteins, there is a competition between proteins for degradation enzymes and degradation becomes nonlinear. Furthermore, when the amount of degradation enzymes is further reduced, degradation becomes saturated, and the number of proteins degraded per unit time becomes independent of protein copy number, because it is in fact limited by enzyme copy number. A more complex degradation mechanism is allosteric degradation [102] where the protein inhibits or enhances the activity of the degradation enzyme by binding to one of its sites. In this case, the kinetics of degradation is much more complicated. We introduce here only two examples of how molecular degradation can be highly nonlinear, but in fact there are much more.

Dimerization and phosphorylation We have yet to introduce two important elemental biochemical processes which will considered later in this thesis work: dimerization and phospho-

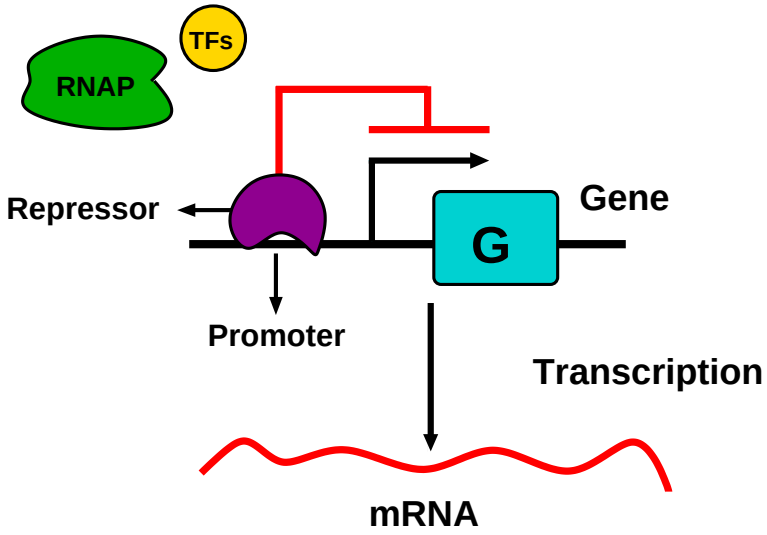

Figure 1.5: **Transcription regulation**. The promoter is a particular region near the beginning of a gene. The binding of a repressor to the promoter prevents attachment of RNAPs and TFs so that it represses the transcription rate.

rylation.

Dimerization is a common process in biological systems [131, 7]. In this process two identified units A called monomers bind to each other and form a new molecule A_2 named a dimer. Many functional proteins in living cells, such as transcription factors, act in the form of dimers, because this provides a way to modify the chemical kinetics. The dimerization process of a monomer A can be explicitly described by:

$$A + A \rightarrow A_2$$

Phosphorylation is also a commonly encountered process in which a phosphate group PO_4^{3-} is covalently attached to the protein. Phosphorylation of many enzymes modifies their conformation and controls their activities.

1.2 Genetic networks

We have reviewed elemental biochemical processes in cells. In this section, we will introduce genetic networks where elemental processes are carried out and discuss their typical dynamical behaviors.

1.2.1 Natural genetic networks

In cells, transcription of genes lead to mRNAs which in turn serve to produce functional proteins. Moreover, the presence of certain proteins can regulate the transcription rates of some genes, including their own genes. Thus genes indirectly promote or inhibit the rate of expression of each other. A collection of genes interacting via mRNA and proteins is called a genetic network [94].

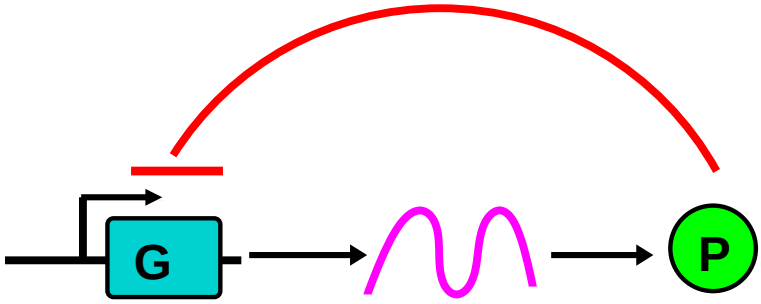

Figure 1.6: **A minimal genetic network**. In this circuit, the gene synthesizes mRNA which produces the associated protein. The proteins represses the transcription of its own gene, which forms a negative feedback loop. This minimal genetic network actually exists in the nature, such as Hes1, Hes7, involved in somite segmentation, or LexA which is the SOS regulon in *E.Coli*.

The simplest genetic network is a loop in which proteins synthesized via transcription of a given gene and subsequent translation regulate the transcription rate of this gene (Fig 1.6). For example, direct binding of protein Hes7 to its own promoter represses the expression of its own gene, which has been demonstrated in mice [12, 127]. Other examples are repressor LexA as the SOS regulon in *Escherichia coli* [168], Hes1 [100] in the segmentation and etc.

The genetic networks based around a single gene are not the rule. As there are thousands of genes in DNA molecules (Fig 1.7), the majority of genetic networks existing in nature are much more complex and involve anything from a few genes to much more. Accordingly, understanding their dynamics can quickly become very complex.

1.2.2 Synthetic genetic networks

Synthetic biology combines biological and engineering approaches to construct artificial genetic networks which comprise only a few components and functions in isolation [79, 37, 102]. This allows one to overcome the complexity and coordination of natural genetic networks. With small synthetic genetic networks as testbeds, typical dynamical behaviors can easily be studied.

Here we take a synthetic network reported [37] as an example (Fig 1.8). This synthetic network involves a negative feedback loop in which the first gene, LacI taken from *E.Coli*, inhibits the transcription of the second gene, TetR, which in turn represses transcription of a third gene, cI from λ phage. Finally, cI prevents the expression of LacI. When parameters of this synthetic network are tuned, two dynamical behaviors are possible. The system may settle onto a stable steady state, or the steady state may be destabilized, leading to sustained oscillations. This example shows the simplicity and the power of the synthetic biology in studying the dynamics of genetic networks by providing testbeds for theoretical modeling.

We should mention that synthetic genetic networks may display another typical behavior: bistability [79, 154, 101], in which the system has two possible steady states and transits between them. However, we have chosen to focus here on oscillatory behaviors. In the next section, we will try to understand the design principles responsible for this behavior, using deterministic modeling.

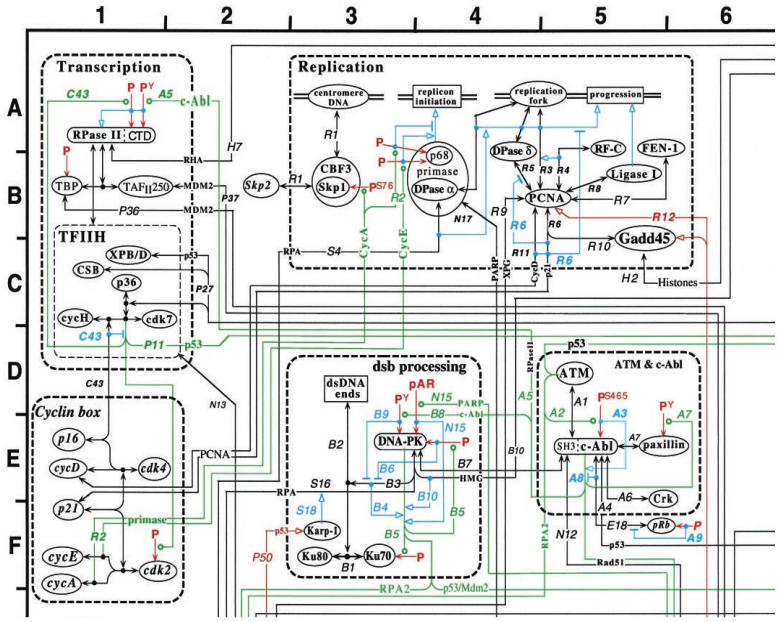

Figure 1.7: **A typical genetic network in mammalian cells**. Components of the genetic network which are presented as a map are macromolecules involved in the cell cycle of mammalian cell nucleus. Lines connecting nodes represent biochemical reactions. Figure taken from [111].

1.3 Deterministic modeling of genetic networks and key ingredients for oscillations

For exploring quantitatively the design principles of genetic oscillators, a mathematical description of genetic networks is required. The commonly used approach is deterministic modeling based on the law of mass action in which the reaction rate is proportional to the concentration of reactants [141, 103, 91]. In this way, we obtain a set of deterministic rate equations whose variables are the concentrations of components of genetic networks, such as concentrations of mRNAs and proteins. Assuming a genetic network comprising N components, a general mathematical description can be written as follows:

$$\frac{dx_i(t)}{dt} = f(x_1, x_2, ..x_i, ..x_N) - g(x_1, x_2, ..x_i, ..x_N)$$

where x_i is the activity of gene i or the concentration of molecule i. Interactions between molecular actors, such as transcription, translation, regulation, degradation and other elemental biochemical processes we have described in Section 1.1.2, are included in these two functions $f(x_1, x_2, ..x_i, ..x_N)$ and $g(x_1, x_2, ..., x_N)$ which describe the gain or loss mechanisms for component i.

1.3.1 Necessary ingredient for biological oscillations: negative feedback

We introduce here a necessary ingredient for biological oscillations, negative feedback, and its mathematical modeling. In classical mechanics, the simple harmonic oscillator displays

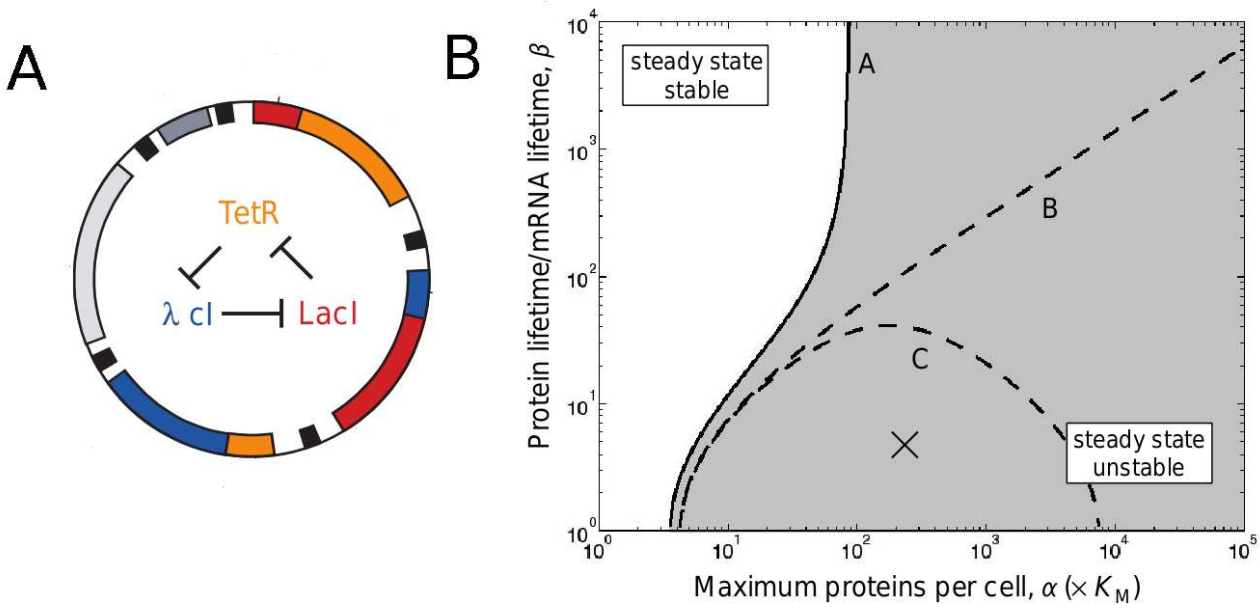

Figure 1.8: **Construction and oscillatory behavior of a synthetic network**. (A) The synthetic genetic network is a regulating loop composed of three genes (LacI, tetR and cI). (B) Two different dynamical behaviors corresponding to two regions in parameter space are found: this simple synthetic genetic network settles onto a stable steady state (top left) or displays sustained oscillation because of unstable steady state (bottom right). Figure taken from [37].

sustained oscillations. When displaced from its equilibrium position, oscillator experiences a restoring force that intends to bring it back to the equilibrium position. In fact, the restoring force plays a role of negative feedback. Quite similarly, the negative feedback is also a necessary condition for biological oscillation appearing in genetic networks [13], as has been shown mathematically [174, 88].

In order to illustrate the mathematical modeling of a negative feedback loop, we consider the example of a minimal genetic network consisting of one self-repressing gene. In this genetic network, proteins inhibiting the transcription of their own gene create the negative feedback. To be precise, when there are only a few proteins, protein synthesis of the gene is not affected. When proteins become abundant, they can bind to the promoter of gene and thereby inhibit the transcription. After proteins are degraded, the gene resumes the protein synthesis. The kinetic of this minimal genetic circuit can be described by [184]:

$$\frac{dp(t)}{dt} = k_1 \frac{K_d^n}{K_d^n + p^n} - g(p) \tag{1.1}$$

where p is the concentration of protein; k_1 is the maximum transcription rate when there are few copies of protein and gene is fully expressed. The term $\frac{K_d^n}{K_d^n+p^n}$ is used to model the negative feedback and describes how gene transcription is repressed by the presence of the protein. In this term, n is an integer indicating whether protein binds to the gene as a monomer ($n = 1$), dimer ($n = 2$), trimer ($n = 3$) or other forms, K_d characterizes the protein-DNA dissociation constant. For small protein copy numbers, $p \ll K_d$, the synthesis rate is approaching a constant. The synthesis rate decreases monotonically as a function of protein concentration p. In the limit

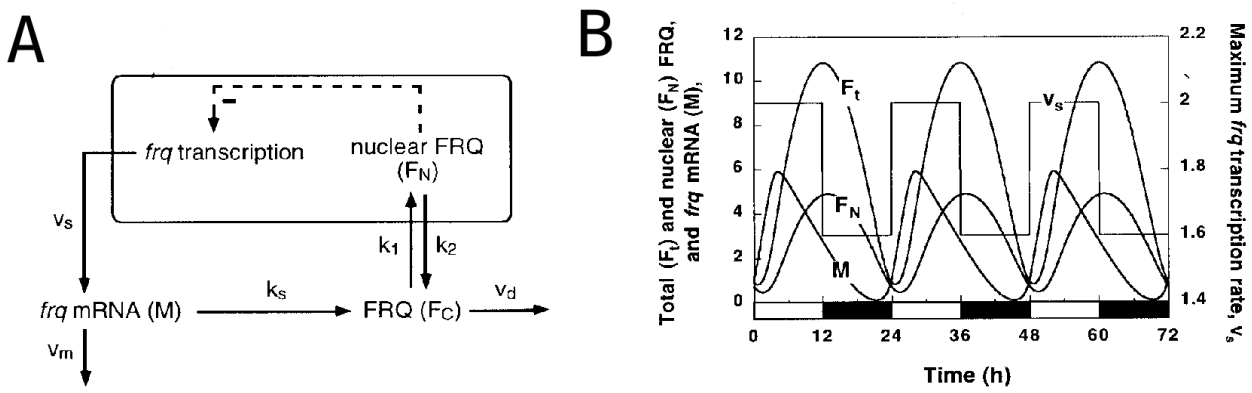

Figure 1.9: **A model of circadian rhythms with a reaction transport delay and its corresponding oscillatory behavior**. (A) Biochemical reactions of the negative feedback loop of protein FRQ in *Neurospora* are presented. The protein FRQ is synthesized in cytoplasm and then transported into nucleus to regulate transcription of the *frq* gene. (B) A numerical simulation of Eqs. (1.2) shows oscillations with a period close to 24 hours. Figure taken from [123].

case where protein concentration is sufficiently large ($p \gg K_d$) the production of proteins is completely inhibited. In this way, the negative feedback can be mathematically described.

1.3.2 Delay as the second key ingredient for oscillations

When analyzing Eq. (1.1), we find that the state of constant protein concentration is always stable. Any perturbation in protein concentration immediately relaxes to zero because of the negative feedback. Therefore, another key ingredient is needed for oscillations to appear. It is found that if the negative feedback is delayed, oscillations appear systematically [137, 100, 127, 135, 123, 11, 180, 133, 175, 80, 145]. This can be understood with a simple example: consider a person who walks straight towards a point marked on the ground. When he is approaching to the point, he will reach it and stop there if he can make the decision instantaneously. This resembles a genetic network in which the negative feedback signal is not delayed. However, if this person have to take some time to realize the fact that he has reached the point, he will cross it, then turn around and walk towards it again. Certainly he will reach and cross that point again and again. Sustained oscillations are thus generated.

Therefore delay is also a dynamical ingredient which plays a crucial role in biological oscillations. There are actually various sources of delay in biological systems, for example, transcription and translation take a minimum amount of time. Molecular transport, phosphorylation and other intermediate steps of biological processes also lead to delays. Two main strategies are used to model a delay. In some studies, the delay originates in a reaction step [123, 80], and we will call such a delay a "reaction delay". In other studies, the delay appears explicitly as a time-delayed value of some dynamical variable [137, 127, 184], without specifying the underlying processes. We shall thereafter refer to such a delay as an "explicit delay".

Ordinary Differential Equation (ODE) model with a delay induced by a reaction step First, we consider a model of circadian rhythms in *Neurospora* [123, 80], in which delay originates from a reaction step. In this model (Fig 1.9), *frq* mRNAs are transcribed in the nucleus and transported into cytoplasm where protein FRQ proteins are synthesized. Proteins are transported reversibly between cytoplasm and nucleus. Negative regulation of gene expression is exerted by proteins in nucleus. Given the biochemical reactions involved in this genetic circuit, the kinetic is described by:

$$\frac{dM(t)}{dt} = v_s \frac{K_I^n}{K_I^n + F_N^n} - v_m \frac{M}{K_m + M} \tag{1.2a}$$

$$\frac{dF_C(t)}{dt} = k_s M - v_d \frac{F_C}{K_d + F_C} - k_1 F_C + k_2 F_N \tag{1.2b}$$

$$\frac{dF_N(t)}{dt} = k_1 F_C - k_2 F_N \tag{1.2c}$$

The set of Ordinary Differential Equations (ODEs) thus obtained describe the temporal evolution of each components according to the law of mass action. In Eqs. (1.2), M, F_C and F_N are respectively *frq* mRNA copy number, FRQ protein copy numbers in the cytoplasm and in the nucleus. The term $v_s \frac{K_I^n}{K_I^n + F_N^n}$ denotes the transcription rate under control of FRQ proteins. The term $v_m \frac{M}{K_m + M}$ (or $v_d \frac{F_C}{K_d + F_C}$) is the degradation function of mRNA (or FRQ proteins in cytoplasm) following a Michaelis-Menten mechanism (see Section 1.1.2) in which v_m is the maximum degradation rate and K_m is the Michaelis-Menten constant. k_s is the translation rate of FRQ protein. k_1 and k_2 characterize the protein transport rates into and out of nucleus.

Here proteins are separated into two different compartments: cytoplasm and nucleus. The protein transport between cytoplasm and nucleus leading to the delay ensures that synthesized proteins cannot inhibit gene transcription instantly. It can be proved that without the delay due to protein transport, the steady state of Eqs. (1.2) remains always stable. Therefore, the delay specified by reaction delays due to protein transport is crucial for this circadian model to display sustained oscillations. Although reaction delays increase the dimension of system, the advantage of models described by ODEs is that they are easily analyzed. Approaches for analyzing ODE dynamics are well developed, such as linear stability analysis.

Delayed Differential Equations (DDEs) model involving an explicit delay The second strategy to model delays is to introduce a variable with an explicit delay in the equations, which are then termed Delayed Differential Equations (DDEs). Such equations were first introduced in [130] and have been since widely used to describe mathematically the kinetics of biological networks [137, 100, 127, 135, 11, 133, 80, 145].

To illustrate this concept with an example, we consider a small genetic circuit involving the transcription factor Hes1 which regulates the expression of its own gene via a feedback loop [95]. The kinetics of this genetic circuit is described by the following DDEs:

$$\frac{dr(t)}{dt} = \frac{\alpha k^h}{k^h + s(t-\tau)^h} - k_r r(t) \tag{1.3a}$$

$$\frac{ds(t)}{dt} = \beta r(t) - k_s s(t) \tag{1.3b}$$

In Eqs. 1.3, s and r are concentrations of protein Hes1 and of the corresponding mRNA, respectively. β is the translation rate. k_r and k_s are the spontaneous degradation rate of mRNA and proteins, respectively. The production of mRNA and regulation by Hes1 are described by the term $\frac{\alpha k^h}{k^h+s(t-\tau)^h}$. The mRNA is fully expressed when the gene is not bound by Hes1 and is produced at a rate α. Otherwise, binding of Hes1 to its own gene reduces the transcription rate. The explicit delay τ accounts for the fact that mRNA synthesis rate depends of the protein concentration at time $t-\tau$ rather than at time t. This delay may be due to time needed to complete transcription, translation or any other intermediate biological step. Importantly, the model described by Eqs 1.3 can display sustained oscillations [100]. Compared with equations featuring a reaction delay, the advantage of using an explicit delay is that we can neglect the detail of biochemical processes inducing a delay. On the other hand, the dynamical analysis of a system described by DDEs requires more sophisticated approaches.

The important point we want to make is that a given process can be modeled either with a reaction delay or with an explicit delay. How do we choose between them in our modeling strategy? Indeed whether or not they have similar dynamical behaviors, especially for inducing oscillations, remains unclear.

1.3.3 Nonlinear degradation and oscillations

As mentioned in Section 1.1.2, various degradation mechanisms can affect macromolecules like mRNA and proteins in cells, such as linear, Michaelis-Menten or allosteric degradation. They lead to very different behaviors of the degradation rate as a function of the concentration (Fig 1.10). In particular, degradation functions describing Michaelis-Menten mechanism have been utilized in Eqs. (1.2). In general, degradation mechanisms play important roles in shaping the dy

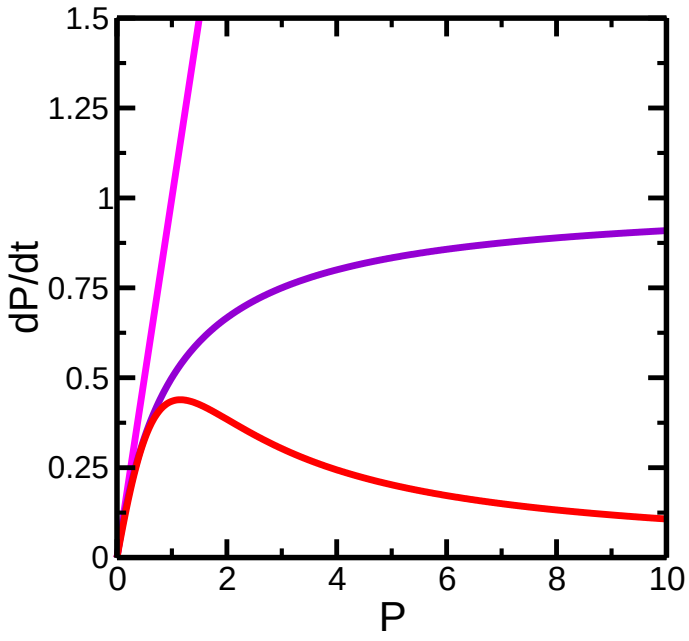

Figure 1.10: **Functions associated with different degradation kinetics**. Three functions, which mathematically represent the kinetics of linear (red curve), Michaelis-Menten (magenta) and allosteric (blue) degradation are shown.

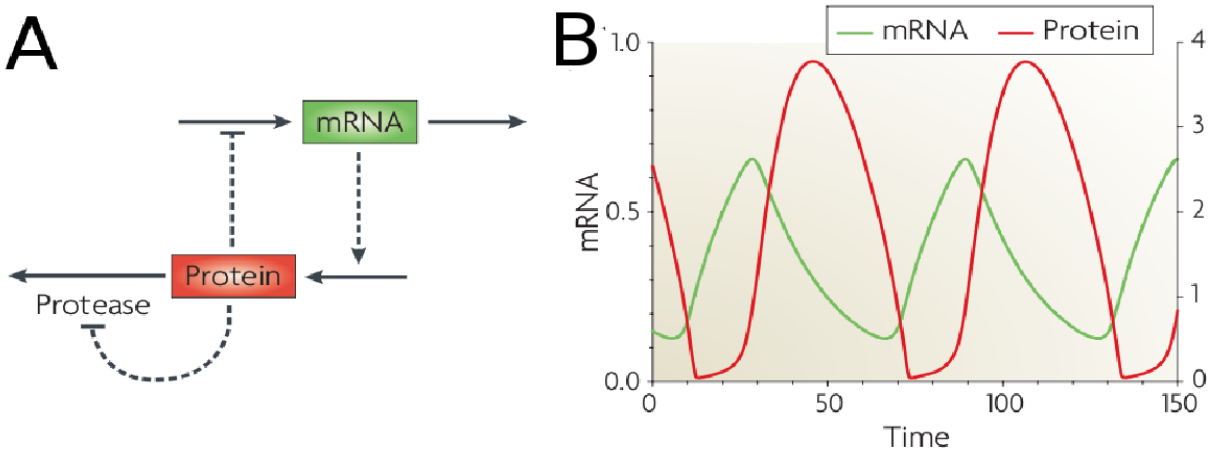

Figure 1.11: **Oscillations driven by nonlinear degradation**. (A) Protein and mRNA are involved in a negative feedback loop where the protein represses its own expression. In addition, it can bind to an allosteric site in its degradation enzyme, the protease, and inhibit protease activity. (B) Sustained oscillations in mRNA and protein concentrations are observed. Figure taken from [145].

To introduce the mathematical modeling of degradation and its role in oscillations, we will focus on a circadian model describing the expression of the *PER* gene in fruit flies [145]. In this example, the protein PER represses the expression of its own gene, but it can also bind to an allosteric site on its degradation enzyme, the protease, and thereby inhibit protease activity (Fig 1.11A). The system is modeled by the following equations [145]:

$$\frac{dX(t)}{dt} = k_{sx}\frac{K_d^p}{K_d^p + Y^p} - k_{dx}X \tag{1.4a}$$

$$\frac{dY(t)}{dt} = k_{sy}X - k_{dy}Y - \frac{k_T Y}{K_m + Y + K_I Y^2} \tag{1.4b}$$

In Eqs. (1.4), Y and X denote respectively the concentration of protein PER and its mRNA. $k_{sx}\frac{K_d^p}{K_d^p+Y^p}$ describes the regulation of *PER* transcription rate by protein PER. k_{sy} is the translation rate of PER proteins. k_{dx} is the spontaneous mRNA degradation rate. k_{dy} is the rate constant for an alternative pathway of protein degradation. Both of them are associated with linear degradation mechanisms. The term $\frac{k_T Y}{K_m+Y+K_I Y^2}$ represents the allosteric degradation and is a second-order function. The presence of this nonlinear degradation term is essential to trigger sustained oscillations in protein and mRNA concentrations (Fig 1.11).

To summarize, we have discussed in this section three key ingredients of biological oscillations: negative feedback as a necessary condition, delay as well as nonlinear degradation. We have also discussed the fact that genetic networks can be mathematically described by models comprising ODEs or DDEs. These deterministic models have been quite successful to predict oscillatory behavior in genetic networks. However, deterministic models are obtained by assuming that the number of reacting molecules are infinite. This assumption is usually not correct because many molecules involved in cellular reactions present very small copy numbers.

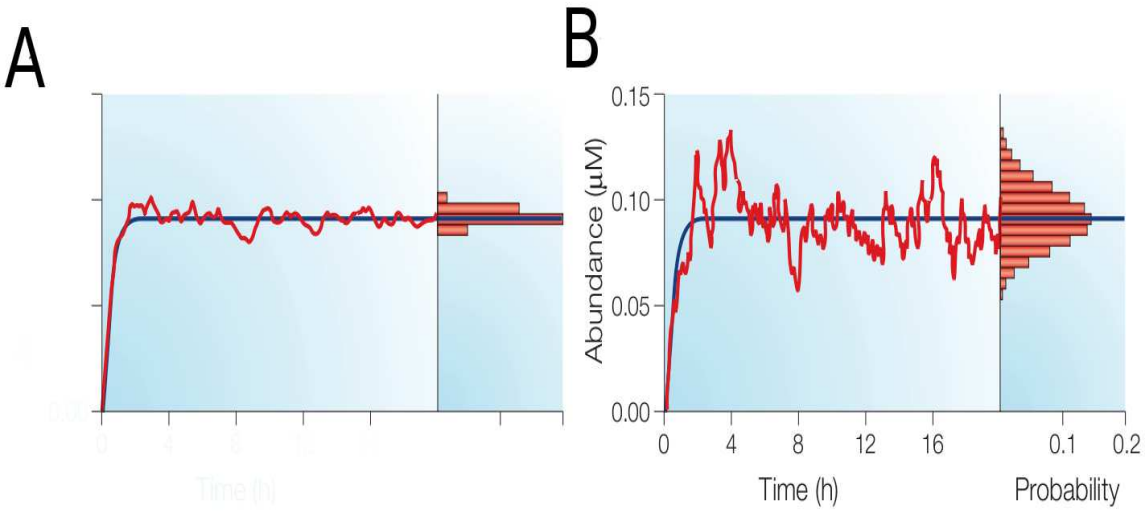

Figure 1.12: **Fluctuations due to small copy numbers**. Time series of protein concentrations generated from a deterministic model and from a stochastic simulation (blue and red curves, respectively). Histograms show the probability that a cell will have a given protein concentration. (A) Low-amplitude fluctuations with high numbers of protein molecules ($\sim 10,000$). (B) Large fluctuations correlate with a decrease in the number of protein molecules (~ 100). Figure taken from [110].

In the next section, we will discuss effects due to the small and integer numbers of molecules, which result in fluctuations (or noises) which can influence the oscillatory behavior of genetic networks.

1.4 Stochastic fluctuations in genetic networks

1.4.1 Fluctuations due to low copy number of molecules

Deterministic rate equations are a widely used approach for modeling the dynamics of genetic networks. In this context, copy numbers of macromolecules are assumed to be very large, so that the relative error made by describing them with continuous variables is very small (Fig 1.12A) in spite of the presence of fluctuations [110, 196]. More precisely, deterministic models usually describe the time evolution of the averages of molecular numbers over an infinite number of realizations of the underlying stochastic process.

However, in cells, many molecules are present at very small levels so that molecular copy numbers have to be considered as small integers, not continuous quantities. For example, there are typically ten copies of a transcription factor and one or two copies of a given gene in a single cell. When the copy number of reacting molecules is low, synthesis or degradation leading to variation of the copy number by one unit induces a large relative fluctuation, and thus may have a significant influence on the system dynamics (Fig 1.12B).

Fluctuations arising from low copy numbers and inducing stochasticity are commonly encountered in many biological processes [72, 159, 150] and have attracted intensive studies since

decades [38, 105, 179, 183, 33, 151]. Recent theoretical work has demonstrated that negative feedback genetic circuits are able to reduce noise [183, 33, 47, 9], with however some fundamental limits [125]. A natural question is whether molecular noise is just a nuisance which should be controlled or whether it could have a functional role?

1.4.2 Functional roles of fluctuations

One may think that fluctuations are harmful to cellular functions, because they can degrade the quality of intra- and inter-cellular signals. However, more and more experimental and theoretical investigations support the hypothesis that molecular fluctuations play crucial roles in the design principles of cellular functions [195, 35, 5, 77]. For example, molecular fluctuations have been proposed as a mechanism responsible for phenotypic heterogeneity [157, 120]. This is expected to be particularly important for microbial cells that need to adapt efficiently to highly changing environments. In addition, fluctuations in gene expression provide mechanisms for achieving distinct physiological states in a given population, and therefore increase the probability of survival without needing genetic mutation [110, 74].

A recent work by Gagatay *et al.* [77] provides new insights into the functional role of fluctuations. Under experimental stress, a small fraction of *B.subtilis* cells transiently switch into a competence state in which they can take up extracellular DNA and incorporate it into their chromosome. This switching dynamics is mainly driven by a relatively simple circuit involving the transcription factor ComK which actives its own expression and inhibits the expression of its activator ComS (Fig 1.13A). By making use of the developed single-cell experimental technology, Gagatay *et al.* construct an alternative circuit called SynEx circuit in which ComK activating always its own expression induces the expression of MecA which is responsible for the degradation of ComK (Fig 1.13C). According to comparisons of simulations and vivo experiments, the native and synthesized SynEx circuits generate transient pulses of ComK (Fig 1.13B and D) with similar frequency, duration and amplitude.

So a fundamental question is why this particular circuit exists natively in cells rather than other potential alternatives that display the same dynamical behavior, such as the SynEx circuit. A key point is that the cellular fluctuation profiles of native and alternative SynEx circuit are very different (Fig 1.13E). Specifically, the cell-cell variability of competence duration induced by fluctuations in the copy number of ComS in the native circuit is much larger than these in the SynEx circuit. It was shown that such higher fluctuations in the native circuit provides a functional advantage which facilitates response to a wide range of extracellular DNA concentration [77].

This suggests that fluctuation dynamics is an important property of a gene circuit architecture besides its deterministic dynamics, and that both must be taken into account when discussing the design principles of gene circuits in living cells.

1.4.3 Approaches to modeling stochastic fluctuations: the master equation and its cumulant expansion

Fluctuations are generally considered as a zero-mean noise superimposed on the deterministic signal. However, genetic networks are usually highly nonlinear, which implies that a zero-mean noise can be transformed into a non zero-mean signal feedback. For example, let us assume that $\langle x \rangle = 0$. Then, the square of this signal verifies $\langle x^2 \rangle \neq 0$ unless x is constantly zero. If another

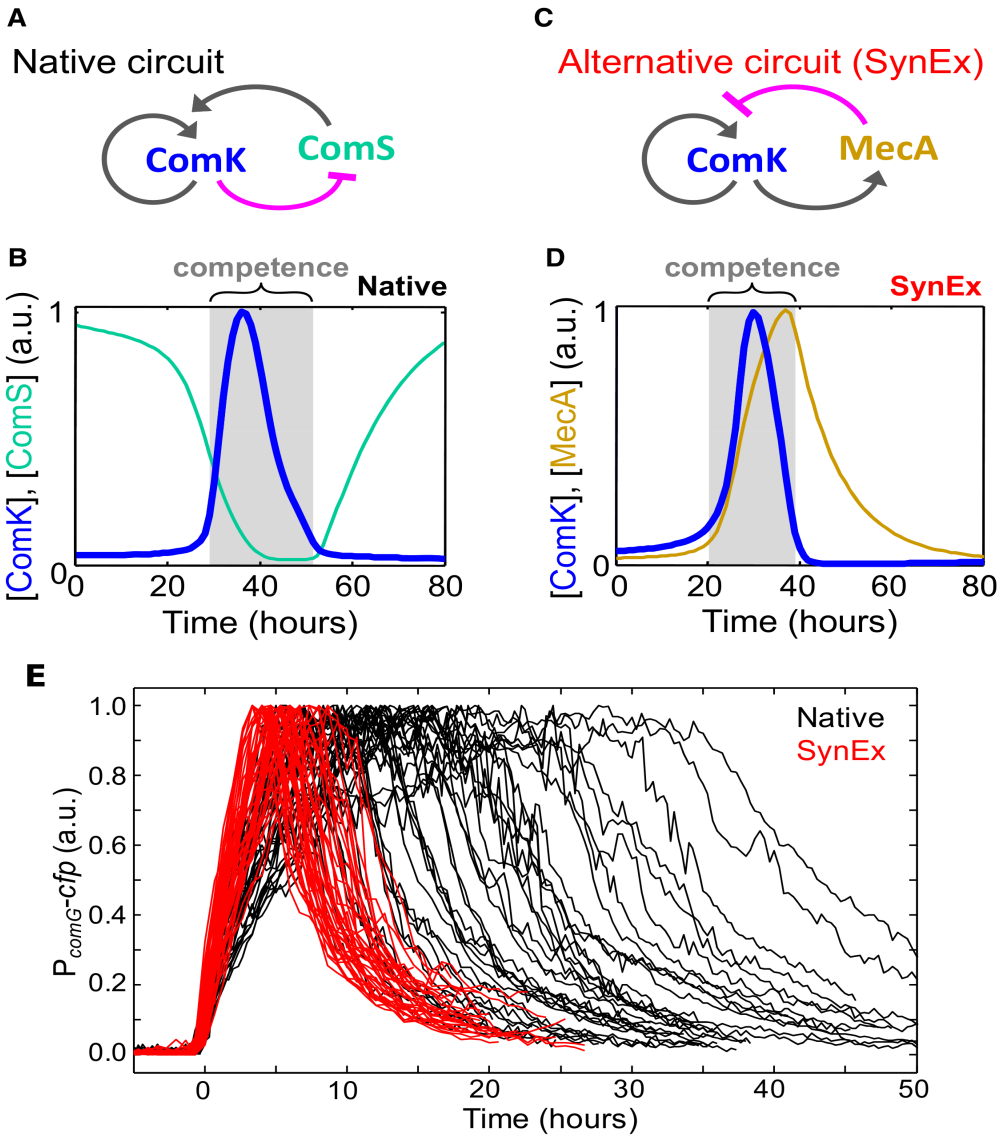

Figure 1.13: **Fluctuation dynamics discriminates functionally analogous genetic circuits.** (A) The simplified diagram of the native circuit displaying the competence dynamics. The ComK positively regulates its own expression and represses the expression of its activator ComS. (B) The native circuit generates a transient increase of ComK corresponding the competence dynamics. (C) The diagram of the alternative circuit: ComK positively regulating its own expression induces the protein MecA which is responsible for its degradation. (D) The alternative circuit generates also a competence dynamics which is equivalent to that of the native circuit. (E) The cell-cell fluctuation of competence duration is distinct between the native and the SynEx circuit. The distribution of competence duration of native circuit is much broader than that of alternative circuit. Figure taken from [77].

variable is driven by x through this square term, say $\dot{y} = f(y) + \alpha x^2$, then fluctuations of the zero-mean signal x will feed back into the average quantities whose evolution is described by the deterministic equations. It can be expected that such feedback may dramatically change the dynamical behavior. We will discuss here approaches to describe genetic networks subject to fluctuations and to characterize the influence of fluctuations on the dynamical behaviors.

The standard approach to describe a stochastic system is the so-called master equation [189]. It consists of coupled differential equations describing the time evolution of probabilities of all microscopic states characterized by the number of molecules in each species. Let $P(\{n\}, t)$ be the probability of the microscopic state $n = (n_1, n_2...)$ where n_i denotes the number of molecules in species S_i. The master equation reads as follows:

$$\frac{dP(\{n\}, t)}{dt} = \sum_{\{n'\}} W_{\{n\}\{n'\}} P(\{n^{'}\}, t) - W_{\{n'\}\{n\}} P(\{n\}, t) \tag{1.5}$$

where $W_{\{n\}\{n'\}} \geq 0$ is the transition rate from microscopic state $\{n^{'}\}$ to $\{n\}$. The master equation is a gain-loss equation for the probabilities of microscopic states. To be specific, the first term is related to the increase in the probability of being in state $\{n\}$ due to transitions from other states $\{n^{'}\}$, and the second term is related to the decrease of this probability due to transitions from $\{n\}$ to other states. The master equation predicts the temporal evolution of all molecular quantities characterizing deterministic averages and fluctuations. However, it can rarely be analytically solved [158], even for the steady states of the probability density function (PDF) $P(\{n\}, t)$ [98], and it is typically solved by numerical integration [169, 139]. Moreover, knowing the time evolution of probabilities may not suffice to characterize certain dynamical behaviors such as oscillations, since the solution of the master equation in this case will typically be a PDF centered around the limit cycle, with the phase dynamics being lost. Note that a commonly used algorithm for stochastic simulation, the Gillespie algorithm [82]), does not solve the master equation but yields a specific realization of the specific process. Evaluating the time evolution of probabilities requires many concurrent simulations of the stochastic process to obtain the ensemble average.

One approach of particular interest to solve the master equation is to construct the cumulant expansion for the PDF of the master equation. Before deriving the cumulant expansion, we define cumulants and moments. First, moments of a random variable x are defined as:

$$\mu_p = \langle x^p \rangle = \sum_{\{n\}} x^p P(\{n\}, t) \tag{1.6}$$

where p is the order of moment and the $\{n\}$ are all the microscopic states. Therefore, all moments are easily computed once the PDF is known. Moments are often used to characterize the shape of the PDF quantitatively. For example, the first-order moment is the mean; the second-order and third-order moments respectively measure variance and the skewness. More generally, a PDF is completely specified if moments of all orders are known. The moment-generating function is then defined in terms of the moments μ_n as:

$$M_x(t) = E(e^{tx}) = 1 + \mu_1 t + \frac{\mu_2 t^2}{2!} + \frac{\mu_3 t^3}{3!} + ... \frac{\mu_n t^n}{n!} + ...$$

As to the cumulants K_n of x, they are defined via the cumulant-generating function expressed as a function of the moment-generating function. More precisely, the cumulant-generating

function of x is expressed as:

$$G_x(t) = \sum_{n=1}^{\infty} K_n \frac{t^n}{n!} = \log M_x(t)$$

Therefore, cumulants are given by:

$$K_n = \frac{d^n G_x(0)}{dt^n} = \frac{d^n \log M_x(0)}{dt^n} \tag{1.7}$$

Eq. (1.7) allows us to express cumulants of any order in terms of moments. A similar principle can be applied to find joint cumulants for different random variables x, y z:

$$\begin{aligned} K_x &= \langle x \rangle \\ K_{xy} &= \Delta_{xy} = \langle xy \rangle - \langle x \rangle \langle y \rangle \\ K_{x,y,z} &= \langle xyz \rangle - \langle x \rangle \langle y \rangle \langle z \rangle - \langle x \rangle \Delta_{y,z} - \langle y \rangle \Delta_{x,z} - \langle z \rangle \Delta_{x,y}. \end{aligned} \tag{1.8}$$

The first-order and second-order joint cumulants are the average, and the covariance, respectively. Note that higher-order cumulants are expressed in terms of the lower-order cumulants. In fact, a cumulant of order n captures the n-point correlations which are not trivially induced by correlations of lower order.

An important case is that of a Gaussian distribution for which all cumulants of order 3 and above are zero. This distribution is thus completely determined by its average and covariance. In general, it may be expected that n-point cumulants of high order are negligible compared to the lowest-order cumulants and thus that only a small number of them suffices to characterize the complete PDF. The idea is then that the master equation (3.3) can be recast in a set of differential equations describing the time evolution of moments:

$$\frac{d\mu_p}{dt} = \sum_{\{n\}} x^p \frac{dP(\{n\}, t)}{dt} \tag{1.9}$$

where the time derivatives of the $P(\{n\}, t)$ are given by the master equation (3.3) and thus can be rewritten in terms of the $P(\{n\}, t)$ themselves. As will be illustrated below, the key point is that when this substitution is done, the right hand side terms of Eqs. (1.9) can be rewritten in terms of moments. In the end, we obtained equations giving the time derivatives of moments in terms of moments. Similar equations can obviously be derived for cumulants, and thus coupled equations for cumulants constitute a infinite hierarchy because of the cumulant order.

In a linear stochastic system, equations for cumulants of a given order involves only cumulants of the same and lower orders, thus the infinite hierarchy of equations for cumulants can be easily truncated. In contrast, for the nonlinear system, equations for cumulant of a given order involves higher-order cumulants. This is the most common case, as the presence of any biochemical reaction will make the system nonlinear. So a crucial question raised is how to truncate this infinite hierarchy of equations for joint cumulants. In the following, we will illustrate the cumulant expansion approach derived from the master equation in the context of a linear and a nonlinear examples. We will also try to show how fluctuations can influence dynamical behavior.

A simple linear model First, we show how to describe a simple linear model by the master equation and the cumulant expansion. This linear model consists of the following chemical reactions:

$$\begin{aligned} \emptyset &\xrightarrow{\alpha} X \\ X &\xrightarrow{X} \emptyset \end{aligned} \tag{1.10}$$

In this system, protein X is synthesized at rate α and is linearly degraded at the rate of one molecule per unity time. The synthesis and degradation are both linear. The deterministic rate equation describing this model is simply given by

$$\frac{dx}{dt} = \alpha - x \tag{1.11}$$

where x is the average copy number of X.

When the synthesis rate is comparable with or smaller than the degradation rate, the copy number of proteins X is very low and thus fluctuations have to be taken into account. The master equation reads:

$$\frac{dP(x,t)}{dt} = \alpha P(x-1,t) + (x+1)P(x+1,t) - \alpha P(x,t) - xP(x,t) \tag{1.12}$$

where $P(x,t)$ is the probability at time t of the microscopic state with x copies of proteins. Note that the evolution of $P(x,t)$ depends only on the probabilities of the two other states $x+1$ and $x-1$, as well as on the transition rate α. Even though Eq. (1.12) is linear, it is still not easy to compute the steady state of probability $P(x)$. The difficulty of finding analytical solutions limits severely the usability of the master equation. However, a numerical solution is easily obtained and is useful to verify the analysis of cumulant expansion that we will construct below (Fig 1.14).

From the definitions of cumulants and the master equation (1.12), we can write equations for the first-order and second-order cumulants associated respectively with the average and the variance [189]:

$$\begin{aligned} \frac{d\langle x\rangle}{dt} &= \alpha - \langle x\rangle \\ \frac{d\Delta_{xx}}{dt} &= \langle a_2(x)\rangle + 2\langle (x-\langle x\rangle)a_1(x)\rangle = \alpha + \langle x\rangle - 2\Delta_{xx} \end{aligned} \tag{1.13}$$

Note that the equation for the average (or first-order cumulant) is the same as the deterministic rate equation. Moreover, it is not affected by the variance. The evolution of the variance depends on the average and on itself. Equations for high-order cumulants similarly involve only cumulants of the same or lower order. Eqs. (1.13) form a closed system. The steady state can easily be obtained, with an average $\langle x\rangle = \alpha$ and a variance-to-mean ratio $\frac{\sqrt{\Delta_{xx}}}{\langle x\rangle} = 1/\sqrt{\alpha}$. Both expressions agree well with the numerical integration of the master equation (Fig 1.14).

In this part, we derived the cumulant expansion from the master equation for a linear model. We showed that equations for cumulants of a given order are uncoupled from those for higher-order cumulants, for example the average quantity is not affected by fluctuations. Two quantities that capture essential features of dynamics, average and the variance, are perfectly predicted by the cumulant expansion. We next analyze a nonlinear example using the cumulant expansion.

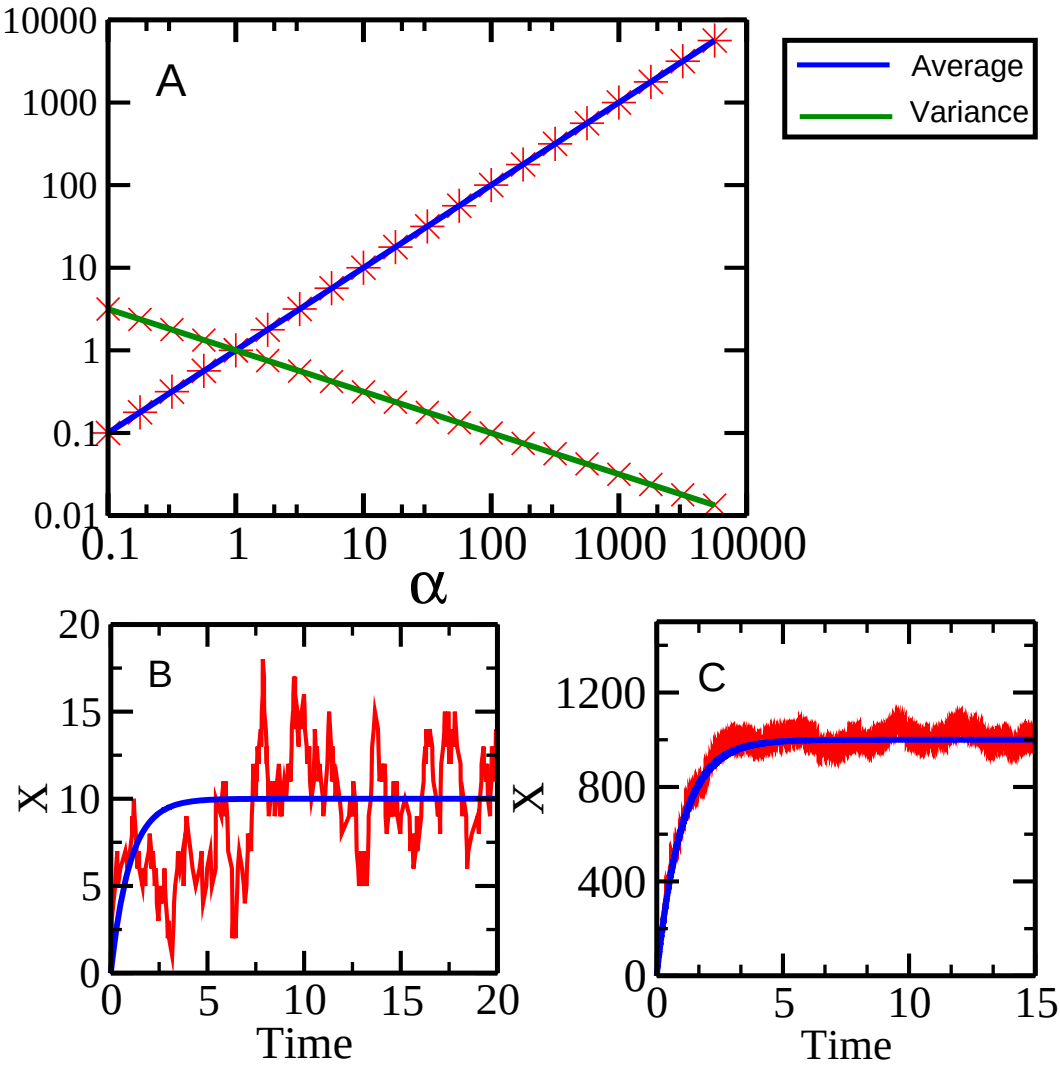

Figure 1.14: **Analysis of a linear model by using cumulant expansion.** (A) The average $\langle x \rangle$ (blue curve) and variance-to-mean ratio (green curve) $\sqrt{\Delta}/\langle x \rangle$) predicted by the cumulant expansion (1.13) are consistent with numerical simulations. (B) The time evolution of the concentration of X displays large fluctuations when the copy number is low ($\alpha = 10$). (C) The time evolution of the concentration of X is subject to small fluctuation for $\alpha = 1000$.

Nonlinear example: dimerization The nonlinear example we will consider here is a reaction network featuring a dimerization process. Dimerization is very common in biological systems and describes a process in which two identical units bind together and form a dimer [188, 7]. The governing kinetic reactions of the unit X are expressed as follows:

$$\begin{aligned} \emptyset &\xrightarrow{\alpha} X \\ X + X &\xrightarrow{X(X-1)} X_2 \end{aligned} \tag{1.14}$$

The protein X is synthesized at constant rate α and is transformed into the dimer X_2 at a rate which is proportional to the number of molecule pairs $x(x-1)/2$. In the kinetic reactions Eqs. (1.14), the factor of $1/2$ is absorbed into the rate constant which is normalized. Note that for simplicity, we neglect degradation of protein X. This dimerization model can be described by the following cumulant expansion:

$$\frac{d\langle x\rangle}{dt} = \alpha - 2\left(\langle x\rangle^2 - \langle x\rangle + \Delta_{xx}\right) \tag{1.15a}$$

$$\frac{d\Delta_{xx}}{dt} = \alpha + 4(\langle x\rangle^2 - \langle x\rangle + 2\Delta_{xx} - 2\langle x\rangle\Delta_{xx} - K_3) \tag{1.15b}$$

Equations (1.15) describe the time evolution of the average and of the variance. The factor of 2 accounts for the fact that dimerization transforms two protein molecules at a time. Compared to the deterministic rate equation $dx/dt = \alpha - 2x^2$, the differential equation for the average (1.15a) describes more precisely how the reaction rate depends on the number of molecule pairs by using the term $\langle x\rangle(\langle x\rangle - 1)$ instead of $\langle x\rangle^2$. This is particularly important in the limit case $\alpha \ll 1$ where fluctuations are significant because of the low copy number of proteins X.

A key point is that variance affects the time evolution of average and that the differential equation for the variance involves the third-order cumulant K_3. Thus, equations for the cumulant of a given order involve higher-order cumulants: all cumulants are coupled in an infinite hierarchy of equations in which Eqs. (1.15) are the first. Thus constructing a strategy to truncate this hierarchy is a key step to solve the cumulant expansion. Several approximations have been suggested [7, 134, 86]. Here we will discuss only two limit cases according to the importance of fluctuations.

When the synthesis rate $\alpha \gg 1$, the protein X is in high level so that we can assume that fluctuations described by second-order and third-order cumulants can be ignored. Under this assumption, the steady state is the same as that predicted by the deterministic rate equation: $\langle x\rangle = \sqrt{\alpha/2}$. This analytical steady state predicts well the regime with low fluctuations (Fig 1.15). However, when α approaches to unity or becomes even smaller, fluctuations are dominant and the steady state is underestimated.

In the limit case $\alpha \ll 1$, the synthesis process is so slow compared to dimerization that once two protein molecules are produced, both will be dimerized rapidly. The fluctuation is significant in this limit, as the protein copy number is either 0 or 1. The temporal protein profile consists of a sequence of spikes. The variance-to-mean ratio equals unity (Fig 1.15) which is related to the fact that protein spikes are distributed in time according to a Poisson process. On average, there are no protein molecules during a time $1/\alpha$, then 1 molecule during the same

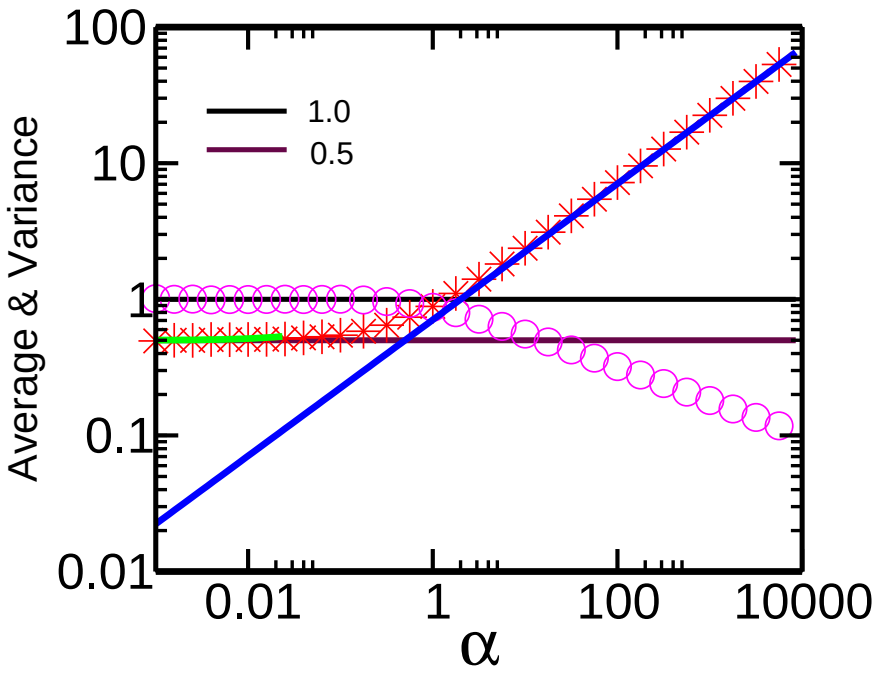

Figure 1.15: **Comparison of numerical average and fluctuation level of stochastic dimerization with theoretical predictions of the cumulant expansion**. Red stars (resp., magenta circles) represent the numerical average concentration $\langle x \rangle$ (resp., variance-to-mean ratio $\sqrt{\Delta_{xx}}/\langle x \rangle$) in the stochastic dimerization process. The variance to mean ratio is approaching to unity for $\alpha \ll 1$, which indicates that the system dynamics is driven by protein production following a Poisson process. The blue (green) curve is the steady state predicted by cumulant expansion for the limit case where $\alpha \gg 1$ (resp., $\alpha \ll 1$). Both limits agree well with numerical simulations.

time. So the average of molecules is found to be $\langle x \rangle = 1/2$ (Fig 1.15). The steady state of the second-order and third-order cumulants is also easily found: $\sqrt{\Delta_{xx}}/ < x > \simeq 1, K_3 = 3\alpha/4$ according to Eqs. (1.15). In this limit, the steady state is exactly solved and is seen to be highly affected by fluctuations.

In the dimerization problem, as well as in any nonlinear system, equations of the cumulant expansion at a given order involve higher-order cumulants. This implies that average quantities governed by deterministic rate equations are perturbed when fluctuations are taken into account, which is especially important in cell circuits. More generally, any finite set of cumulant equations does not form a a closed system, and it is therefore necessary to elaborate a strategy to truncate the hierarchy of equations.

In this section, we discussed the important role of fluctuations due to low molecular copy numbers in genetic networks. We presented a promising approach, the cumulant expansion of the master equation, which allows us to analyze the influence of fluctuations on dynamical behaviors. In chapter 3, we will apply this approach to the example of a self-repressing gene, showing how fluctuations can induce oscillations. In the next section, we will still consider the effect of stochastic fluctuations, but more specifically those arising in the highly regulated and multiple-step process of transcription [165, 44, 104, 142, 143, 121] (or, equivalently, translation [20]).

1.5 Stochastic aspect of the transcription

Transcription is a process where mRNA molecules are synthesized by macromolecules called RNA polymerases (RNAPs), which copy the gene sequence carried by DNA. In many studies, transcription is described as a simple Poisson process with a constant rate [184, 150]. It is actually a very complex process [106, 192, 76, 18, 178]. Transcription can be divided into three separate steps: initiation, elongation and termination, and each of them is exquisitely regulated [59, 14, 132, 167] by transcription factors present in small copy numbers and stochastic events. Precise single-molecule experiments have allowed us to characterize directly the dynamics of transcription by monitoring mRNA synthesis in real time, one copy at a time.

1.5.1 Transcriptional bursting directly observed in experiments

Golding and colleagues used MS2-GFP fusion proteins to tag transcripts in living individual cells of *E. Coli* when the target gene was expressed. They have directly observed individual transcription events by measuring the copy number of mRNA transcripts. They showed that transcription occurs in an unexpectedly irregular fashion rather than at constant rate.

Fig 1.16 shows the copy number of mRNA transcripts in individual cells as a function of real time. it is seen that transcription is characterized by periods of inactivity of duration Δt_{OFF} followed by periods of activity of duration Δt_{ON}. During the ON periods, transcripts are synthesized, whereas during the OFF periods, transcription remains silent. The averages of the ON and OFF durations are respectively 6 min and 37 min. This phenomenon is called transcriptional bursting and has been widely observed in many prokaryotes [85, 20, 71], in yeast [198, 181, 19] and in eukaryotes including mammalians [156, 155, 22, 42, 44, 176].

Numerous mechanisms have been suggested to explain transcriptional bursting [136, 85, 152, 31, 177, 164, 199, 187, 186, 62]. A commonly considered idea is that transcriptional

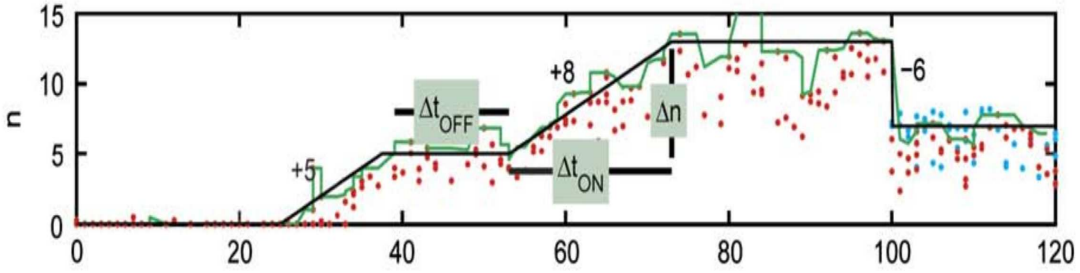

Figure 1.16: **Transcriptional bursting in individual cells**. The mRNA transcripts tagged by fluorescence proteins are counted in real time t in single cell. The experimental result directly demonstrated that the gene is transcribed during short periods of activity of duration Δt_{ON} interspersed by silent intervals of duration Δt_{OFF}. The averages of these two periods are $\langle \Delta t_{ON} \rangle \approx 6$ min and $\langle \Delta t_{OFF} \rangle \approx 37$ min. Therefore, the autonomous dynamics of gene activity has a time scale comparable to that of other important processes. Figure taken from [85].

bursting is the consequence of initiation dynamics, in particular, regarding access to promoter [85]. In this hypothesis, the silent periods are the periods where the promoter is bound by repressors, and the bursts correspond the periods of activity when the promoter is free. However, studies in vitro have shown that initiation time can be as fast as a few seconds [136], which conflicts the hypothesis. Meanwhile, as typical genes possess easily thousands or millions of nucleotides that need to be transcribed by RNAPs and the elongation process is subject to complex regulations, people begin recently to investigate the elongation as a developmental checkpoint. Elongation is actually not a constant forward process. RNAPs read the template of DNA in a stochastic stepwise fashion, and moreover they can display some unusual dynamical behaviors which enhance the stochasticity, such as transcriptional pausing [126] which is one mechanism possibly explaining the transcriptional bursting [31, 62, 67].

1.5.2 Pausing of RNA polymerases during transcription

Unusual dynamical behaviors of RNAPs [182, 43, 163, 25, 97] may seriously affect the transcription dynamics. During elongation in which RNAPs move forwards along the DNA template in stochastic single-nucleotide steps [190], the movement of RNAP in vivo and in vitro is usually interrupted by an event where the RNAP is halted at a nucleotide. This behavior is defined as transcriptional pausing [26, 118, 126, 108]. Single-molecule assays have recently achieved base-pair resolution for following motion of RNAPs and allow us to directly visualize the RNAP pausing [57, 54, 56] (Fig 1.17). Transcriptional pausing has been widely found in both prokaryotes and eukaryotes [10, 70, 118].

RNAP pausing occurs in a stochastic fashion and is spontaneously reversible, after which the RNAP continues to move forwards. The majority of pauses are short with an average duration of ≈ 1 s. They are referred to as "ubiquitous" or "elemental" short pauses [56, 107, 119, 55, 57] and must be distinguished from the so-called prolonged backtracking pauses [55, 57, 4]. The latter are rare (represent less than $5\% - 10\%$ of all pauses) and their durations are usually over 20 s. In fact, prolonged pauses are involved in the backtracking process [197, 171, 78, 48, 113, 49, 15, 50, 192] in which RNAPs move backwards and may proofread the

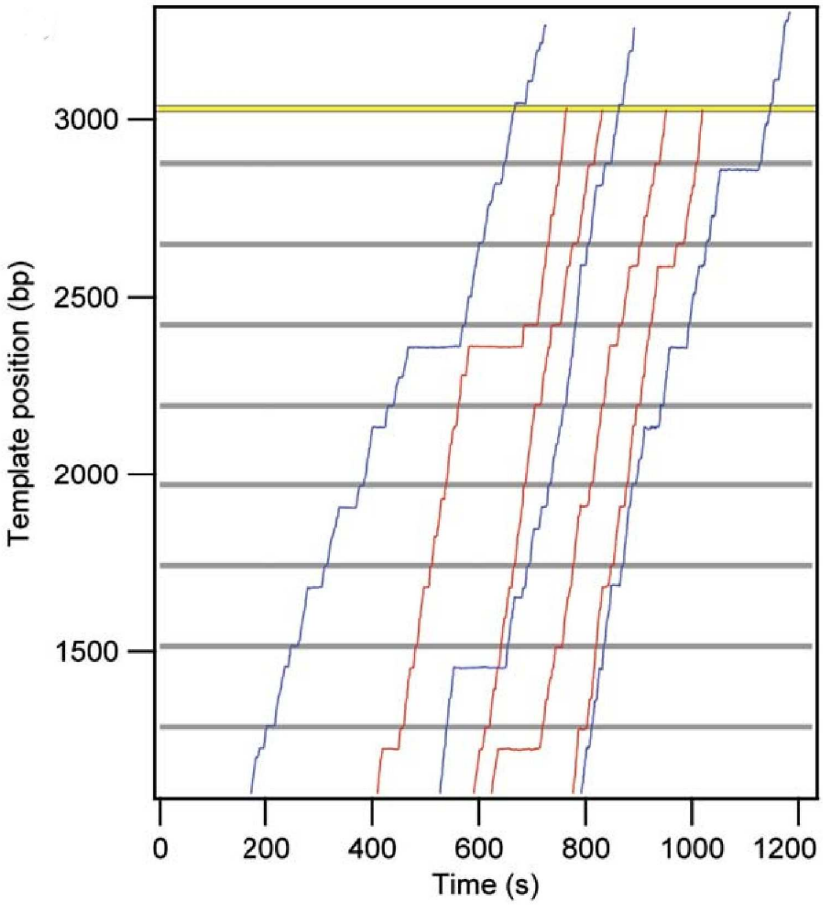

Figure 1.17: **RNAP pausing observed in single-molecule transcription on synthetic template**. Advanced single-cell experiments allow people to monitor the dynamics of single RNAP molecule in real time. Here seven representative records of positions of single RNAP molecule along the ≈ 3 kb *ops* repeat templates as a function of time are shown. It is found that in most records, RNAP is halted at some sites for a small or large amount of time. Figure taken from [57].

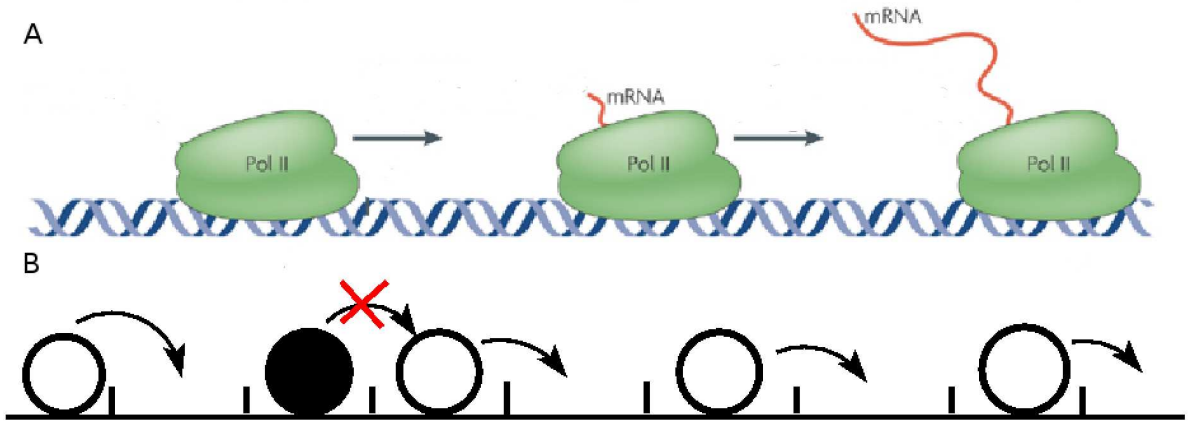

Figure 1.18: **Transcription and TASEP model**. TASEP model is used to describe the transcription process. (A) In transcription, RNAPs cannot overlap with each other and move forwards along DNA template in stochastic single-nucleotide steps. (B) In the TASEP model, each site can only accommodate at most one particle which hop down with a constant probability related to the elongation rate only if the next site is not occupied.

sequences of copied mRNA [63]. In addition, it is observed that short pauses depend little on template sequences. They are not affected by mechanical pushing forces exerted by trailing RNAPs, so that the pause frequency and duration are the same even for a large density of RNAPs. In contrast, prolonged backtracking pauses exhibit a strong sequence dependence, and a RNAP in prolonged pause can be pushed forwards by a trailing RNAP [56, 171, 65].

Recent investigations show that transcriptional pausing plays important functional roles in coordination and accuracy of gene expression [51, 45, 1, 61]. Intuitively, transcriptional pausing leads to traffic jams of RNAPs. Not only the average transcription rate is seriously reduced, but also the temporal dynamics of transcription is dramatically affected. When small groups of RNAPs induced by traffic jams arrive at the termination site, they produce a group of mRNA transcripts, which can correspond to a burst of transcription we have mentioned before.

As the transcription is a complex process, the question is how to describe mathematically the transcriptional pausing so as to study its influence. To this aim, we first introduce a classical system in out-of-equilibrium statistical physics, which has been proposed to model transcription.

1.5.3 TASEP (Totally Asymmetric Simple Exclusion Processes) as the model of transcription

During transcription, there are two main features in RNAP dynamical behavior. They move forwards along DNA template in a stochastic fashion, and they can not overlap with each other so that they are constrained by the "excluded volume constraint". Several studies have suggested to model transcription [62, 164, 31, 109] (as well as translation [23, 17, 161, 172, 24]) as a driven lattice gas in one dimension (Fig 1.18), known as the Totally Asymmetric Simple Exclusion Process in out-of-equilibrium statistical physics (TASEP for short) [115, 112, 29, 28, 34, 30, 117, 27, 153].

In this TASEP model, RNAPs are modeled as particles moving in one direction along a discrete lattice comprising N sites. For the sake of simplicity, we will not take into account RNAP footprint, which does not qualitatively change the results. Particles hop down with a constant probability associated with elongation rate, only if the next site is not occupied by another particle, so that each site can accommodate at most one particle (the excluded volume constraint is implemented). The dynamics of TASEP model highly depends on the boundary conditions. TASEP models with open boundary conditions [29], where a new particle can enter the lattice with rate α representing the transcription initiation and particles can also exit from the end of lattice with rate β for the transcription termination, describe naturally the transcription process. The analysis of TASEP with open boundary condition is relatively involved and has been an achievement of statistical physics.

For mathematical simplicity, transcription can also be modeled as a TASEP model with periodic boundary conditions, where the lattice is a ring around which particles move indefinitely [28, 172]. The number of particles is thus constant. Also, all sites are equivalent and the density is the same in each site by symmetry.

The central quantity of interest in the TASEP model is the particle current, which is the average number of particles going through one site per unit time. This current may display large variations depending on the values of parameters because of the possible formation of "traffic jams" blocking the flow across the lattice and thereby dramatically lowering the current. From a statistical point of view, the particle current is a representative of all transport phenomena between two bodies which are not in thermal equilibrium. As in any transport phenomenon, there is a conservation law relating the variation of density at one site to the difference between the incoming and outgoing currents. When steady state has been reached, densities at all sites are constant, which implies that the current is the same across the entire lattice.

In the TASEP model with periodic boundary condition, the current, which measures the mean number of particles advancing from a given site during unit time and represents the transcription rate, is given by:

$$J = \rho(1 - \rho) \tag{1.16}$$

where ρ is the average density of RNAPs at each site. Remarkably this expression can be obtained from a simple mean-field approach. Formula (1.16) expresses the fact that the probability of a particle advancing from one site depends both on the presence of a particle on the site (with probability ρ) and on the absence of particle on the next site (with probability $1-\rho$). Note that the transcription rate is a symmetric function of density. Low particle densities lead to small transcription rate, which is not surprising. For high densities of particles, only particles with an empty site in front of them can move forwards and contribute to the current, so that the transcription rate is reduced as well. The current reaches its maximum value of $J_{\max} = 1/4$ when the ring system is half occupied.

Although TASEP with periodic condition, where particles travel on a ring, has little resemblance with the transcription process, it is sufficiently simple to allow analytical studies. Some analytical results can be derived and provide us crucial insights, such as the current-density relation, into the transcription dynamics influenced by RNAPs.

Here we have discussed a typical feature of transcription, RNAP pausing, as a possible explanation for transcriptional bursting. The TASEP model, as one classical system in out-equilibrium statistical physics, has then be presented as a model of transcription. We will

analyze it further in chapter 4 to investigate dynamical effects of RNAP pausing on transcription.

1.6 Conclusion and aims of this thesis

In this chapter, we first introduced the intriguing oscillatory behaviors displayed in many biological systems, which motivated us to begin this thesis work. Then we discussed briefly elemental biochemical processes occurring in the cell and how major functional molecules, such as DNA, mRNA and proteins, interact with each other via biochemical reactions to form genetic networks which can display oscillations. Next, we showed how to describe genetic networks by deterministic modeling and discussed three key ingredients for oscillations: negative feedback, delay and nonlinear degradation. Stochastic fluctuations due to low molecular copy numbers are then considered as well as their influence on genetic networks. An approach combining the master equation and a cumulant expansion were presented as a method to describe fluctuations and to explore their influence on oscillatory behavior. Finally we focused on the stochastic properties of transcription which have attracted much interest during the past decade. In particular, transcriptional pausing was introduced. It may highly affect the dynamics of transcription and contributes to transcriptional bursting.

This thesis mainly focuses on dynamical effects of transcription. As the title suggests, it consists of three different studies devoted to (i) the combination of various and multiple time delays in a self-repressing gene circuit and its influences on oscillations, (ii) dynamical effects of stochastic fluctuation on the oscillation and (iii) the dynamical influences of RNAP pausing on transcriptional dynamics.

In the first part of this work, we study theoretically the combination of various delays in a small gene circuit as well as the dynamical influence of different types of delays in the context of a deterministic modeling of genetic networks. Time delays have been highlighted as an important ingredient for sustained oscillations. Here we consider in particular time delays arising from transcription. Transcription is usually considered as an instantaneous process [137, 100, 127]. However, we have presented in Section 1.5.1 experimental results showing that gene activity has an autonomous dynamics with a time scale comparable to that of other important processes. Thus we will recall in Section 2.1.1 a model proposed by Morant *et al.* [138] taking into account a transcriptional dynamics of gene activity with a finite gene response time. In this model, it can be analytically shown that there is a time scale of the gene response at which the system is most destabilized. Also, the trade-off between the delay and saturation of degradation can be studied precisely. However, there are various sources of delay in genetic networks, such as translation, molecule transport, etc. as we mentioned in Section 1.3.2.

A natural question is whether the results of [138] remain valid when other sources of delay are taken into account. In Chapter 2, we will study extensions of the model of [138] where a second delay derived from protein transport is added. Another question is raised: how do delays from two different sources interact, does their combined influence only depend on their sum or not? Is it correct to hypothesize that the more delays there are, the more easily oscillations are obtained is correct? In Section 1.3.2, we had introduced two types of delay modeling: explicit delay or reaction delay in which the delay is modeled as reaction step. The following question is whether the two different types of delay exert the same influence on the appearance of oscillations. If no, how do they differ? Under which conditions do they play the same role

in dynamics? We will answer all these questions in Chapter 2 by deriving analytical expression of the oscillation threshold in models with explicit or reaction delays.

The second part of this thesis work is motivated by the observation that the dynamical behaviors predicted by deterministic models, such as steady state levels and regions of oscillation in parameter space, can be dramatically modified by fluctuations. In order to elucidate the influence of fluctuations on the dynamics of genetic networks, we use the minimal genetic circuit comprising one self-repressing gene with linear degradation of mRNA and protein. In this reduced system, sources of fluctuations are not only the low copy numbers of mRNA and protein molecules, but also the dynamics of gene activity which is characterized by transitions between "ON" and "OFF" states. The question is then whether the cumulant expansion can be applied to describe the binary gene states. If so, how do fluctuations change the steady state predicted by deterministic model? Moreover, it is known that there is no deterministic oscillations in this small genetic circuit when degradation mechanisms are linear [138]. Are fluctuations then able to induce oscillations ? Another question raised is how to characterize stochastic oscillation in numerical simulations to verify our analytical results. All these questions will be discussed in Chapter 3.

The third part studies dynamical effects of RNAP pausing on transcription. We introduced In Section 1.5.3 the classical TASEP model which is used to model the transcription process, as well as pausing which increases the stochasticity of transcription in Section 1.5.2. Thus the question is how to incorporate pausing into the TASEP system, and how to analyze its dynamical effects. The TASEP model is usually studied under the mean-field approximation [23, 17, 161]. Does the mean-field approximation always work in the model with pausing ? If not, how can we study this model ? With pausing, what is the typical behavior of transcription and how does pausing drive the transcription rate ? We will study these questions in Chapter 4.

Chapter 2

Oscillation arising from combination of various and multiple time delays with nonlinearity in a self-repressing gene

Many studies have highlighted the importance of oscillations in various biological systems, such as circadian rhythms [53, 75, 160, 40, 3, 32], cell division cycle [129], immune response [66], cell growth/death [93] and embryo development [2, 60]. The physiological role of biological oscillations may stem from the fact that they encode more information than steady states, allowing them to contribute efficiently to robust biological regulations of cellular functions at different levels. Oscillations also allow cells to respond flexibly to variations of their environments [21, 148]. In order to meet the growing need of understanding the mechanisms underlying these oscillations, intensive work has been carried out [174, 88, 184, 137, 100, 127, 135, 123, 11, 180, 133, 175, 80, 145]. It was found that negative feedback of molecular signals is a necessary ingredient to support sustained oscillations displayed in genetic networks. Secondly, these molecular signals should be sufficiently delayed so that the steady state of genetic networks is destabilized. Additionally, highly nonlinear molecule degradation mechanisms may also trigger oscillations. The subtle relation between delay and nonlinear degradation in the design principle of biological oscillators has been often mentioned but is not yet fully understood [135].

There are in fact various sources of time delay, in particular originating in transcription [85, 155, 22], translation, phosphorylation, molecule transport, etc. The combination of these various and multiple time delays may be key to explain the robustness of experimentally observed oscillations [175], however how these delays interact remains unclear. In order to address this question, mathematical modeling of delays is required. There are actually two types of delay used in modeling. In some studies, the delay appears in an explicit manner, as the time-delayed value of some dynamical variables [137, 127, 184], and the underlying processes are not specified. We shall thereafter refer to such a delay as an "explicit delay". In other studies, the delay originates in a reaction step [123, 80], and we shall call such a delay a "reaction delay". Although explicit and reaction delays are commonly used to model biological oscillations, the similarities and differences of these two types of delay modeling have not yet studied systematically.

The purpose of chapter 2 is to investigate how the combination of various and multiple time delays affects oscillations and to compare the influences of explicit and reaction delays on

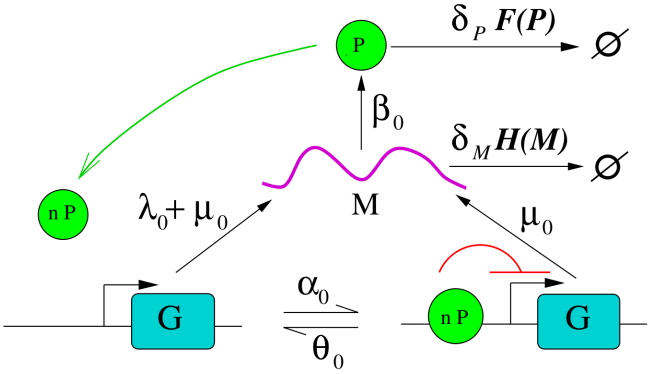

Figure 2.1: **Reaction diagram of the self-regulating gene circuit**. Transcription and translation are involved. The autonomous dynamics of gene is approximated by the protein-DNA binding-unbinding. Transcription rate of bound gene by proteins is perfectly repressed. Arbitrary degradation functions are chosen for mRNA and protein.

oscillatory behavior of genetic networks. To this aim, we consider the minimal genetic network comprising a self-repressing gene [12, 127, 100, 137], which is a very common element of biological networks. For example, in *E. Coli* cells, 40% of genes for transcription factors are repressed by their own protein products [162]. This chapter is organized as follows: in the first section, we review a mathematical model for the self-repressing gene that takes into account the autonomous dynamics of gene activity [138]. Then we extend that model to combine the gene response time with transport delays modeled as reaction delays or as explicit delays. These two cases will be discussed respectively in the second and third sections. In the last section, we discuss similarities and differences in the dynamical behaviors induced by the two types of delay.

2.1 Description of time delay in the model of a self-repressing gene with transcription memory

2.1.1 Basic model accounting for the slow dynamics of gene activity

We first review the model proposed by Morant *et al.* [138]. It describes the minimal genetic circuit consisting of a self-repressing gene (Fig 2.1), and takes into account the fact that a slow transcriptional dynamics can affect the system behavior. Besides transcriptional regulation, the synthesis and degradation of proteins and messager RNAs are the basic biochemical processes of the model. For the sake of simplicity, we only consider the case where transcription rate is reduced by the binding of a protein monomer.

Except for the degradation mechanisms which will be left unspecified, we apply the law of mass action according to which the kinetic rate is proportional to the number of reacting molecules. The kinetic equations of the genetic circuit represented in Fig 2.1 are then the

following:

$$\frac{dG}{dT} = \theta_0(1-G) - \alpha_0 PG \tag{2.1a}$$
$$\frac{dP}{dT} = \frac{dG}{dt} + \beta_0 M - \delta_P F(P) \tag{2.1b}$$
$$\frac{dM}{dT} = \mu_0 + \lambda_0 G - \delta_M H(M) \tag{2.1c}$$

G, P and M represent respectively gene activity, protein and RNA copy numbers. Eq. (2.1a) describes the kinetics of protein-DNA binding at rate α_0 and unbinding at rate θ_0. The first term of Eq. (2.1b) corresponds to protein-DNA binding and unbinding, the second and third terms describe translation and protein degradation. Eq. (2.1c) describes the total transcription rate $\mu_0 + \lambda_0 G$ (as the unbound gene is expressed as G with the transcription rate $\lambda_0 + \mu_0$ and the bound gene is $1 - G$ with the transcription rate μ_0) and RNA degradation. Eq. (2.1a) can also be viewed as a minimal description of the dynamics of an effective gene activity G slowly relaxing towards an equilibrium value given by the gene regulation function $G = 1/(1 + P/P_0)$, with $P_0 = \theta_0/\alpha_0$ the half-expression threshold. Such a model can be viewed as the leading approximation of a mechanistic model of transcription including all processes concurring to gene expression when there is a dominant limiting step. The negative feedback in this system arises from the fact that the activity G, or the transcription rate is reduced when the DNA is bound by a protein. In order to study how oscillations can be induced by tuning protein and RNA degradation, we consider arbitrary degradation functions $H(M)$ and $F(P)$, without specifying detailed mechanisms. These degradation functions are assumed to have unit derivative at zero, so that a deviation of their slope from 1 will characterize the nonlinearity in degradation.

Note that a common interpretation of the circuit in Fig 2.1 is that the gene has two states, bound and unbound. Our deterministic model is valid when the mean response time of gene is small compared to the oscillation period, so that there are many binding/unbinding events by cycle. G can then be viewed as a temporal average of gene activity.

By renormalizing time, variables and parameters (more details in the Appendix A.1), Eq. (2.1) can be rewritten in dimensionless form :

$$\frac{dg}{dt} = \theta\,(1 - g(1+p)) \tag{2.2a}$$
$$\frac{dp}{dt} = \alpha\,(1 - g(1+p)) + \delta(m - f(p)) \tag{2.2b}$$
$$\frac{dm}{dt} = \mu + \lambda\, g - h(m) \tag{2.2c}$$

Eqs. (2.2) have a single steady state (g_*, p_*, m_*). Note that the steady state depends only on parameters λ and μ as well as on functions f and h, whereas parameters θ, α, δ control time scales. Therefore, tuning gene response time does not change the steady state. The behavior of the degradation in the neighborhood of the steady state is described by the slopes:

$$s = \frac{df(p)}{dp}\Big|_{p=p_*}, u = \frac{dh(m)}{dm}\Big|_{m=m_*} \tag{2.3}$$

In the case of linear degradation, we have $u = s = 1$. Small or even negative values of the slopes s and u generally denote strongly nonlinear degradation mechanisms, including saturation.

To assess whether system (2.2) can display sustained oscillations, we have searched for parameter values where the fixed point loses its stability to a periodic solution via the Hopf bifurcation. For the sake of simplicity, we assume here perfect repression when the gene is bound by a protein ($\mu = 0$) and a large threshold ($P_0 \gg 1$) leading to $\alpha \sim 0$. Under this approximation, the Routh-Hurwitz stability criterion [89] indicates that the Hopf bifurcation occurs when:

$$\mathcal{H}_\epsilon(\Sigma, T) = \Sigma \left(\frac{\epsilon^2 \Sigma^2}{4} T^2 + (\Sigma - \frac{1}{\Sigma})T + 1 \right) < 0 \tag{2.4}$$

where $\epsilon = \frac{2\sqrt{\delta s u}}{\delta s + u} \in [0, 1]$ indicates whether the protein degradation rate δs and the RNA degradation rate u are balanced. When these two rates are equal, $\epsilon = 1$. ϵ becomes zero as one or the other rate vanishes. $\Sigma = \delta s + u$ represents the sum of degradation rates; $T = g_*^2 \sqrt{\delta \lambda}/\theta$ characterizes the time scale with which gene responds to a sudden variation in protein copy number.

The criterion 2.4 allows one to identify key ingredients for oscillations and to assess their influence quantitatively. In particular it shows how delay, characterized by T and nonlinear degradation, characterized by Σ and ϵ, combine to generate oscillations. When the delay is small, the degradation needs to be strongly saturated ($\Sigma \to 0$). When degradation is closer to a linear mechanism, then the delay needs to be larger. In this sense, a saturated degradation can be viewed as equivalent to a time delay.

Eq. 2.4 defines a series of curves $\Sigma_\epsilon(T)$ specifying the degradation rate Σ at oscillation threshold as a function of response time T and balance index ϵ. For a given ϵ, oscillations are found for $\Sigma < \Sigma_\epsilon(T)$. For fixed T, $\Sigma_\epsilon(T)$ decreases monotonously with ϵ. Fig 2.2 shows the limit curves $\Sigma_1(T)$ and $\Sigma_0(T)$ which are particularly important to understand the bifurcation diagram.

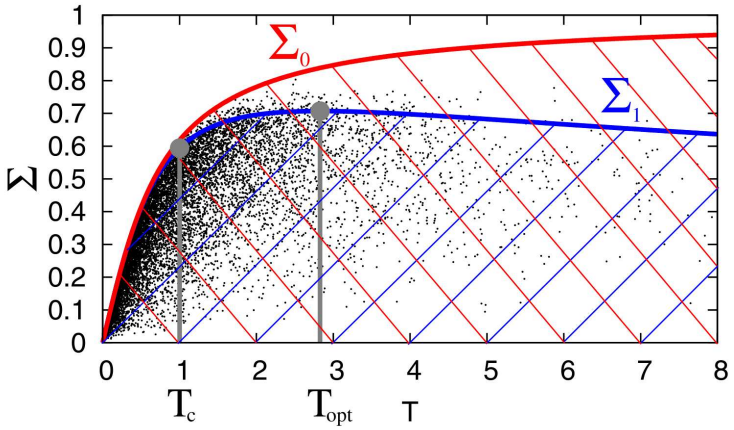

Figure 2.2: **Bifurcation diagram of the model (2.2) in the (Σ, T) plane**. The system with balance index ϵ oscillates for $\Sigma < \Sigma_\epsilon(T)$. The limit curves $\Sigma_1(T)$ and $\Sigma_0(T)$ are shown. Black dots correspond to oscillating parameter sets of (2.2) with specific degradation functions. $T_c = 1$ denotes the time scale beyond which transcriptional dynamics cannot be neglected. Figure taken from [138].

Two regions can be distinguished. For $T < 1$, the instability threshold $\Sigma_\epsilon(T)$ increases rapidly with T and its dependence on ϵ is negligible. In the region of $T > 1$, $\Sigma_\epsilon(T)$ reaches its

maximum value and then decreases for larger values of T except in the case $\epsilon = 0$.

This analysis indicates that there is a resonance-like phenomenon in the dynamics of a self-repressing gene with slow transcription response: this minimal genetic circuit bifurcates most easily to periodic behavior, or more generally is most instable, at a finite value of the gene relaxation time given by $T_{opt} = 2\sqrt{2}$.

However, there are other various sources of delay besides the gene response time, such as those arising in translation, molecule transport, etc. A natural question is whether the conclusions of Ref. [138] remain valid when other various delays are taken into account. Another question is how to describle mathematically these delays. In the next section, we give examples of various sources of delay and introduce two types of delay modeling which are commonly used in the literature: reaction delays or explicit delays.

2.1.2 Reaction and explicit delays

The work of Morant *et al.* [138] analytically demonstrates that delay is a key ingredient for oscillatory behavior. However, besides delays due to slow transcriptional response, various sources of delay are found in biological systems, even in such a minimal genetic network comprising a self-repressing gene. Translation takes a minimum amount of time. There are many intermediate steps giving rise to delays, such as molecular transport of mRNAs and proteins between nucleus and cytoplasm, phosphorylation, etc. (Fig 2.3). These delays are probably comparable to or even larger than the gene response time.

This raises thus several questions: do delays from other sources promote or weaken oscillations induced by the slow gene activity? Can these delays be combined into a global delay that promotes oscillatory behavior? If yes, how ? A further question is whether delays from different sources play similar roles or not.

To answer these question, a mathematical modeling of delay is required. However, two different strategies of modeling delays exist in the literature: explicit and reaction delays, as mentioned in Section 1.3.2. Explicit delays allow us to ignore details of underlying mechanisms, but bring up delayed differential equations (DDEs), which are difficult to solve analytically, while reaction delays allow us to specify process details giving birth to delay and are easily analyzed. When a delay has to be introduced in a model, researchers most often seem to choose between those two types of delay according to personal preferences, without discussing their different influence on oscillatory behavior. Thus, we may ask whether the influence of one type of delay is stronger than that of the other. If it is so, which one does promote oscillations more easily ? What are the conditions under which they will play similar roles for triggering oscillations ?

2.1.3 Extended models incorporating molecular transport delays

In order to answer these questions, we extend the self-repressing gene model studied by Morant *et al.* [138] in which there is already a delay modeled as reaction delay: the gene response time. We will combine this delay with an additional delay, here induced by protein transport between nucleus and cytoplasm. Such kind of delay has been for example considered in Ref. [123].

In this extended models, transport delay will be modeled as a reaction delay and as an explicit delay in turn. For the sake of simplicity, we still assume perfect repression ($\mu = 0$) and

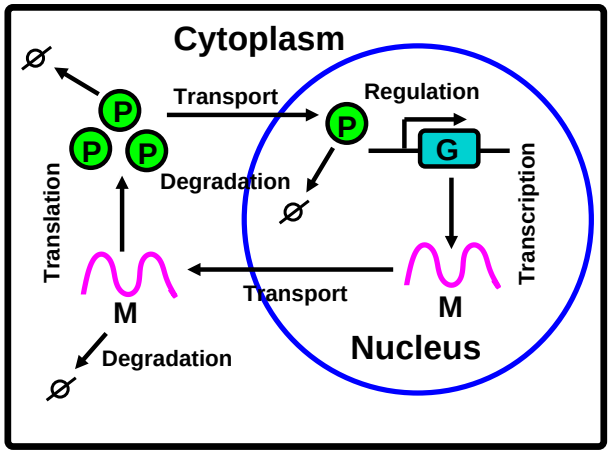

Figure 2.3: **Various sources of delay**. Transcription and translation take a minimum amount of time. Many intermediate steps lead to delays, such as the transport of mRNA and protein between nucleus and cytoplasm.

a large half-repression threshold ($\alpha \sim 0$). Numerical simulations indicate that the final results we will obtain are relevant even when these conditions are not met.

We first assume that protein transport from cytoplasm to nucleus is irreversible and model it as a reaction delay. Then the governing kinetic equations in dimensionless form are:

$$\frac{dg}{dt} = \theta\,(1 - g(1 + p_n)) \tag{2.5a}$$

$$\frac{dm}{dt} = \lambda\, g - h(m) \tag{2.5b}$$

$$\frac{dp_c}{dt} = \delta\, m - \frac{p_c}{\tau_p} \tag{2.5c}$$

$$\frac{dp_n}{dt} = \frac{p_c}{\tau_p} - \delta\, f_n(p_n) \tag{2.5d}$$

where g, m, p_c, p_n are respectively gene activity, RNA copy number, protein copy numbers in cytoplasm and nucleus. We assume single-protein regulation for simplicity, with DNA being bound by nuclear protein (2.5a). $h(m)$ ($f_n(p_n)$) is an arbitrary function describing the kinetics of RNA degradation (protein degradation in the nucleus) which can be linear or nonlinear. λ and δ characterize the maximum transcription rate and the protein degradation rate; θ denotes the rapidity of gene response to a sudden variation of protein copy number. Since Eqs. (2.5) are extended from Eqs. (2.2) by adding the transport of proteins, the same degradation parameter δ is kept in (2.5c) and (2.5d) as in the Eq. (2.2b). The term $\frac{p_c}{\tau_p}$ represents the protein transport from cytoplasm to nucleus, with τ_p the protein transport time. Note that such a reaction delay could also describes delays arising from intermediate steps of several biological processes, such as protein folding, phosphorylation, etc. Here we suppose that the protein degradation occurs only in the nucleus so that the steady state of system does not depend on the transport delay, which much simplifies the following computation.

Protein transport can be irreversible, if it results from an active process, or reversible if it results from mere diffusion [123]. To include reverse transport, kinetic equations are modified

as follows:

$$\frac{dg}{dt} = \theta\ (1 - g(1 + p_n)) \tag{2.6a}$$
$$\frac{dm}{dt} = \lambda\, g - h(m) \tag{2.6b}$$
$$\frac{dp_c}{dt} = \delta\, m - \frac{p_c}{\tau_p} + \frac{p_n}{\tau_r} \tag{2.6c}$$
$$\frac{dp_n}{dt} = \frac{p_c}{\tau_p} - \frac{p_n}{\tau_r} - \delta\, f_n(p_n) \tag{2.6d}$$

where g, m, p_c, p_n are respectively gene activity, RNA copy number, protein copy number in the cytoplasm and in the nucleus; the term $\frac{p_c}{\tau_p}$ represents the protein transport from cytoplasm to nucleus; similarly, the term $\frac{p_n}{\tau_r}$ is the reverse transport of protein from nucleus to cytoplasm, and τ_r characterizes the reverse delay.

This delay due to protein transport can also modeled as explicit delays. Explicit delays in the form of $x(t-\tau)$ where $x(t)$ is a dynamical variable and τ is a fixed delay are widely used to describe the dynamics of genetic networks. They are generally used to take into account biological processes which take a minimum amount of time, such as transcription and translation, etc. It is clearly also admissible to use such terms to describe a delay induced by molecular transport. In this case, the kinetic equations become:

$$\frac{dg}{dt} = \theta\ (1 - g(1 + p(t - \tau_1))) \tag{2.7a}$$
$$\frac{dm}{dt} = \lambda\, g - h(m) \tag{2.7b}$$
$$\frac{dp}{dt} = \delta\, (m(t - \tau_2) - f(p)) \tag{2.7c}$$

where θ is still the parameter characterizing gene response time, and τ_1 and τ_2 represent respectively the protein transport time from cytoplasm to nucleus and RNA transport time from nucleus to cytoplasm.

In the next section, we will analyze the above extended models with reaction and explicit delays taking into account molecular transport. We will try to uncover the principles governing the combination of these delays with the gene response time and to understand the differences in the influences of a reaction delay and of an explicit delay on oscillatory dynamics. Note that the model extended with a reaction transport delay has one more variable than the original model (4 variables in Eqs. 2.5 vs. 3 variable in model 2.1). In contrast with this, the model with the explicit transport delay has the same dimension as the original one, however the ordinary differential equations (ODEs) become delay differential equations (DDEs).

2.2 Analysis of the model with a reaction delay

2.2.1 Analytical criterion of oscillations

First of all, we study the model 2.5 with a reaction delay describing the irreversible transport of protein from cytoplasm to nucleus. Eqs. (2.5) have a single steady state (g_*, m_*, p_{n*}) which

does not depend on the parameter τ_p and θ characterizing respectively the time scale of protein transport and gene response. In other words, varying τ_p and θ does not change the fixed point (g_*, m_*, p_{n*}), which is an important property that it greatly facilitates our analysis. The behavior of the degradation mechanisms in the neighborhood of the steady state is still denoted by the slopes:

$$s = \frac{df(p_n)}{dp_n}\Big|_{p_n=p_{n*}}, u = \frac{dh(m)}{dm}\Big|_{m=m_*} \tag{2.8}$$

As the steady state does not depend on the τ_p and θ, the slopes s and u does not depend on them either. In the case where degradation functions of protein and RNA are linear, we have $s = 1, u = 1$. Values of s and u differing from 1 generally denote strongly nonlinear degradation mechanisms, including saturation.

To assess whether the system (2.5) can display sustained oscillations, we search for parameter values where the fixed point losses its stability to a periodic solution via a Hopf bifurcation. The analytical Routh-Hurwitz stability criterion [89] indicates the Hopf bifurcation occurs when:

$$\begin{aligned}\mathcal{H} = (\delta\, s + u)(\tau_g + \tau_p)(\delta\, s\tau_g + 1)(u\tau_g + 1)(\delta\, s\tau_p + 1)(u\tau_p + 1) \\ -g_*^2\delta\lambda[(\tau_g + \tau_p) + (\delta\, s + u)\tau_g\tau_p]^2 < 0\end{aligned} \tag{2.9}$$

where $\tau_g = g_*/\theta$ denotes the gene response time. When the criterion (2.9) crosses zero to become negative, sustained oscillations are observed. In the limit of $\tau_p = 0$, we recover the criterion given by Morant *et al.* [138]. Note that the effect of the transport delay is exactly the same as that of gene response time, as shows the symmetry between τ_p and τ_g in the criterion (2.9). Therefore, we can conclude that these two reaction delays from different sources play completely identical roles for the appearance of oscillations. This is a remarkable result because it allows us to extend our results for the self-repressing gene with finite response time to the case of a self-repressing gene with finite transport time, which has been much more studied.

Then, how do these two reaction delays combine ? To answer this, we represent the criterion as a function of the sum of delays $\tau = \tau_g + \tau_p$ and of their product $\tau'^2 = \tau_g\,\tau_p$, which are both symmetric functions. The degradation rate of mRNA and protein $\delta\, s$ and u are also expressed in their sum $\sigma = \delta\, s + u$ and product $\gamma = \delta\, su$ (see Section 2.1.1).

$$\mathcal{H} = \sigma + (\sigma^2 - g_*^2\delta\,\lambda)\,\tau + \sigma\,\gamma\,\tau^2 + \sigma\,\tau'^2(\frac{\gamma^2\,\tau\,\tau'^2 - \sigma - 2\gamma\,\tau}{\sigma\,\tau'^2 + \tau} - g_*^2\delta\,\lambda) < 0 \tag{2.10}$$

We then renormalize criterion (2.10) as follows:

$$\sigma_c = g_*\sqrt{\delta\,\lambda},\ \sigma = \sigma_c\,\Sigma,\ \gamma = \frac{\epsilon^2\,\Sigma^2\,\sigma_c^2}{4},\ \tau = \frac{T}{\sigma_c},\ \tau'^2 = \frac{\eta\,T}{2\sigma_c} \tag{2.11}$$

where $\eta = 2\sqrt{\tau'^2}/\tau \in [0, 1]$ is a balance indicator that quantifies the asymmetry of the two delays. For example, $\eta = 1$ implies that these two delays are equal and $\eta = 0$ when one of delay is zero; $\epsilon = 2\sqrt{\gamma}/\sigma \in [0, 1]$ is another balance indicator introduced before, which plays a similar role as η for the degradation rates.

The oscillation criterion then becomes:

$$\mathcal{H}_{\epsilon,\eta}(\Sigma, T) = \frac{\epsilon^2\,\Sigma^2}{4}T^2 + (\Sigma - \frac{1}{\Sigma})T + 1 + T\eta^2\,(\frac{\epsilon^4\,\Sigma^4\,T^3\,\eta^2 - 64\Sigma - 32\epsilon^2\,\Sigma^2\,T}{64(\Sigma\,T\eta^2 + 4)} - \frac{T}{4}) < 0 \tag{2.12}$$

where T is the sum of the two reaction delays derived from different biological mechanisms (protein transport delay and gene response time), and Σ is the sum of degradation rates.

When only one of the delays is present and the other is zero, then we have $\eta = 0$. In this case, the criterion reduces to the first 3 terms in (2.12) are the same as the criterion (2.4), as expected. In the general case where both delays are present, these terms quantify how the total sum of the two reaction delays influences the appearance of sustained oscillations. The last term in criterion (2.10), however, demonstrates that these delays can interact together in a non-trivial manner. Thus the influence of two coexisting reaction delays on the appearance of oscillations cannot be deduced simply from their sum. In particular, their interaction can either promote or reduce oscillations depending on the sign of the interaction term. In the limit case where one delay is much smaller than the other, the interaction term is neglectful compared to the first 3 terms so that it is the sum of delays that influences oscillations. However, in general cases, the term of delay interaction may become significant, which has to be taken into account.

In conclusion, we found that the analytical criterion (2.12) not only depends on the two key parameters which are the sum of delays T and the sum of degradation rates Σ but also on how the delays and degradation rates are balanced, which is quantified respectively by the indicators ϵ and η. In the limit case where one of reaction delays is zero ($\eta = 0$), the last term of criterion (2.12) becomes zero and the oscillation criterion 2.4 is recovered, regardless of whether the delay originates from gene response or transport. In the general case, the dependence of criterion (2.12) on η demontrates that the dynamics is not only influenced by the sum of delays T but also on their relative importances. In particular, note that in the case where degradation is saturated ($\epsilon = 0$), the delay interaction term is always negative, which means that delay interation then always promotes oscillations.

In order to understand more closely how oscillations are affected by the combination of two reaction delays, we now study the bifurcation diagram with more detail, and in particular the behavior of oscillation thresholds in different limiting cases.

2.2.2 Bifurcation diagrams and interaction of reaction delays

The criterion (2.12) defines a series of curves $\Sigma_{\epsilon,\eta}(T)$ specifying the degradation rate Σ at oscillation threshold as a function of delay sum T and balance indicators ϵ and η. For fixed ϵ and η, oscillations can be found for $\Sigma \leq \Sigma_{\epsilon,\eta}(T)$. Fig 2.4A shows four curves corresponding to four limit cases of criterion (2.12), where the two indicators ϵ and η take their maximal and minimal value (respectively 1 and 0). These limiting curves are particularly important to understand the bifurcation diagrams. To provide numerical support for our analysis, we have searched for the parameter space of equations (2.5) for oscillatory behavior in the case where $\alpha, \mu \neq 0$ and where protein degradation and RNA degradation obey respectively allosteric and Michaelis-Menten enzymatic kinetics, as described by:

$$f(p) = \frac{p(a + p/\kappa)}{a + 2a(p/\kappa) + (p/\kappa)^2}\,,\ h(m) = \frac{\chi m}{\chi + m} \tag{2.13}$$

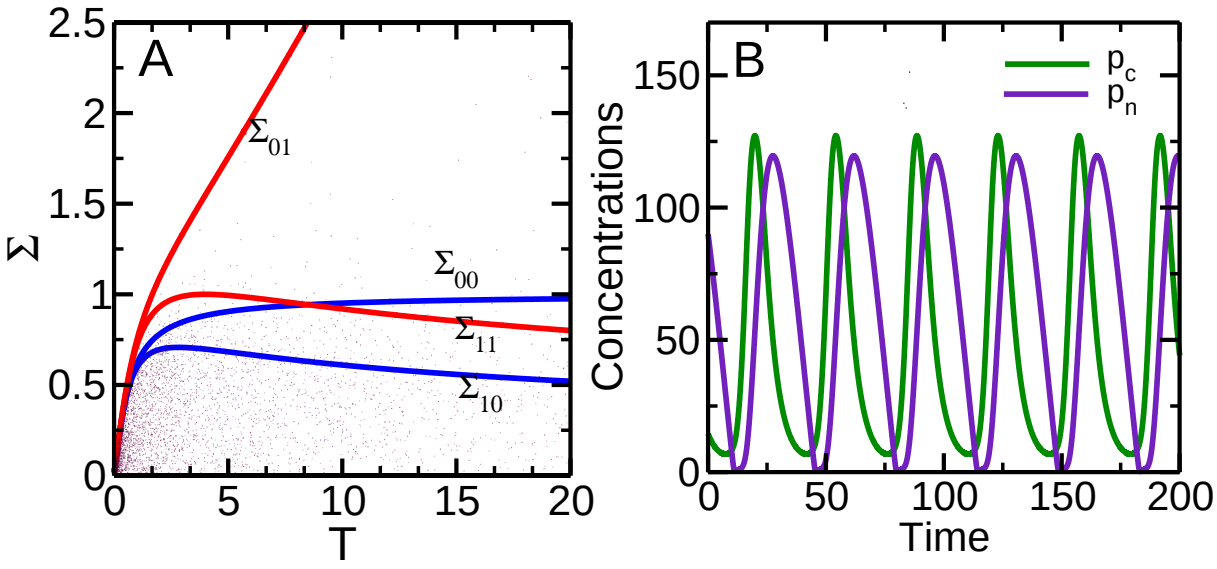

Figure 2.4: **Bifurcation digrams of the model with a reaction transport delay and finite gene response time**. (A) Oscillation threshold specifying the total degradation rate Σ as a function of sum of the two reaction delays T for limiting values of the indicators ϵ and η. The two blue curves correspond to the limit cases in which one of delay is zero ($\eta = 0$) as in [138]; the two red curves correspond to the limiting cases where the two delays are equal. Black dots indicate parameter sets giving rise to oscillations, with $\theta \in (0.001, 1000), \alpha, \mu \in (0.001, 10), \lambda, \delta \in (0.1, 1000), \tau_p \in (0.001, 1000)$, assuming degradation mechanisms described by equation (2.13) with $a, \kappa, \chi \in (0.1, 100)$. (B) Illustration of oscillatory behavior observed in model (2.5) for the parameter set: $\theta = 0.1,\ \lambda = 50,\ \delta = 1,\ \tau_p = 5,\ a = 0.01,\ \kappa, \chi = 10$.

Parameter sets giving rise to sustained oscillations are shown as black dots in Fig 2.4A. Note that numerical simulations agree very well with our analytical criterion even though they do not assume perfect repression ($\mu = 0$) or a large half-repression threshold ($\alpha \sim 0$). Indeed, all dots are below the $\Sigma_{01}(T)$ and only a few of them, which correspond to parameter sets where one degradation rate is small are found above $\Sigma_{11}(T)$.

The curves $\Sigma_{1,0}$ and $\Sigma_{0,0}$ correspond to the cases where there is just one delay (either gene response time or protein transport) and are the same as those discussed by Morant *et al.* [138] for the self-repressing gene with finite response time. In order to illustrate effects of combination of two reaction delays, we first compare the curves $\Sigma_{1,0}$ and $\Sigma_{1,1}$, in which degradation rates are equal ($\epsilon = 1$). In the first curve there is a single delay ($\eta = 0$). In the second one, the total time delay T is divided into two identical delays corresponding to two different biological steps ($\eta = 1$). When T is small, these two curves are superimposed, which implies that it is the sum of delays that drives the system dynamics. However, when T becomes sufficiently larger, these curves separate and the one with two identical delays is always above the other, which means that oscillations can be achieved at higher degradation rates. Thus, delays distributed in intermediate steps of a genetic loop promote oscillations more easily than in a single step, which contributes to the robustness of oscillations. In addition, the resonance-like effect described by Morant *et al.* [138] is preserved: there is always one gene response time for which the system is most easily destabilized.

Secondly, we compare the curves $\Sigma_{0,0}$ and $\Sigma_{0,1}$, for which either protein or mRNA degra-

dation is saturated ($\epsilon = 0$), in which we observe a notable phenomenon. It is found that the oscillation threshold indicated by $\Sigma_{0,1}$ with two identical reaction delays can become remarkably higher than in the single delay case. For practical purposes, it is almost as if oscillations can always be obtained for a sufficient large delay, given that the maximum of the curve is very high.

In the above discussion, we have analyzed the model with a reaction delay describing unidirectional protein transport from cytoplasm to nucleus, whereas transport can be reversible if it results from ordinary diffusion. We consider the latter case in the next section.

2.2.3 Dynamical effect of reversible transport delay

Here, we analyze model (2.6) to discuss the dynamical effect of delays resulting from bidirectional transport of proteins between cytoplasm and nucleus. The steady state (g_*, m_*, p_{n*}) of Eqs. (2.6) does not depend on θ, τ_p or τ_r which characterize the time scales of delays. Thus, tuning gene response time or the two transport delays does not change the steady state, and therefore not the degradation rates either. We introduce the quantity $\mu = \tau_p/\tau_r$ which is the ratio between the transport delay from cytoplasm to nucleus and the one from nucleus to cytoplasm.

To assess how the backward delay τ_r influences the appearance of sustained oscillation, we use again the same strategy of searching for parameter values where the fixed point loses stability via a Hopf bifurcation. After a suitable parameter normalization, we can obtain an analytical oscillation criterion $\mathcal{H}_{\epsilon,\eta}(\Sigma, T, \mu)$, which is very complicated so that we will not reproduce it here. This criterion reduces to (2.12) when the new parameter μ verifies $\mu = 0$. The curves $\Sigma_{\epsilon,\eta}(T, \mu)$ giving the oscillation threshold in the particular case where gene response time is equal to protein transport delay from cytoplasm to nucleus are shown in Fig 2.5 for different values of parameter μ.

When the parameter μ is increased, the oscillation region in parameter space is progressively reduced, requiring smaller and smaller degradation rates. This implies that adding a backward transport delay obstructs sustained oscillations. Note that $\mu > 1$ indicates that the backward transport is more rapid than the protein transport from cytoplasm to nucleus, in this case, the backward transport decreases coupling between gene activity and protein and stabilizes the system.

In this section, we have investigated the dynamical influence of combining two reaction delays on the oscillatory behavior of a minimal genetic network. Remarkably, an analytical criterion was found, which allowed us to show that reaction delays interact in a non-trivial way so that how reaction delays are distributed can affect the dynamics. In the next section, we will discuss the case where transport is modeled using explicit delays.

2.3 Analytical oscillation criterion of the model with explicit delay

In model (2.7), there are two explicit transport delays: τ_1 denotes the protein transport from cytoplasm to nucleus; τ_2 is the RNA transport form nucleus to cytoplasm. As with reaction delays, we first find the steady state (g_*, m_*, p_*), which is the same as for the model (2.2) without transport delays. We also find that the steady state does not depend on the parameters

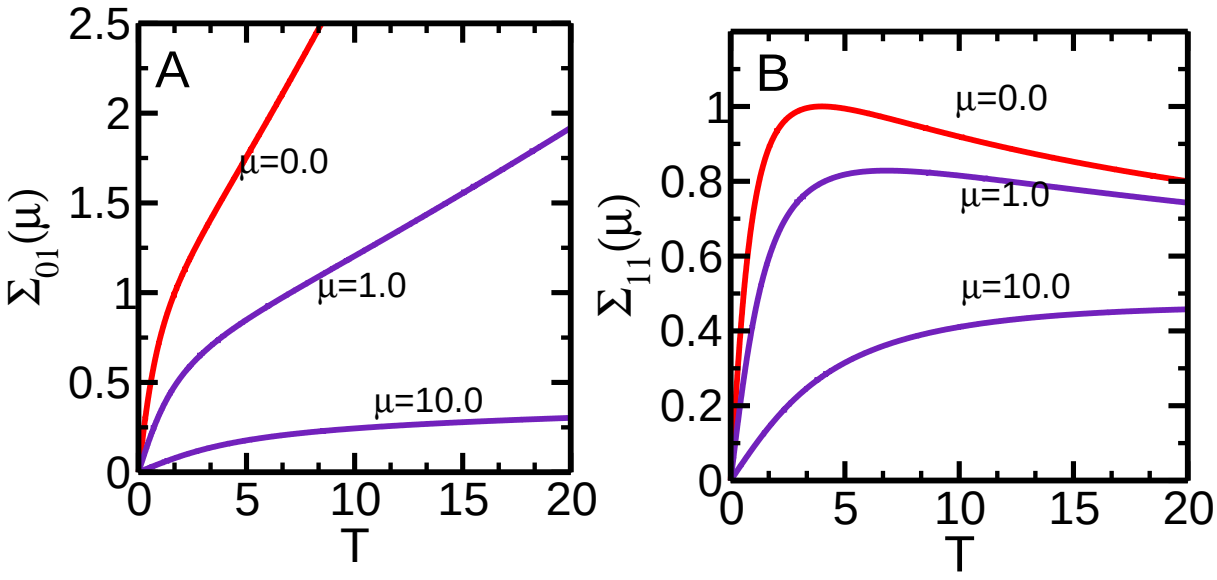

Figure 2.5: **Influence of the reverse transport delay**. (A)Oscillation thresholds $\Sigma_{0,1}(\mu)$ for different values of reverse transport delay in the case where at least one degradation is saturated and where gene response time (τ_g) and the transport delay of protein from cytoplasm to nucleus (τ_p) are equal. $\mu = 0$ means that transport is irreversible. (B)Oscillation thresholds $\Sigma_{1,1}(\mu)$ for different values of reverse transport delay in the case where two degradation rates are equal and where gene response time (τ_g) and the transport delay of protein from cytoplasm to nucleus (τ_p) are also equal.

τ_1 and τ_2 which characterize the time scale of transport delays. Degradation rates of proteins and RNAs are still defined as the slopes of degradation function in the neighborhood of steady state, and they are independent of transport delays τ_1, τ_2 as well.

To assess whether the model (2.7) can display sustained oscillations, we apply also the linear stability analysis in the neighborhood of the steady state by writing the Jacobian matrix [73, 116]. The characteristic equation for eigenvalues obtained is the following:

$$\xi^3 + \frac{1 + g_*\tau_g(\delta\, s + u)}{g_*\tau_g}\xi^2 + \frac{\delta\, s + u + g_*\tau_g \delta\, s\, u}{g_*\tau_g}\xi + \frac{\delta\, s\, u + \delta\, \lambda\, g_*^2\, \exp\left(-(\tau_1+\tau_2)\xi\right)}{g_*\tau_g} = 0 \tag{2.14}$$

where ξ denotes the eigenvalue. If it is positive, the steady state is stable; otherwise, the steady state is unstable and sustained oscillations are observed. It appears in Eq. (2.14) that explicit delays τ_1 and τ_2 only influence the dynamics through their sum $\tau_1 + \tau_2$, unlike with reaction delays.

Now we further analyze Eq. (2.14). We search for values of the total delay $\tau_e = \tau_1 + \tau_2$ for which Eq. (2.14) has a pair of conjugate pure imaginary roots. Therefore, we assume $\xi = \pm i\omega$ ($\omega > 0$) as roots of Eq. (2.14). By separating the real and imaginary parts, we obtain the following equations:

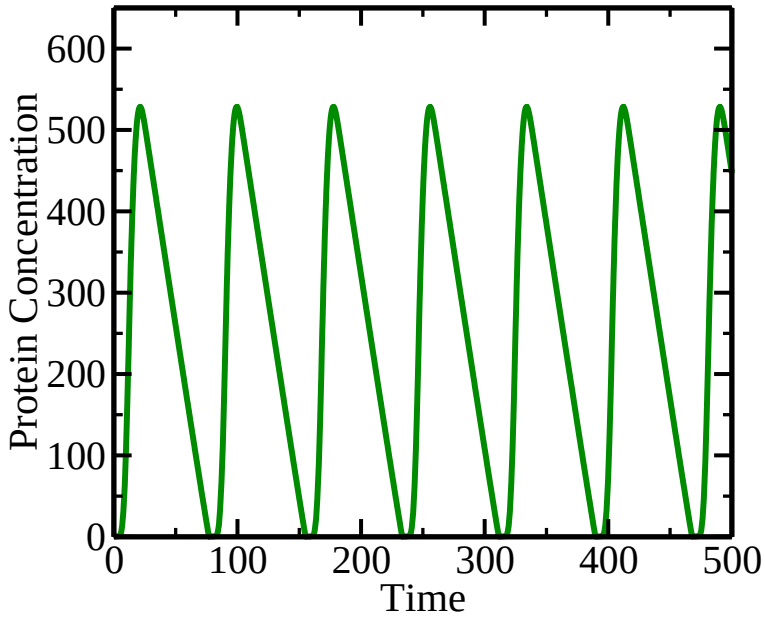

Figure 2.6: **Illustration of oscillations of model (2.7)**. Numerical simulation is carried out with parameters $\theta = 0.1$, $\lambda = 50$, $\delta = 1$, $\tau_p = 5$, $a = 0.01$, $\kappa, \chi = 10$ and specific degradation of mRNA and proteins described by Eqs. (2.13).

$$\cos(\omega\,\tau_e) = \frac{(1 + g_*\tau_g(\delta\,s + u))\omega^2 - \delta\,s\,u}{g_*^2\delta\,\lambda} \tag{2.15a}$$

$$\sin(\omega\,\tau_e) = \frac{-g_*\tau_g\,\omega^3 + (\delta\,s + u + g_*\tau_g\,\delta\,s\,u)\omega}{g_*^2\delta\,\lambda} \tag{2.15b}$$

where τ_e is the sum of explicit delays; $\tau_g = g_*/\theta$ characterizes the time scale of gene response; $\delta\,s$ and u are respectively the protein and RNA degradation rates. Eqs. (2.15) provide relations between the parameters at which the system losses the stability via a Hopf bifurcation and the frequency of periodic solution ω (Fig 2.6). The latter may be in principle eliminated from the equations to yield the equation of an hypersurface in parameter space.

For the sake of simplicity, we renormalize parameters as follows:

$$\sigma = \delta\,s + u, \gamma = \delta\,su, \sigma_c = g_*\sqrt{\delta\,\lambda},\ \sigma = \sigma_c\,\Sigma,\ \gamma = \frac{\epsilon^2\,\Sigma^2\,\sigma_c^2}{4},\ \tau_e = \frac{T_e}{\sigma_c},\ \tau_g = \frac{T_g}{\sigma_c},\ \omega = \Omega\,\sigma_c \tag{2.16}$$

where $\epsilon = 2\sqrt{\gamma}/\sigma \in [0, 1]$ is defined as the balance indicator of degradation rates as before. Thus, Eqs. (2.15) become:

$$\cos(\Omega\,T_e) = (1 + \Sigma\,T_g)\Omega^2 - \frac{\epsilon^2\,\Sigma^2}{4} \tag{2.17a}$$

$$\sin(\Omega\,T_e) = -T_g\,\Omega^3 + (\Sigma + \frac{\epsilon^2\,\Sigma^2\,T_g}{4})\,\Omega \tag{2.17b}$$

T_e is the normalized total delay; T_g is the normalized gene response time; Σ represents the sum of degradation rates; Ω is the frequency of periodic solution when Hopf bifurcation occurs.

Eqs. (2.17) give the oscillation threshold at which the system loses stability to a periodic solution via Hopf bifurcation. They define a series of curves $\Sigma_\epsilon(T_g, T_e)$ which specify the degradation rate Σ at threshold as a function of gene response time T_g and explicit transport delay T_e. Oscillations are found in the region $\Sigma \leq \Sigma_\epsilon(T_g, T_e)$

In order to assess how the bifurcation frequency depends on delay and degradation rate, we square both sides of Eq. (2.17) and add them, then obtain:

$$F(\Psi) = \Psi^3 + (\frac{1}{T^2} + (1 - \frac{\epsilon^2}{2})\,\Sigma^2)\Psi^2 + (\frac{\epsilon^4\,\Sigma^4}{16} + (1 - \frac{\epsilon^2}{2})\,\frac{\Sigma^2}{T^2})\,\Psi + \frac{1}{T^2}(\frac{\epsilon^4\,\Sigma^4}{16} - 1) = 0 \quad (2.18)$$

where $\Omega = \sqrt{\Psi}$. Since $\epsilon \in [0, 1]$, all coefficients of non zero degrees of Ψ are positive. According to the Descartes rule of signs, $\frac{\epsilon^4\,\Sigma^4}{16} - 1 < 0$ leading to $\Sigma < 2/\epsilon$, or equivalently $\sigma_c^2 > \gamma$, is a necessary condition to have a positive root for Ψ and thus a necessary condition for oscillatory behavior as well. The fact that only the product, and not the sum, of the degradation rates, is constrained is interesting because it implies that oscillations will always be observable if one of the degradation rates is sufficiently small. Once the value Ω of the Hopf frequency is obtained, it can be substituted in Eqs. (2.17) to solve for Σ as a function of total delay T_e and response time T_g.

In conclusion, we have analyzed a model with two explicit delays and demonstrated that these two explicit delays have the same effect as a single total delay equal to their sum. We also derived a simple analytical criterion for oscillations given by Eqs. (2.17) and (2.18). This criterion will allow us to compare the influences of reaction and explicit delays on the appearance of oscillations in the next section.

2.4 Comparison of dynamical influences between reaction and explicit delays

In the two previous sections, analytical criteria for oscillating behavior in models both with reaction and explicit transport delays have been obtained. It was found that the combination of two explicit delays is equivalent to a single delay equal to their sum, whereas reaction delays interact in a non-trivial way. This gives a hint that the dynamical influences of reaction and explicit delays are different. In the scientific literature, however, reaction or explicit delay are used in a relatively arbitrary way, seemingly according to the personal preferences of the authors. Therefore, we discuss here similarities and differences in the dynamical behaviors of these two types of delay.

As a starting point, we consider bifurcation diagrams for a sequence of increasing transport delays, modeled either as a reaction delay or as an explicit delay (Fig 2.7). In these diagrams, the degradation rate at oscillation threshold Σ is given as a function of gene response time T_g for a given transport delay. Oscillation are found for the region $\Sigma \leq \Sigma(T_g)$. We show two limiting cases where $\epsilon = 1$ or $\epsilon = 0$, corresponding respectively to the cases where two degradation rates are equal or where one degradation rate is zero. The Σ_{00} and Σ_{10} curves in blue correspond to the cases where there is no transport delay and serve as references.

For the limiting case $\epsilon = 0$ (Fig 2.7A), the larger the transport delay, either in reaction or explicit form, the more extended the oscillation zone is, which indicates transport delay always promotes oscillation. Moreover, because the oscillatory zone is much more extended

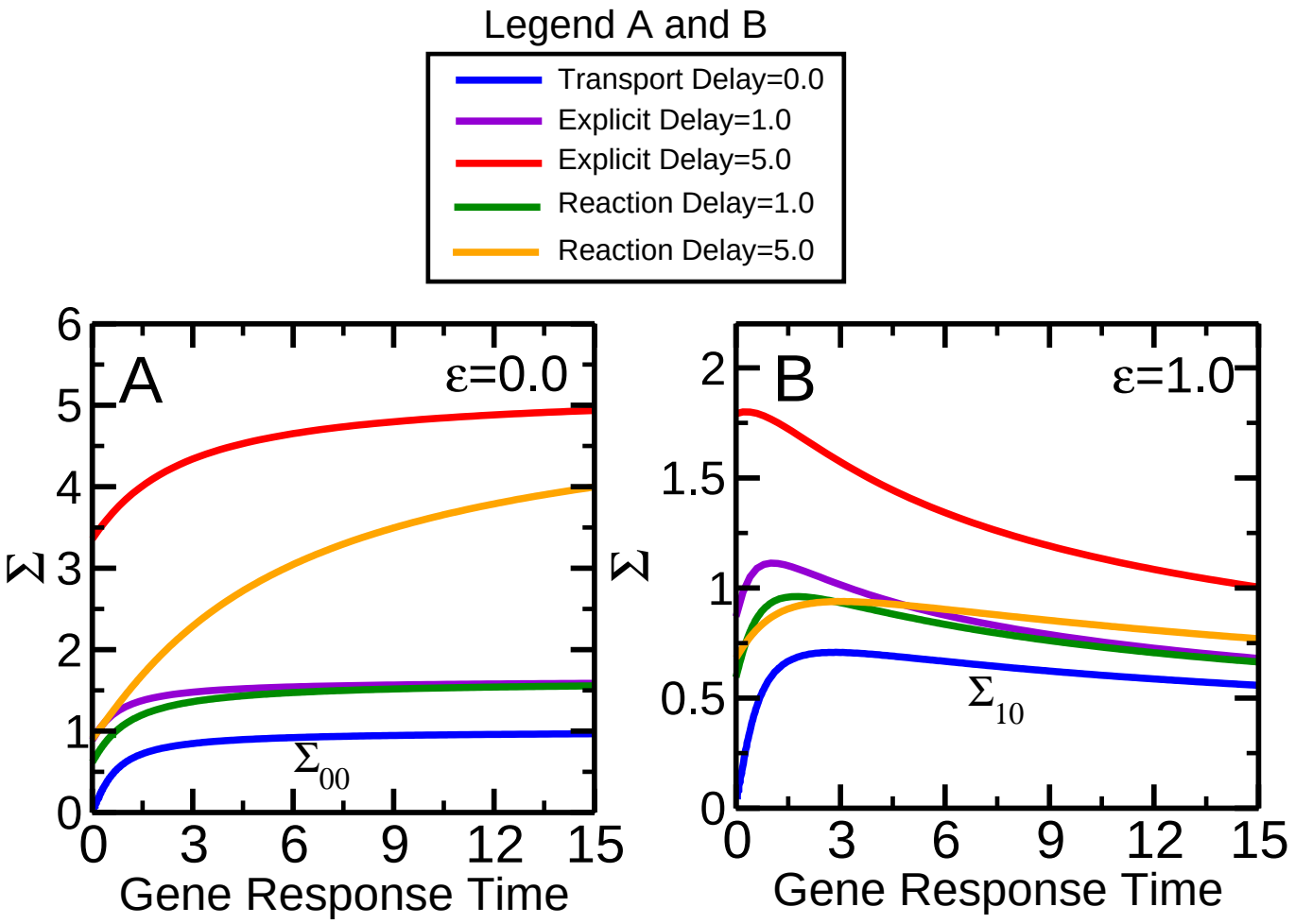

Figure 2.7: **Oscillation threshold in the (Σ, T_g) plane for fixed values of reaction or explicit transport delay**. (A) The case $\epsilon = 0$ means one degradation is saturated. In the blue curve Σ_{00} the transport delay is zero. (B) The case $\epsilon = 1$ is two degradation rates are equal. In the blue curve Σ_{10} the transport delay is also zero.

for explicit delay, we can conclude that the explicit delay is more destabilizing than a reaction delay. Despite quantitative differences, the influences of both types of delays are qualitatively similar.

Fig (2.7)B represents the limiting case $\epsilon = 1$ where the two degradation rates are equal. The resonance-like effect in which the genetic circuit is most destabilized at a finite gene response time can still be observed, but its behavior depends clearly on the type of delay. By increasing gradually the explicit delay, the oscillatory zone becomes more and more extended and the resonance point approaches rapidly the Σ axis, which means that increasing the gene response time will typically stop oscillations. In contrast, when we increase gradually the reaction delay, the resonance point first moves towards the Σ axis and then towards large values of T_g. Moreover the degradation rate Σ at the maximum of the $\Sigma(T_g)$ curve (the "resonance point") saturates at 1. The oscillation regions for small and large reaction delays superimpose relatively well. In this limiting case, influences of explicit and reaction delays seem to be quite different except in the case where they are both small compared to the gene response time. Increasing the explicit delay always promotes oscillations, much more clearly than for a reaction delay.

In order to have a global point of view on the influences of reaction and explicit delays, we now consider bifurcation diagrams in the plane of the gene response time and of the transport delay modeled as either reaction or explicit delay (Fig 2.8). These bifurcation diagrams are computed for different values of the total degradation rate Σ.

We start with Figs 2.8A and C corresponding to the limiting case $\epsilon = 0$ where one degradation mechanism is saturated. It is seen that when the sum of degradation rates Σ is increased, oscillation regions are gradually reduced for both explicit and reaction delays. When both degradations rates are approaching to saturation ($\Sigma \sim 0$), oscillation regions occupy almost the first quadrant of plane, which confirms that saturated degradations facilitate sustained oscillations very much. The oscillation region for a reaction delay is symmetric, resulting from the fact that reaction transport delay and gene response play symmetric roles in the analytical criterion. In contrast, the oscillation region for explicit delay is not symmetric, which is not surprising, given that bifurcation diagrams involve both an explicit delay and a reaction delay, the response time. Note that for some small values of Σ, the instability threshold crosses the axis and enter into the second quadrant of the plane where the explicit delay is negative, which corresponds to the fact that the circuit already oscillates at zero delay. Besides the differences in the symmetries of oscillatory regions, the influences of explicit and reaction delay are qualitatively similar.

When degradation is saturated, the bifurcation occurs for the same value of reaction and explicit delays when $\Sigma < 0.7$ (inset of Fig 2.8A). For $\Sigma > 0.7$, reaction and explicit delays are also equivalent when the gene response time is sufficiently large. However, it can be seen in Fig 2.8A that when the gene response time is smaller than 2, oscillations are much more easily reached with an explicit delay than with a reaction one. More generally, it is interesting to note that even when the gene response time is large, a small added transport delay can trigger oscillations and vice versa. Once again, the fact that oscillations appear for a total delay much smaller when there are two delays rather than one indicates that a sequence of biological processes inducing delays can greatly emphasize oscillations.

Figs 2.8B and D correspond to the limiting case $\epsilon = 1$, where protein and RNA degradation rates are equal. There are significant qualitative changes in the bifurcation diagrams compared to the case $\epsilon = 0$, especially for reaction delays. This suggests that the degradation rate balance

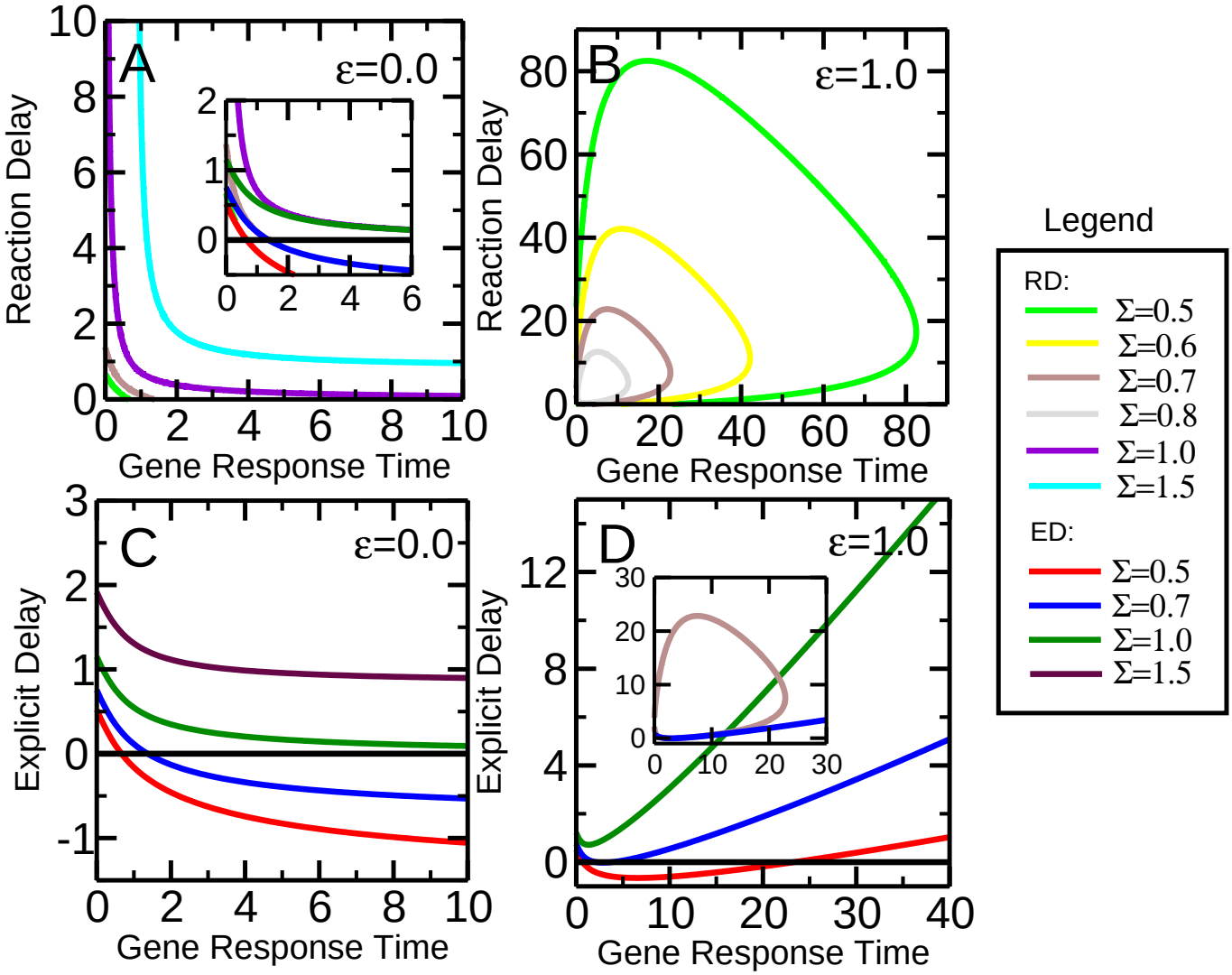

Figure 2.8: **Bifurcation diagrams in the plane of gene response time and reaction or explicit transport delay**. (A) Oscillation thresholds for reaction transport delay at $\Sigma = 0.5, 0.7, 1.0, 1.5$ in the case $\epsilon = 0$ where one degradation is saturated. Oscillation regions are above curves. The inset is to compare oscillation thresholds for small explicit and reaction delays at $\Sigma = 0.5, 0.7, 1.0$. (B) Oscillation thresholds for reaction transport delay at $\Sigma = 0.5, 0.6, 0.7, 0.8$ in the case $\epsilon = 1$ where two degradation rates are equal. Oscillation regions are inside of curves. (C) Oscillation thresholds for explicit transport delay at $\Sigma = 0.5, 0.7, 1.0, 1.5$ in the case $\epsilon = 0$ where one degradation is saturated. Oscillation regions are above curves. (D) Oscillation thresholds for explicit transport delay at $\Sigma = 0.5, 0.7, 1.0$ in the case $\epsilon = 1$. Oscillation regions are above curves. The inset is to compare oscillation thresholds of explicit and reaction delays for $\Sigma = 0.7$.

measured by ϵ is also a key factor for deciding the appearance of oscillations. By increasing gradually the degradation rate sum Σ, oscillation regions for reaction and explicit delays are both reduced, and no oscillations are observed for $\Sigma > 1$ (reaction delays) or $\Sigma > 2$ (explicit delays). However, a marked difference between reaction and explicit delays is that for explicit delays, oscillations can always be observed whenever the transport delay is larger than the threshold whereas for reaction delays, oscillations disappear when the delay is too large. This corresponds to the fact that in the latter case, the oscillation region is contained inside a closed curve.

To summarize, Fig 2.8 allows us to understand the similarities and differences in the dynamical influences of reaction and explicit delays on oscillatory behavior. Whether degradation rates are balanced or not is clearly an important factor for the appearance of biological oscillations. When degradations of protein and mRNA are approaching to saturation or transport delay is small, reaction and explicit delays have similar effects. A sufficiently large explicit delay always destabilizes the system, however a large reaction delay can restabilize the oscillating system. This has to be taken into account when designing the mathematical modeling.

2.5 Conclusions

Previous work [138] showed that a delay due to the slow dynamics of gene activity can induce oscillations in a minimal genetic circuit comprising a self-repressing gene. In this chapter, we extended this model by adding delays due to molecule transport described both as reaction and explicit delays. Our results, based on analytical criteria, confirm the crucial role of delays in the design principle of biological oscillators and the trade-off between delay and degradation nonlinearity. How delays combine was also revealed: the influence of two explicit delays only depends on their sum, while reaction delays interact in a non-trivial way. Thus, how delays are distributed can affect the oscillatory dynamics. Reaction and explicit delays have the same dynamical influences only if delays are small or protein and RNA degradation are saturated. In the general case, their influences are dramatically different. For example, an explicit delay always favors oscillations, in contrast, a reaction delay can either destabilize or stabilize genetic circuits.

Chapter 3

Influences of stochastic fluctuations on the oscillation of a self-repressing gene

3.1 Introduction

The genetic networks that are responsible for cellular dynamics and functions are subject to large fluctuations (or noise) in the number of molecules, as has been revealed by numerous experimental and theoretical investigations [38, 105, 179, 183, 33, 151]. Indeed, most biochemical reactions in cells involve very low copy numbers of reacting molecules. Noise may be a nuisance, because it makes it more difficult to obtain a deterministic behavior, however an interesting question is why cells are designed with it. Recent studies provide strong supports for the hypothesis that cellular noise may be involved in design principles of living cells [195, 35, 5, 77, 140, 185].

In particular, noise in gene expression have been identified as a key biological effect underling observed phenotypic variability of genetically identical cells in homogeneous environments [157, 69, 173, 128, 157, 120]. Furthermore, it plays a crucial role in coordinated expression of a large set of genes and in probabilistic differentiation strategies [195, 35, 5, 77]. Although negative feedback loops can in principle reduce noise [183, 33, 47, 9], they display fundamental limitations [125]. Therefore, molecular noise, wanted or unwanted, is unavoidable in biological systems. However, how noise influences the dynamics of genetic networks is not yet fully understood in spite of the numerous works devoted to its study.

Here, we will investigate the dynamical influences of noise in the same minimal genetic circuit consisting of a self-repressing gene that we have studied before [137, 127, 138]. In this simple circuit, most of the stochasticity resides in the fact that there is a single copy of the gene and that the dynamics of its activity is complex and should be taken into account [85]. In a molecular point of view, the gene dynamics is usually modeled as a series of transition between "ON" and "OFF" states [85, 151]. When the promoter site is bound by repressor (OFF state), transcription is repressed, while the transcription occurs when the repressor is dissociated from gene (ON state). The fluctuations resulting from the stochastic alternation between these two gene states may decisively affect the system stability and oscillatory behavior [124, 170, 147, 194, 166, 99, 64, 68, 193].

The common approach to modeling gene expression using mean-field deterministic rate equations provides insights into the time evolution of molecular average quantities, but it is not appropriate to describe fluctuations [122]. Moreover, it is often overlooked that these fluctua-

tions will in fact impact the deterministic dynamics described by the mean-field equations and may in some cases dramatically change the behavior predicted classically. Although noise is general viewed as a zero-mean fluctuation superimposed on the deterministic average, this is no longer true as soon as this noise interacts with the nonlinearities present in the genetic circuit. For example, if x is a zero-mean stochastic variable, then $(1+x)^2$ will not have a mean of 1.

Therefore, stochastic models of genetic circuits, where the integer nature of copy numbers are fully taken into account, are required. Such models can easily be simulated numerically, however this provides generally little insight because it is difficult to assess the influence of the various parameters.

The most general theoretical approach is writing the master equations which describe the time evolution of probability distribution of microscopic states [189] characterized by the integer copy numbers of molecules. The master equation describes precisely the temporal evolution of the system and is typically solved by numerical integration [169, 139]. However, the disadvantage is that there are only very few examples which may be analytically solved, even when only the steady state is desired [158, 98, 146, 46]. For instance, Hornos et al. [98] derive an exact solution of the master equation for the self-repressing gene under a hypothesis that an adiabatic elimination of the mRNA variable. However this hypothesis is not realistic. Some approximations of the master equations can also be considered, such as the Fokker-Plank or Langevin equations [7, 8, 36, 189], however they are not appropriate to describe fluctuations linked to a binary variable such as the gene state.

The aim of this chapter is to derive a cumulant expansion of the master equation for the genetic circuit of a self-repressing gene. In this way, we will obtain equations for the dynamics of the mean-field averages taking into account the effect of fluctuations by incorporating a few higher-order cumulants as dynamical variables. This will allow us to investigate how fluctuations dramatically shift the steady state and to compare oscillation regions to what is predicted by usual deterministic rate equations.

This chapter is organized as follows: First, we write the master equations for the genetic circuit. Second, we show how to derive the cumulant expansion from the master equation. In the third section, we examine the steady state of a set of equations for cumulants, and compare it with the one predicted by usual deterministic rate equations. The last section is devoted to the dynamical influences of fluctuation on oscillatory behavior.

3.2 Master equation of the model of a self-repressing gene

We first consider the master equation that describes the dynamics of self-repressing gene circuit. In Fig (3.1) we show all biochemical reactions in this genetic network.

The gene synthesizes mRNA with transcription rate λ when active. Protein monomers bind the gene with a constant rate α, and then repress transcription completely. The bound monomer dissociates from the gene at rate θ. β is the protein translation rate. Because we are interested in understanding how fluctuations can spontaneously induce oscillations, we minimize the amount of nonlinearities in the circuit, so as to stay away from configurations where oscillations appear deterministically. We thus consider linear degradation mechanisms for mRNA and proteins, with the associated degradation rates δ_m and δ_p.

The parameter $P_0 = \theta/\alpha$ represents the number of proteins at which gene activity is repressed to half the maximum level. This threshold is assumed to be large, $P_0 \gg 1$ leading to

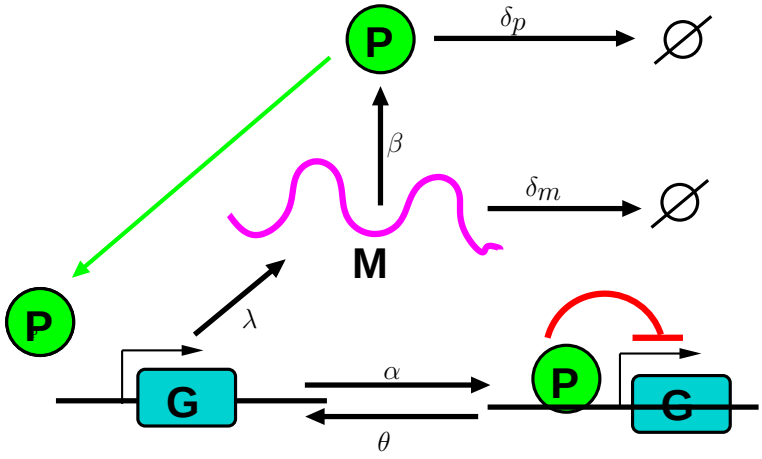

Figure 3.1: **Reaction diagram of the self-repressing gene**. In this minimal genetic circuit, a single protein binding to gene perfectly represses the transcription. Degradations of protein and mRNA are both linear.

$\alpha \simeq 0$, so that binding or unbinding of a single protein does not change protein copy number significantly. According to the analysis in previous chapter, the genetic circuit studied can then be described by the following deterministic rate equations with reduced variables and parameters:

$$\frac{dG}{dt} = \Theta\,(1 - G(1+P)) \tag{3.1a}$$

$$\frac{dP}{dt} = \delta(M - P) \tag{3.1b}$$

$$\frac{dM}{dt} = \Lambda\, G - M \tag{3.1c}$$

where G, P and M are respectively the average gene activity, and protein and mRNA copy numbers. Note that the notion of gene activity average is natural in the limit case where binding-unbinding is much faster than other processes. The steady state of Eqs. (3.1) is given by

$$G_* = \frac{-1+\sqrt{1+4\Lambda}}{2\Lambda},\ P_* = M_* = \Lambda G_* \tag{3.2}$$

and is found to be unconditionally stable. Thus deterministic rate equations predict that there are no oscillations in the case of linear degradation.

In stochastic simulations, gene activity alternates randomly between the "ON" and "OFF" states and can be described by a binary variable, with value 0 or 1. There is no macroscopic limit in which this variable becomes continuous, in contrast with protein or mRNA copy numbers which become large when Λ is increased. Because of this important property, common theoretical descriptions, such as deterministic rate equations under mean-field approximation, Fokker-Plank or Langevin equations, are in principle not appropriate to describe the dynamics of this genetic circuit. However, the master equation describing the time evolution of probabilities for all microscopic states characterized by gene activity, copy numbers of mRNA and protein, is still valid. To formulate the master equation more easily, we consider separately the

time evolution of the quantities $P_{0,m,p}$ and $P_{1,m,p}$ which are respectively the probabilities of finding the gene bound ($g = 0$) or unbound ($g = 1$) with m copies of mRNA and p copies of protein:

$$\begin{aligned}
\frac{dP_{0,m,p}}{dt} &= \alpha(p+1)P_{1,m,p+1} - \theta\, P_{0,m,p} \\
&\quad +\delta_m(m+1)P_{0,m+1,p} - \delta_m m P_{0,m,p} \\
&\quad \beta\, m P_{0,m,p-1} - \beta\, m P_{0,m,p} \\
&\quad +\delta_p(p+1)P_{0,m,p+1} - \delta_p\, p P_{0,m,p} \\
\frac{dP_{1,m,p}}{dt} &= \theta\, P_{0,m,p-1} - \alpha\, p P_{1,m,p} \\
&\quad \lambda\, P_{1,m-1,p} - \lambda\, P_{1,m,p} \\
&\quad \delta_m(m+1)P_{1,m+1,p} - \delta_m m P_{1,m,p} \\
&\quad \beta\, m P_{1,m,p-1} - \beta\, m P_{1,m,p} \\
&\quad \delta_p\,(p+1)P_{1,m,p+1} - \delta_p\, p P_{1,m,p}
\end{aligned}$$

The master equation is actually a gain-loss equation. Consider for example the equation of $P_{0,m,p}$. The probability of having the system in microscopic state $\{0, m, p\}$ can increase due to transition to $\{0, m, p\}$ from other states and decrease due to transition from $\{0, m, p\}$ to other states. More precisely, the term $\alpha(p+1)P_{1,m,p+1}$ describes how the probability of the state $\{0, m, p\}$ increases due to transition from $\{1, m, p+1\}$ by protein binding; on the other hand, $-\theta\, P_{0,m,p}$ is the loss of this probability due to transition from state $\{0, m, p\}$ into $\{1, m, p+1\}$ by the dissociation of protein from DNA.

Even for such a simple genetic circuit, the master equation is not easily solved analytically and must be studied numerically, except under some approximations [98].

In the next section, we propose an approach which will allows us to take into account fluctuations while obtaining deterministic dynamical equation: a cumulant expansion of the master equation. From the latter, differential equations are obtained for the time evolution of the cumulants of the distributions, the lowest-order cumulants being the averages involved in the mean-field equations. This allows us to incorporate the influence of fluctuations in a hierarchical way.

3.3 Incorporating variances in a deterministic model: cumulant expansion to the second order

Before deriving the cumulant expansion of the master equation, we first recall how moments and low-order joint cumulants are defined. The set of moments of a probability distribution function (PDF) characterizes it completely. Given the probability distribution $P_{g,m,p}$ defined above, the moments of a quantity $x(g, m, p)$ depending on the circuit state are defined as the expectation values of the powers of x:

$$\mu_n = \langle x^n \rangle = \sum_{g,m,p} x^n(g, m, p) P_{g,m,p}$$

where n is the order of the moment. The moment-generating function is a series whose coefficients are the moments μ_n:

$$M_x(t) = \langle e^{tx} \rangle = 1 + \mu_1 t + \frac{\mu_2 t^2}{2!} + \frac{\mu_3 t^3}{3!} + ... \frac{\mu_n t^n}{n!} + ...$$

For example, the moment-generating function of a Gaussian PDF for variable x with average μ and variance σ^2 is $M_x(t) = e^{\mu t + 1/2\sigma^2 t^2}$.

The idea behind a moment expansion of the master equation is that if we know the time evolution of the probability distribution, we know the time evolution of the moment values:

$$\frac{d\mu_n}{dt} = \sum_{g,m,p} x^n(g,m,p) \frac{dP_{g,m,p}}{dt} \tag{3.4}$$

with $dP_{g,m,p}/dt$ being expressed in terms of the $P_{g',m',p'}$ and possibly of the state variables when the biochemical network features reactions where there are two or more reagents. For example, the variables p and m appear in the time derivates given in Eqs. (3.3). In this case, the time derivative of a given moment will typically involve moments of higher order so that Eqs. (3.4) are not closed but form an infinite hierarchy of equations. When reactions involve only one reagent, the time derivates of moments of order 1 are expressed in terms of moments of order 2, etc.

Solving the moment equation is equivalent to solving the master equation since the PDF can be reconstructed from its moments. The idea behind the moment expansion is that if only a few moments suffice to describe the dynamics, solving the corresponding equations (3.4) may be simpler than solving the master equation. However, this is not true if equations at all order are coupled, as mentioned above. It is then necessary to truncate the hierarchy of equations, which means to choose a way to approximate the moments of certain orders, without solving for them [189].

Actually, this approach is more efficiently formulated in terms of cumulants, a notion closely related to moments. The cumulants K_n of x are defined via the cumulant-generating function, which is obtained as the logarithm of the moment-generating function:

$$G_x(t) = \sum_{n=1}^{\infty} K_n \frac{t^n}{n!} = \log M_x(t)$$

Thus cumulants are given by:

$$K_n = \frac{d^n G_x(0)}{dt^n} = \frac{d^n \log M_x(0)}{dt^n} \tag{3.5}$$

Eq. (3.5) allows us to express cumulants at any order as a combination of moments. Assuming a Gaussian-distributed variable x, we have $G_x(t) = \mu t + \sigma^2 t^2/2$, so that cumulants are given by $K_1 = \mu$ and $K_2 = \sigma^2$, with higher-order cumulants being zero. More generally, we obtain the following expressions for the lowest-order joint cumulants of random variables x, y z:

$$\begin{aligned} K_x &= \langle x \rangle \\ K_{xy} &= \Delta_{xy} = \langle xy \rangle - \langle x \rangle \langle y \rangle \\ K_{x,y,z} &= \langle xyz \rangle - \langle x \rangle \langle y \rangle \langle z \rangle - \langle x \rangle \Delta_{y,z} - \langle y \rangle \Delta_{x,z} - \langle z \rangle \Delta_{x,y}. \end{aligned} \tag{3.6}$$

Note that the first-order cumulant is the average and the second-order is the covariance. It can be seen that the second- and third-order cumulants can be expressed in term of lower-order cumulants, which is actually true at any order. In fact, joint cumulants defined by Eqs. (3.6) quantify the amount of correlation at a given order which does not arise from correlations at lower orders. For example, the second-order moment $\langle xy \rangle$ can be non-zero even though x and y are not correlated, simply because the averages $\langle x \rangle$ and $\langle y \rangle$ are not zero. In the cumulant Δ_{xy}, which is actually the joint moment of $x - \langle x \rangle$ and $y - \langle y \rangle$, the contribution from the averages is removed so that the cumulant is zero whenever the two random variables are not correlated. Thus, Δ_{xy} measures much better the correlation of x and y than the moment in some sense. Similarly the third-order cumulant $K_{x,y,z}$ is zero if random variables x, y and z are correlated only pairwise.

The presence of a binary variable, such as gene activity g (which takes values 1 or 0), can lead to special mathematical treatments [147, 194, 166, 99, 64, 68, 193], and also simplifies cumulant computation. More precisely, moments of any order of a binary variable g are equal to its average.

$$\langle g^n \rangle = \sum_{g,m,p} g^n P_{g,m,p} = \sum_{g,m,p} g P_{g,m,p} = \langle g \rangle$$

Consequently, all high-order cumulants of g can be expressed in terms of the first-order moment (the average), which reduces the dimension of the cumulant expansion. For example, the variance is written as:

$$\Delta_{gg} = \langle g \rangle (1 - \langle g \rangle) \tag{3.7}$$

As described above, cumulant expansion of the master equations yields a set of differential equations which describe the temporal variations of joint cumulants. We will now derive the equations for the lowest-order cumulants of the self-repressing gene circuit. We first renormalize the time, joint cumulants and parameters according to:

$$\begin{aligned}
&t \to T\delta_m, \quad \delta \to \frac{\delta_p}{\delta_m}, \quad \langle g \rangle \to G, \quad \langle p \rangle \to \frac{\theta}{\alpha} P, \quad \langle m \rangle \to \frac{\delta\theta}{\beta\alpha} M, \\
&\Delta_{pp} \to (\frac{\theta}{\alpha})^2 \Delta_{PP}, \quad \Delta_{mm} \to (\frac{\delta\theta}{\beta\alpha})^2 \Delta_{MM}, \quad \Theta \to \frac{\theta}{\delta_m}, \\
&\Delta_{gp} \to \frac{\theta}{\alpha} \Delta_{GP}, \quad \Delta_{gm} \to \frac{\delta\theta}{\beta\alpha} \Delta_{GM}, \quad \Delta_{mp} \to \frac{\delta}{\beta} (\frac{\theta}{\alpha})^2 \Delta_{MP}, \\
&\frac{\lambda\beta\alpha}{\delta\theta} \to \Lambda, \quad \frac{\theta}{\alpha} \to P_0
\end{aligned} \tag{3.8}$$

We also define the quantity $P_0 = \theta/\alpha$ which represents the number of proteins required to reduce the transcription rate by half. Again, we assume this threshold is large ($P_0 >> 1$) which is equivalent to $\alpha \simeq 0$.

The time evolution equations for the first-order cumulants (i.e., averages) and second-order

cumulants (i.e., covariances) are given by:

$$\begin{aligned}
\frac{dG}{dT} &= \Theta(1 - G - GP - \Delta_{GP}) && \text{(3.9a)}\\
\frac{dM}{dT} &= \Lambda\, G - M && \text{(3.9b)}\\
\frac{dP}{dT} &= \delta(M - P) && \text{(3.9c)}\\
\frac{d\Delta_{GP}}{dT} &= \delta(\Delta_{GM} - \Delta_{GP}) - \Theta(K_{GPP} + G\Delta_{PP} + \Delta_{GP} + P\Delta_{GP}) && \text{(3.9d)}\\
\frac{d\Delta_{GM}}{dT} &= \Lambda\, G(1 - G) - \Delta_{GM} - \Theta(K_{GMP} + \Delta_{GM} + G\Delta_{MP} + P\Delta_{GM}) && \text{(3.9e)}\\
\frac{d\Delta_{MP}}{dT} &= \Lambda\Delta_{GP} - (\delta + 1)\Delta_{MP} + \delta\Delta_{MM} && \text{(3.9f)}\\
\frac{d\Delta_{MM}}{dT} &= 2\Lambda\Delta_{GM} - 2\Delta_{MM} && \text{(3.9g)}\\
\frac{d\Delta_{PP}}{dT} &= 2\delta(\Delta_{MP} - \Delta_{PP}) && \text{(3.9h)}
\end{aligned}$$

Eqs. (3.9) describe the time evolution of joint cumulants of order 1 and 2 in the form of deterministic differential equations, incorporating the effect of fluctuations through cumulants of second order and higher. These are only the first equations among an infinite hierarchy. When covariances and higher-order cumulants are set to zero, which amounts to neglect fluctuations, we recover the deterministic mean-field model based on averages of copy numbers, which are the first-order cumulants. The important point is that the time evolution of averages depend on covariances, which demonstrates that fluctuations affect the temporal dynamics of averages. In addition, the time evolution of second-order cumulants depends on third-order cumulants K_{GPP} and K_{GMP}. If we wrote the differential equations for third-order cumulants, we would find that they depend on even higher orders, forming a infinite hierarchy of equations for cumulants.

Therefore, Eqs. (3.9) do not form a closed system because of the presence of the third-order cumulants K_{GPP} and K_{GMP}. The standard solution to this problem is to make approximations allowing us to truncate this hierarchy and close it. Several propositions for achieving this have been discussed [7, 134, 86]. The most natural strategy is to set the cumulants of third order to zero, which closes the set of equations. This amounts to represent the probability distributions of copy number variables as Gaussian distributions whose averages and variances are the dynamical variables. This assumption is valid in the limit case where molecular fluctuations are small enough.

In the next section, we will study the steady state of (3.9) in the limiting cases of a fast and of a slow gene, according to the value of the unbinding rate θ. In these two cases, we will be able to truncate easily the hierarchy of equations for cumulants or to solve it analytically. This will allow us to show that fluctuations can severely shift average copy numbers compared to the prediction of the mean-field model.

3.4 Steady state analysis

3.4.1 Fast gene limit case

We first discuss the steady state of the system (3.9) in the simple limit case where the binding and unbinding rates are infinitely large, with $\Theta \to \infty$ keeping the ratio θ/α constant. This expresses the fact that the gene remains in a given state, bound or unbound, only for very short amounts of time, so that the production of mRNA and proteins is little affected during that time. The memory of molecular states is well preserved across one binding-unbinding cycle, so that the average gene activity can reach an equilibrium state with respect to the current protein and mRNA copy numbers, and can be described by a continuous variable varying between 0 and 1. In this limit, fluctuations of mRNA and protein copy numbers are relatively small. Their distribution can be expected to be Gaussian, leading to the third-order cumulants K_{GPP} and K_{GMP} being zero, so that Eqs. (3.9) become closed.

We further assume that $\Lambda \gg 1$, which implies that the mRNA and protein copy numbers are large compared to 1 and thus can be considered as macroscopic variables. The steady state of the cumulant equation system (3.9) is expressed as follows:

$$\begin{aligned}
&G_* \simeq \frac{1}{\sqrt{\Lambda}}(1+\frac{1}{4}\frac{\delta}{\Theta(\delta+1)}) \\
&P_* = M_* \simeq \sqrt{\Lambda}(1+\frac{1}{4}\frac{\delta}{\Theta(\delta+1)}) \\
&\Delta_{GP*} \simeq -\frac{1}{2}\frac{\delta}{\Theta(\delta+1)} \quad \Delta_{MP*} = \Delta_{PP*} \simeq \frac{\Lambda}{2}\frac{\delta}{\Theta(\delta+1)} \\
&\Delta_{MM*} \simeq \frac{\Lambda}{2\Theta}(1+\frac{1}{\delta+1}) \quad \Delta_{GM*} \simeq \frac{1}{2\Theta}(1+\frac{1}{\delta+1})
\end{aligned} \tag{3.10}$$

Note that first-order corrections to the steady-state values of G, M and P only involves a combination of parameters, $\Theta(\delta+1)/\delta$. This shows that the protein degradation rate is also important for controlling the level of fluctuations. Furthermore, the covariance Δ_{GP*} is negative, which results from the negative protein regulation. The covariances Δ_{PP*} and Δ_{MM*} are positive, as expected, since they are the variances of P and M. Eqs. (3.10) indicate that when the unbinding rate Θ is increased to infinity, all covariances tend to zero and the steady state predicted by deterministic model (3.2) is recovered.

In conclusion, third-order cumulants can be neglected in the limit case where the transition between gene activity states is rapid. Then the system (3.9) becomes closed and an analytical expression of the steady state can be obtained. This expression allows us to study how the steady state is affected by fluctuations. Our analytical results agrees well with numerical simulation (Fig 3.3).

3.4.2 Slow gene limit case

Conversely, we now consider the opposite limit case where the unbinding rate is very small: $\Theta \ll 1$. In this limit, the time scale of gene binding and unbinding is much larger than that of other processes. The dynamics of system is simply driven by the gene jumping between "ON" and "OFF" states according to a Poissonian stochastic process. When the gene is unbound,

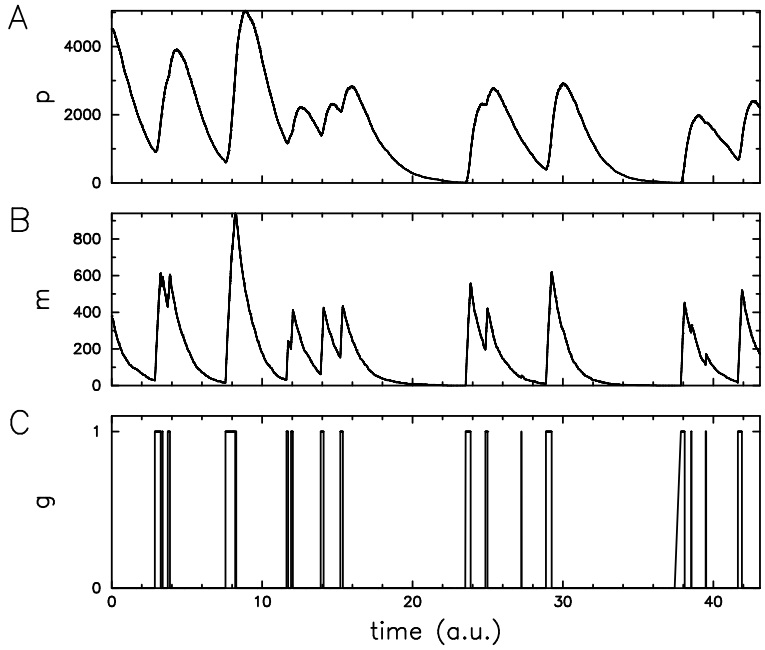

Figure 3.2: **Correlated temporal variations of copy numbers of protein (A), mRNA (B) and gene (C) in the slow gene limit case**. Numerical estimation is carried out with $\delta = 1$, $\theta = 0.5$, $P_0 = 100$, $\Lambda = 200$ and $\beta = 10$. The gene activity is equal to 0 (OFF state) or 1 (ON state). When the gene is in the ON state, mRNA and protein copy numbers increase rapidly. They decrease abruptly when the gene switches to OFF state. The dynamics of system is mainly driven by the transition of the gene between two states, and temporal variations of copy numbers of protein, mRNA and gene are highly correlated.

mRNA and protein copy numbers reach rapidly a high level $M_{ON} = P_{ON} = \Lambda$ which is then maintained. When the gene is bound and transcription is repressed, mRNA and protein copy numbers are abruptly decreased until they reach a low level $M_{OFF} = P_{OFF} = 0$. Once the binding protein dissociates from DNA, an another dynamical cycle repeats. When the system starts a new cycle, it has lost the memory of the previous one. Thus, temporal variations of copy numbers of the gene, mRNA and protein are all characterized by a sequence of spikes, distributed in time according to a Poisson process.

In this limit case, we cannot simply assume that the third-order cumulants are zero, because fluctuations are extremely important. The mRNA and protein copy numbers are at a high level for some time and then are decreased to zero, according to transitions between gene states. However, this simple behavior, without memory from one cycle to the next, makes it easier to determine the steady state. On average, the gene is unbound during an amount of time proportional to $1/(\theta\Lambda)$ and it is repressed during an amount of time proportional to $1/\theta$. Thus we can easily find that the average gene activity is given by $G_* = 1/(1+\Lambda)$. Given the average durations of the two gene states and the mRNA and protein levels in each state, we can compute the values of averages, covariances and third-order cumulants in the limit $\Lambda \gg 1$:

$$\begin{aligned} &G_* \simeq 1/\Lambda; \quad P_* = M_* \simeq 1; \quad \Delta_{PP*} = \Delta_{MM*} = \Delta_{PM*} \simeq \Lambda \\ &\Delta_{GP*} = \Delta_{GM*} \simeq 1; \quad K_{GPP*} = K_{GMP*} \simeq \Lambda \end{aligned} \tag{3.11}$$

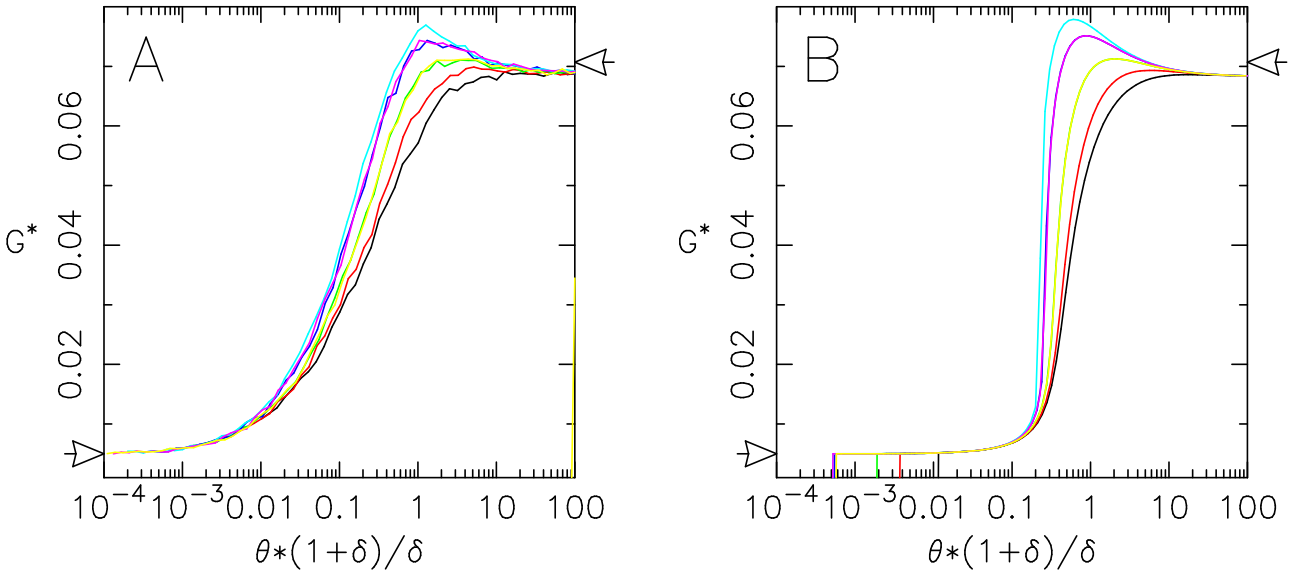

Figure 3.3: **Average gene activities as a function of the reduced parameter** $\Theta(1+\delta^{-1})$**.** (A) Numerical estimation using stochastic simulations is computed with parameter values: $\theta/\alpha = 100$, $\Lambda = 200$, $\lambda = \beta$. Each color corresponds to a given value of δ, which varies from 10^2 (black) to 10 (cyan). (B) Fixed point values of gene activity in model (3.9) with vanishing third-order joint cumulants. Arrows on the left and on the right indicate the gene activity limiting values given by Eqs. (3.10) for large Θ and by Eqs. (3.11) for small Θ.

This expression for the steady state (3.11) is consistent with numerical simulations (Fig 3.3). Compared to the fast gene case, averages (G_*, M_*, P_*) are much lower and decrease with the quantity $\theta(1+\delta)/\delta$. In addition, the values of various cumulants differ very much in magnitude (we assume that Λ is large). This is related to the fact that G, which takes only two values in any case, remains a microscopic variable while M and P become macroscopic variables in the limit of large Λ. Accordingly, averages and covariances involving the microscopic variable G are small compared to their counterparts involving only macroscopic variables and to the third-order cumulants with two macroscopic variables. Note however that even though fluctuations involving the gene are smaller in absolute value, they are more important in relative proportions, as one can see by computing the ratio of the standard deviation (square root of variance) to the average (is proportional to Λ for G and to $\sqrt{\Lambda}$ for M and P).

In order to compare the predicted values for the steady state of Eqs. (3.9) in fast and slow gene limit cases, we have evaluated it numerically using stochastic simulations of the biochemical reaction network. In Fig 3.3A we show the numerical estimation of average gene activities using stochastic simulations. They agree quite well with the values predicted in the limiting cases of a fast gene (large unbinding rate Θ) and of a slow gene (small unbinding rate Θ). We find that the steady state value derived from deterministic rate equations (indicated by the arrow for large Θ) is dramatically shifted when Θ is small and fluctuations are large.

Fig (3.3)B shows the steady state values of gene activity in Eqs. (3.9), computed by assuming that third-order joint cumulants are zero. It can be seen that model (3.9) with vanishing third-order cumulants reproduces well the variation of average gene activity with unbinding rate. The global shape of the curve, with a maximum around $\Theta(1+\delta)/\delta$, is recovered as well as the limiting values for the fast and slow genes. The main discrepancy is a faster transition

from high to low values of gene activity for Eqs. (3.9) than for stochastic simulations.

In summary, we have studied in this section the average behavior of the self-repressing gene circuit using a cumulant expansion of the master equation. In the limit of a fast gene, first-order corrections to the mean-field values can be obtained by assuming that third-order cumulants are zero. This assumption is not valid in the limit of a slow gene, however, simple expressions for averages and covariances can be obtained by noting that mRNA and protein copy numbers are slaved to the gene state. Yet, the average gene activities predicted by cumulant expansion with zero third-order cumulants reproduce surprisingly well these obtained from numerical simulations across the entire range of unbinding rates. Thus the approach followed here allows us to study precisely how average quantities can be dramatically shifted by fluctuations. Further studies are the analysis of instability of the steady state of Eqs. (3.9) and what is the dynamical significance of this instability. We discuss them in the next section.

3.5 Dynamical influence of fluctuations

3.5.1 Stability of steady states predicted by cumulant model

An analysis of a deterministic model of the self-repressing gene by Morant *et al.* had concluded that the steady state of this model is always stable in the case of linear degradation. This model is equivalent to that obtained from cumulant expansion (3.9) when all cumulants of second order and higher orders are set to zero, in other words the mean-field model (3.1). When fluctuations are incorporated into the dynamics by adding higher-order cumulants as dynamical variables, and the associated differential equations, the stability of the fixed point may be modified. Thus, the approach based on the cumulant expansion allows us to study how fluctuations feed back into the deterministic part of the dynamics to induce oscillations of the average quantities.

We first consider the fast gene limit. In this case corresponding to $\Theta \to \infty$, the genetic circuit is subjected to weak fluctuations, and protein and mRNA copy numbers are expected to follow a Gaussian distribution. The third-order cumulants can be neglected, allowing to truncate the hierarchy of equations for cumulants. A linear stability analysis of the fixed point of these equations demonstrates that it is always stable, which is consistent with the results of Morant *et al*. for the mean-field model [138].

The stability analysis of the steady state in the slow gene limit case is more interesting, because the system is then dominated by fluctuations. The stability of the 8 differential equations of the cumulant expansion (3.9) is difficult to study analytically due to the high dimension. For the sake of simplicity, we will analyze this system in the limit where $\Theta\Lambda << \delta, 1$. It then happens that the following equations uncouple from the other equations:

$$\frac{dP}{dt} = \delta\left(M - P\right) \tag{3.12a}$$
$$\frac{dM}{dt} = \Lambda G - M \tag{3.12b}$$
$$\frac{dG}{dt} = \theta\left(1 - G - GP - \Delta_{G,P}\right) \tag{3.12c}$$
$$\frac{d\Delta_{G,P}}{dt} = \delta\left(\Delta_{G,M} - \Delta_{G,P}\right) \tag{3.12d}$$
$$\frac{d\Delta_{G,M}}{dt} = \Lambda G\left(1 - G\right) - \Delta_{G,M}. \tag{3.12e}$$

The reduced system (3.12) consists of only five differential equations describing the dynamics of three averages and two covariances. Its steady state coincides with that of the cumulant expansion (3.9) in the slow gene limit but its stability analysis is much simpler. By further assuming that $\Lambda >> 1$ a Routh-Hurwitz criterion associated to a Hopf bifurcation can be obtained analytically. Thus we find that oscillations appear in the reduced system when

$$\mathcal{H} = 4\Theta^2 + \Theta\left[2\left(1+\delta\right) - \frac{\delta}{\left(1+\delta\right)}\Lambda\right] + \delta < 0. \tag{3.13}$$

The analytical criterion (3.13) indicates that if the unbinding rate is infinitely slow ($\Theta \simeq 0$), the steady state is stable because $\mathcal{H} = \delta > 0$. For intermediate values of Θ, the criterion can become negative provided Λ is sufficiently large, which predicts that the system can display oscillations. However, this can only occur when $\Theta\Lambda \geq \delta + 1$ which conflicts with our preceding assumption $\Theta\Lambda << \delta, 1$. The question raised is thus whether the analytical criterion (3.13) is valid to capture the appearance of oscillatory behavior. In the following part of this section, we therefore check the relevance of this criterion by comparing its predictions with stochastic simulations which are performed according to an implementation of next reaction method (Gibson-Bruck algorithm [81]) mainly by my colleague Quentin Thommen.

3.5.2 Numerical stochastic oscillation

In order to verify whether the criterion (3.13) obtained from the cumulant expansion correctly predicts oscillatory behavior, we will search for parameter values at which the circuit displays a dynamical behavior which can be classified as oscillations. However, it is difficult to characterize oscillations in the stochastic regime because there is no periodicity in a strict sense. We will therefore search for regimes where one observes a series of protein copy number peaks more or less regularly spaced. We thus first need to discuss approaches allowing us to characterize the regularity of such stochastic oscillations in a system subject to fluctuations.

Given a temporal stochastic signal, it is usual to assess the regularity of oscillatory behavior by computing the autocorrelation function of the signal. However, this function is sensitive to variations both in time and amplitude. In our case, it is natural to allow the signal to be highly variable in amplitude while requiring that it is regular in time. It is indeed often the case that biological functions only detect that a signal goes over some threshold regardless of the actual amplitude reached, so that only the peak timing is significant. Therefore, the autocorrelation function may fail to capture the regularity in time of the signal if there are important variations in amplitude.

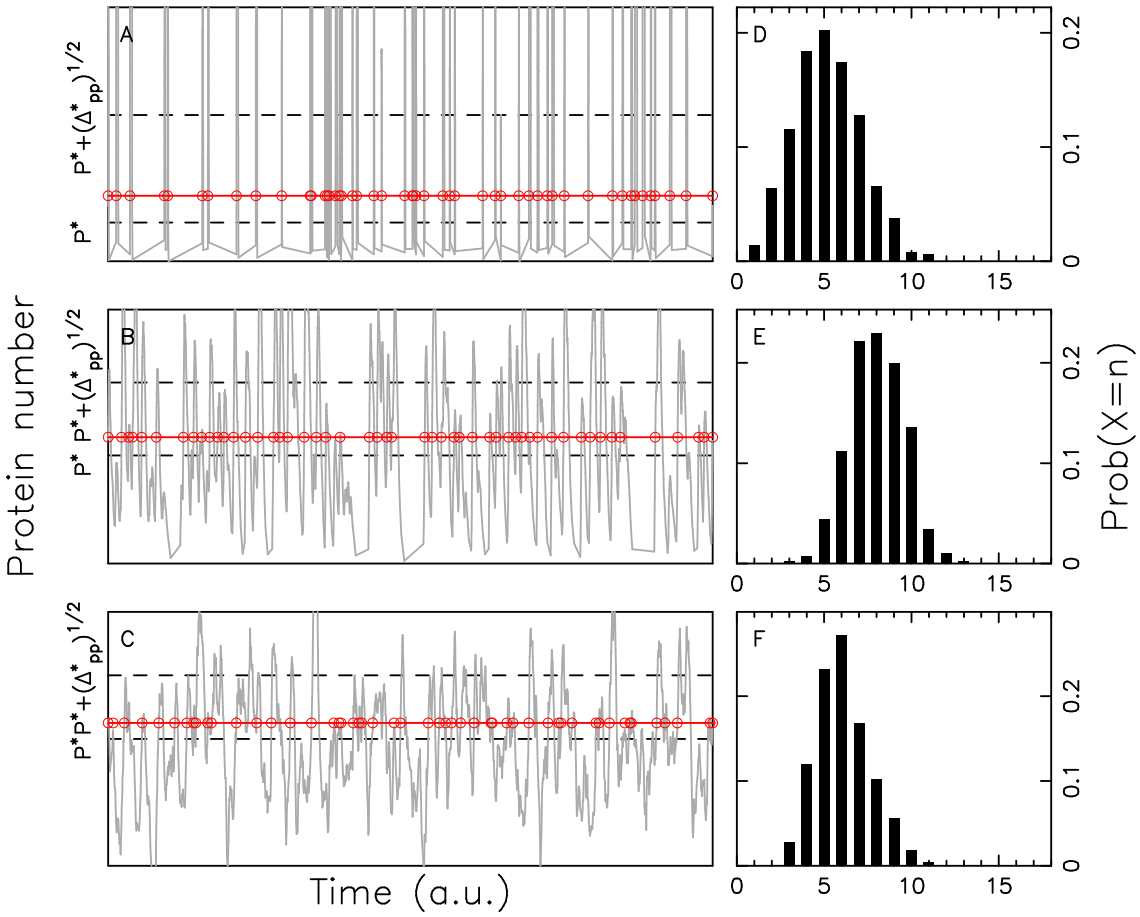

Figure 3.4: **Stochastic dynamical behavior of the self-repressing gene reaction network**. Time evolution of protein copy number for $\Lambda = 200, \delta = 1$ and (A) $\theta = 5.10^{-3}$ (B) $\theta = 0.5$ (C) $\theta = 500$. The dashed lines indicates the mean protein level and mean protein level plus variance. Red lines correspond to the high trigger level and spiking events are indicated by red circles. (D,E,F) Probability of observing n spikes during a given time windows for each of the three regimes.

A standard technique for assessing the temporal regularity of oscillations is to divide the state space into two regions, and to record the times of transition from one region to another [149]. Once a series of transition time is obtained, we can characterize the dynamical behavior by an approach also used to quantify the statistics of photons from a light source [149]. More precisely, we determine the probability to observed n events within a time interval of fixed duration and quantify the temporal regularity by computing the Fano factor which is defined as the ratio of variance to mean. A Fano factor close to unity is often associated with a process governed by a Poissonian statistics, while a Fano factor less (resp., larger) than unity indicates anti-bunching (resp., bunching) of events. Bunching means that peaks tend to cluster and one observes groups of peaks followed by long intervals without peaks. On the contrary, anti-bunching is associated with evenly spaced peaks which remain well separated from each other. Here, we will consider that anti-bunching of protein peaks can be viewed as a stochastic counterpart of deterministic oscillations.

Now we apply this approach to characterize the stochastic oscillatory behavior of the self-

repressing gene circuit. The temporal signal we consider here is the fluctuating protein copy number. We could record the times at which the signal crosses a prescribed level, however, we would get many spurious crossings in sequence because stochastic fluctuations would drive the signal rapidly above and below the threshold. Thus we fix two significantly separated levels given by the average protein level P_* and by $P_* + \sqrt{\Delta_{PP}}/4$ where $\sqrt{\Delta_{PP}}$ is the standard deviation. We record the times where the system crosses the first level, then the second before falling back below the first one. This amounts to require a sufficient excursion above the mean level to count one spike.

Fig 3.4 shows how the series of transition time is recorded (see red cycles) as well as the probability distribution of observing n events within a time interval of given duration for a slow, an intermediate and a fast gene. In the slow gene limit ($\Theta \simeq 0$), the transition events occur independently of protein numbers (Fig 3.4A) and their statistics is Poissonian as indicated by the distribution of the number of event number within a given time window (Fig 3.4D). When the binding rate of gene Θ is intermediate, protein spikes are mostly anti-bunched (Fig 3.4B), and the probability distribution of the number of transitions is Gaussian-like (Fig 3.4E). In contrast, for the fast gene case $\Theta \gg 1$, the regularity of stochastic oscillations is degraded (Fig 3.4C and F). We compute the Fano factor for the slow, the intermediate and the fast gene in Fig 3.4. The Fano factor of the slow gene is close to unity, confirming that the transition events is Poissonian. The relative regularity of transitions for the intermediate gene is well captured by the Fano factor that is around 0.35. The Fano factor for the fast gene rises again to about 0.5. Therefore, we observe a resonance-like effect where the protein signal is most regular when the gene response time is close to a finite scale. As we show now, this resonance is also controlled by the protein and mRNA lifetimes, which determine the time during which a memory of previous gene states persists.

We have studied systematically how the Fano Factor depends on the gene unbinding rate Θ and the relative protein decay rate δ in stochastic simulations of the self-repressing gene circuit. As Fig (3.5) shows, the regularity of protein spikes is enforced when (1) the decay rates δ_p and δ_m are comparable ($\delta \sim 1$) and (2) the reduced parameter $\Theta * (1 + \delta^{-1})$ is close to unity. Quite remarkably, the parameter space region where protein spikes are more regularly spaced superimposes very well with the region where the reduced model (3.12) displays deterministic oscillations. This suggests that this model captures well the dynamical interaction of mean-field variables and fluctuations, although it possesses only the average gene activities of the cumulant expansion (3.9) in the slow gene limit. This probably indicates that the dynamically important joint cumulants are those involving the gene state variable. This is not surprising given that gene state remains binary in all limits and is thus the most stochastic variable.

To conclude this section, we have quantified the regularity of stochastic oscillations in the self-repressing gene circuit using the Fano Factor. This allowed us to show that the region where protein spikes are more evenly spaced is well predicted by the oscillation region of a reduced model determined from the cumulant expansion.

3.6 Conclusion

To investigate the dynamical influence of fluctuations on biological oscillations, we developed a cumulant expansion of the master equation that describes a minimal genetic circuit consisting of a self-repressing gene. We have found that fluctuations significantly shift the average values

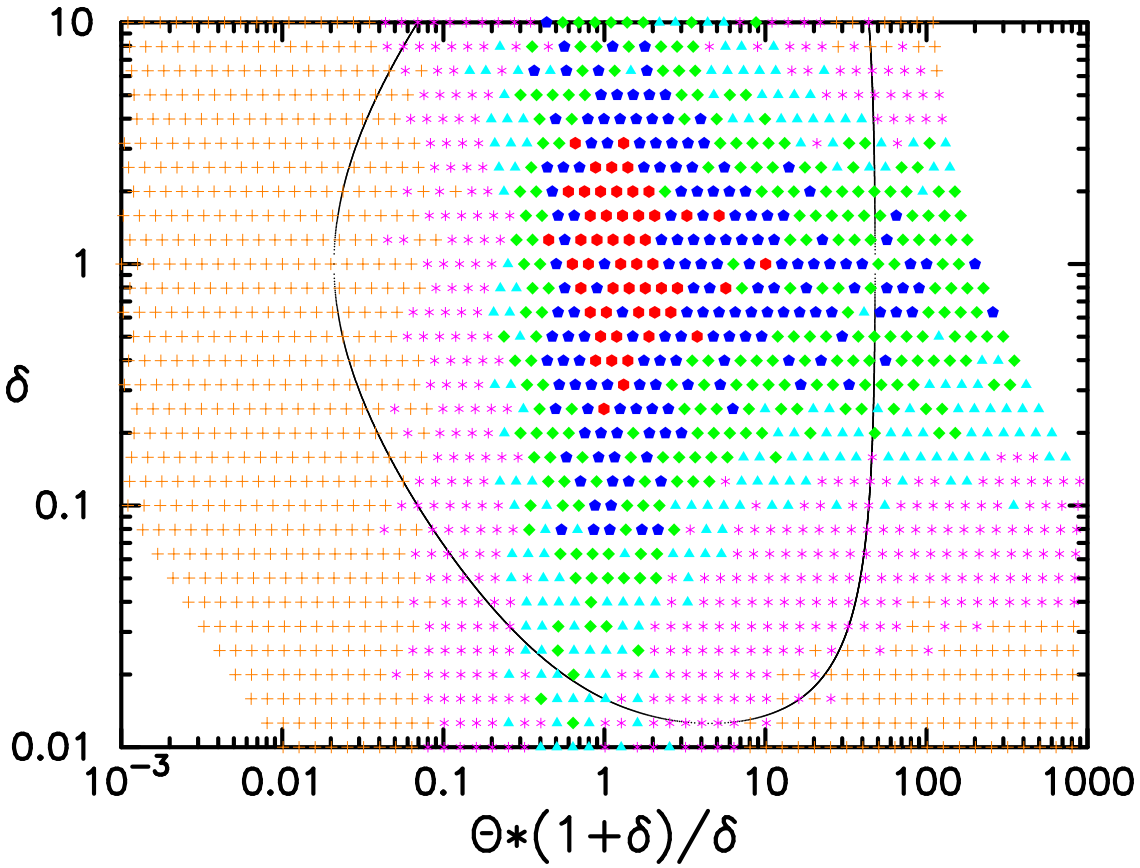

Figure 3.5: **Stochastic oscillations characterized by Fano factor**. Dependence of the Fano factor F quantifying spiking regularity on δ and $\Theta*(1+\delta^{-1})$. Stochastic simulations of genetic circuit have been carried out with $\Lambda = 200$, $\beta = \lambda$, $\theta/\alpha = 100$. Different values of Fano factor are indicated by red hexagon ($F < 0.35$), blue pentagons ($0.3 < F < 0.4$), green diamonds ($0.4 < F < 0.45$), cyan triangles ($0.45 < F < 0.5$), magenta stars ($0.5 < F < 0.7$) and orange crosses ($0.7 < F$). The black line encloses the region where the analytical criterion (3.13) predicts oscillations.

compared to those predicted by deterministic rate equations, and that this is well approached by the cumulant expansion. We also have studied the stochastic time evolution of this circuit and characterized the regularity of protein spikes using a Fano-like indicator, showing the parameter region with most regular protein spikes can be obtained from an analytical oscillation criterion derived from cumulant expansion. Together, our results demonstrate how fluctuations can induce oscillatory behavior in a simple genetic circuit. Our approach to characterize the regularity can also be applied for other biological problems or even experimental signals, for example, the mRNA or protein signals of circadian genes.

Chapter 4

Dynamical effects of stochastic RNA Polymerase pausing on transcription

4.1 Introduction

Transcription is one of the most complex and tightly regulated processes in gene expression [106, 192, 76, 18, 178]. It consists in transcribing the genetic information stored in DNA molecules into mRNA molecules, and is carried out by macromolecules called RNA polymerase (RNAP). Transcription can be essentially divided into 3 different steps: initiation, productive elongation and termination. RNAP and transcription factors (TFs) are recruited by promoter during initiation. Then the assembly of RNAP and TFs leaves the promoter and enters into productive elongation in which RNAP moves forwards along the DNA sequence in stochastic single-nucleotide steps and produces the mRNA transcript. At termination, RNAP releases the complete mRNA transcript and dissociates from the DNA sequence.

Specific dynamical behaviors of RNAP procession along the DNA template during productive elongation can highly modulate transcription rate. This includes pausing [31, 62, 67], which is a phenomenon where the RNAP is halted at a nucleotide and is widely observed in both prokaryotes and eukaryotes [10, 70, 118, 57, 54, 56]. RNAP pausing occurs in a stochastic fashion and is spontaneously reversible, after which the RNAP continues to move forwards along the DNA template. The majority of pauses are short with a period of ≈ 1 s in average, with weak sequence dependence, and they are not affected by trailing RNAPs. They are usually referred to as "ubiquitous" or "elemental" short pauses [56, 107, 119, 55, 57], to distinguish them from the so-called prolonged backtracking pauses [55, 57, 4] which are rare (less than $5\% - 10\%$ of all pauses). Prolonged pauses only occur at specific sites and their duration is usually over 20 s. In contrast to elemental short pauses, prolonged pauses can be suppressed by the pushing of trailing RNAPs. In addition, prolonged pauses involve the backtracking phenomenon [197, 171, 78, 48, 113, 49, 15, 50, 192] in which RNAPs move backwards and may proofread the sequences of copied mRNA [63]. Recent investigations show transcriptional pausing plays important functional roles in coordination and accuracy of gene expression [51, 45, 1, 61]. However, the question as to how RNAP pausing affects transcription dynamics remain unclear.

In this chapter, we focus on elemental short pauses. In order to investigate their influences on the dynamics of transcription, we will use a model well studied in out-of-equilibrium statistical physics, a driven lattice gas in one dimension, known as the Totally Asymmetric Simple

Exclusion Processes (TASEP) [115, 112, 29, 28, 34, 30, 117, 27, 153]. This TASEP model is commonly used to study the different dynamical aspects of transcription [62, 164, 31, 109] and translation [23, 17, 161, 172, 24]. However, in most of these studies, the TASEP model is analyzed under the mean-field approximation which neglects the particle correlations and simplifies computations. Meanwhile, it is shown that, such as in [23], when the particle correlations in the model become nonneglectable, the mean-field approximation biases the results of interest. In particular, Klumpp et al. [109] found that RNAP pausing seriously increases the correlation between RNAPs by numerical simulations. Therefore, a more genelized approache is necessarily needed for understanding the dynamical role of RNAPs pausing. Here we will propose a statistical approach to study the TASEP model and provide insights into the parameter space where the mean-field approximation are not able to be applied.

This chapter is organized as follows: first, we will explain how elemental pauses can be modeled by a TASEP system. We will show that the dynamics of the system can be classified into three regimes according to the dependence of transcription rate on pause duration: short pause case where mean-field approximation works and an analytical transcription rate is given, intermediate and long pause cases where mean-field approach fails. Then we suggest a statistical approach in which transcription rate is contributed by all system configurations characterized by the number of paused particles. After that, we will apply this approach and obtain expressions of transcription rate for short pause and long pause regimes. At last we discuss our understanding of intermediate pause regime and give a general expression of transcription rate for three regimes.

4.2 Mean-field approach and its limitation

First, we explain how to model elemental pauses using the TASEP model. We remind that elemental pauses occur in a stochastic fashion with weak sequence dependence and are not affected by the pushing of trailing RNAPs. These important features allows us to consider an homogeneous system. In TASEP model, the gene sequence is modeled as a one-dimensional discrete lattice of N sites; RNAPs are modeled as particles hopping from site to site with a constant probability ϵ determining the elongation rate, only if the next site is not occupied, to take into account the fact that RNAPs can not overlap. For the sake of simplicity, we will neglect the fact that RNAP footprint may extend over several sites, and we moreover assume that the occurrence of pausing is completely random at all sites. RNAPs thus are modeled to have two states in each site: active or paused. In the active state, RNAPs may move forwards along the lattice provided the next site is not occupied by another RNAP, or they may enter into the paused state with a frequency f. RNAPs remain in paused states for an average time of τ before returning the active state (Fig 4.1).

We now indicate the orders of magnitude of some biological parameters involved in our model. The frequency of elemental pauses is about $0.1s^{-1}$ and the duration is $\approx 1s$ [56, 107, 119, 55, 57]. The average elongation rate ϵ in vivo is found in the range of $20 - 80$ nt/s [191, 39]. In fact, these rate measured are already slowed down by the pauses, so that the elongation rate without pauses should be much larger. In our model, we choose an elongation rate given $\epsilon = 100s^{-1}$ as in [109].

We recall that the fundamental quantity that we will study is the particle current. The current is the mean flow of particles across the lattice, that is, the average number of particles

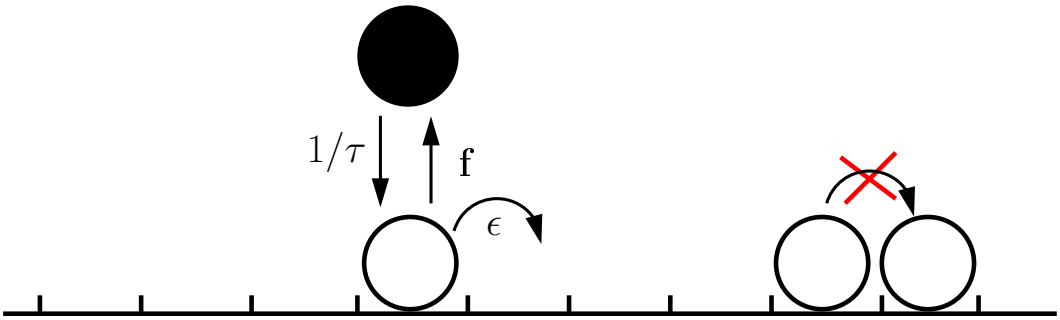

Figure 4.1: **Model of transcription and RNAP pausing**. The size of RNAP footprint is neglected. Active RNAPs (white circles) hop from site to site by stochastic single-nucleotide steps with elongation rate ϵ provided the next site is not occupied. Active RNAPs can switch to an paused state (black) with a pause frequency f, and remain τ amount of time in average before returning to the active state. Figure modified from [109].

going through a site per unit time. Since it can be interpreted as the average number of RNAPs going through the sequence per unit time, the current corresponds to the transcription rate. Our goal is to understand how pauses affect the transcription rate and by how much. In steady state conditions, the current is the same at each site so that a global transcription rate can be defined.

Under the mean-field approximation where the probability of one site being occupied does not depend on that of its next, the time evolution of active and paused states at site i can be described by the following differential rate equations:

$$\frac{da_i}{dt} = \epsilon\, a_{i-1}(1-a_i-p_i) - \epsilon\, a_i(1-a_{i+1}-p_{i+1}) - f a_i + \frac{1}{\tau}p_i \tag{4.1a}$$

$$\frac{dp_i}{dt} = f a_i - \frac{1}{\tau}p_i \tag{4.1b}$$

where a_i and p_i are respectively the density of particles in active and paused states in site i. ϵ is the elongation rate. The term $\epsilon\, a_{i-1}(1-a_i-p_i)$ describes the arrival at site i of active particles coming from site $i-1$ (also named incoming current) and $\epsilon\, a_i(1-a_{i+1}-p_{i+1})$ describes the loss of active particles leaving from site i to $i+1$ (outgoing current). In steady state conditions, the incoming and outgoing currents are the same at each site, so that a global current can be defined. $f a_i$ is related to active particles entering into paused state, and $\frac{1}{\tau}p_i$ to paused particle becoming active.

In steady state, the densities of active and paused particles can be expressed by:

$$a_i = \frac{1}{f\tau}p_i = \frac{1}{1+f\tau}\rho_i \tag{4.2}$$

where $\rho_i = a_i + p_i$ is the total density for site i. The current (transcription rate) of the TASEP model depends on boundary conditions [115, 112, 29, 28], which can be open or periodic. With open boundary conditions, particles enter the one-dimensional lattice with a constant probability α representing the initiation rate and exit from the end of lattice with termination rate β. Likewise, for simplicity, we can use the TASEP system with periodic boundary conditions (also called ring TASEP) which is a uniform and closed system. The total number of particles is conserved and all sites are identical because of symmetry. As recalled in Chapter 1, the

current of the TASEP without pauses is correctly predicted by mean-field approximation and is given by

$$J = \epsilon\rho(1-\rho)$$

where ρ is RNAP density for all sites. So the current is proportional to the elongation rate and reaches the maximum value for the RNAP density of 0.5; For small density, transcription rate is low because there are few particles; when the density becomes large, the probability that the site next to a particle is occupied is large so that the current is reduced by particle exclusion.

We thereafter adopt the TASEP system with periodic boundary condition to study the influence of RNAP pausing. By incorporating RNA pausing, the TASEP system is still uniform because the occurrence of an elemental short pause does not depend on location. Although the TASEP model with periodic condition has little resemblance to the transcription process, it is sufficiently simple for us to derive some analytical results. Their results will provide us with crucial insights into the influence of RNAP pausing, such as how the current-density relation is modified.

In the TASEP system with periodic boundary condition, all sites are identical, and thus $\rho_{i-1} = \rho_i = \rho$. The expression of current in the mean-field approximation, obtained by solving Eqs. (4.1) and recognizing that $J = \epsilon a(1-\rho)$ (i.e., there can be an advance if site is occupied by an active particle and next site is empty) becomes:

$$J = \frac{\epsilon}{1+f\tau}\rho(1-\rho) \tag{4.3}$$

Eq. (4.3) predicts the transcription rate taking into account the RNAP pausing under mean-field approximation. Note that it depends only on the product of pause frequency f and pause duration τ. Compared with the case without pausing, the transcription rate is reduced by RNAP pausing with a factor of $\frac{1}{1+f\tau}$, which results from the fact that only the particles in active state contribute the current. So the TASEP model with pausing can be mapped to the TASEP model without pausing by changing the elongation rate.

In order to assess whether the mean-field approximation predicts correctly the transcription rate, we will compare Eq. (4.3) with the numerical stochastic simulation. Here we specify that in this chapter, all numerical simulations are performed according to Gillespie Algorithm [82].

Before that, we normalize rate constants of TASEP system with respect to the elongation rate ϵ, so that the new expression of current is (4.3) with $\epsilon = 1$. By using the Gillespie Algorithm [82], the current of the TASEP model with pausing and periodic boundary conditions is numerically estimated for values ranges of pause duration τ. Moreover, the value of $f\tau = 0.1$ is fixed, so that the fraction of paused particles remains constant and variations in current will only be due to correlations between particles. If the mean-field approximation was valid across the entire range of parameter values, the value of the current should be constant.

As Fig 4.2 shows, the expression (4.3) derived under mean-field approximation predicts well the current when the pause duration τ is smaller than the characteristic time scale of elongation which is here normalized as one (this regime we call small-tau regime). In this regime, pausing only affects the residence time of particles at each site and does not induce traffic jams, thus the mean-field approximation is valid. However, when the pause duration τ becomes comparable to or larger than the characteristic time scale of elongation (called intermediate-tau regime), a paused particle can remain in one site for enough time to block other particles coming behind it, and traffic jams are observed frequently in the lattice. Correlations between

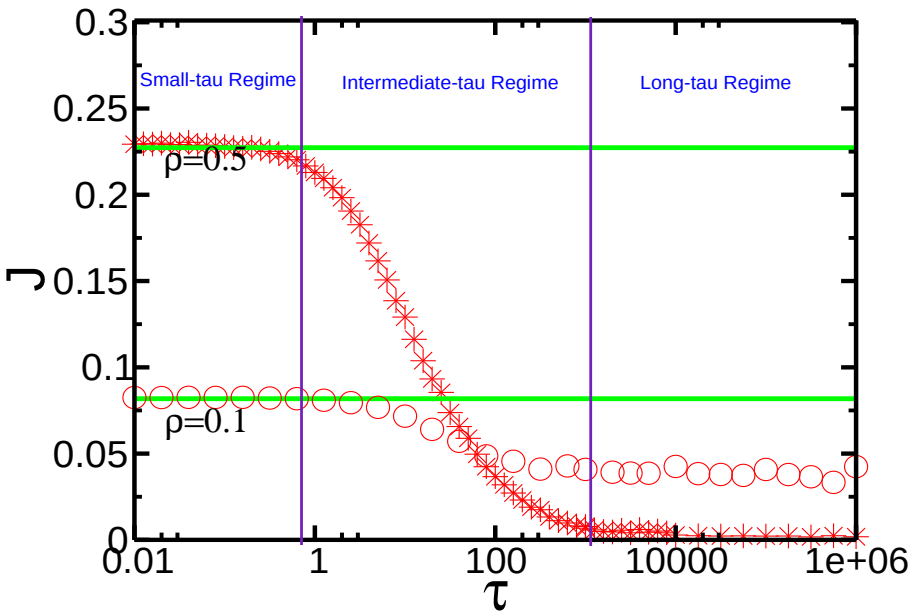

Figure 4.2: **Transcription rate as a function of pause duration τ with $f\tau$ constant**. By using a TASEP model with periodic boundary condition comprising $N = 100$ sites, transcription rates (red stars and circles) are numerically computed by the Gillespie algorithm [82] for RNAP densities $\rho = 0.1$ and $\rho = 0.5$, with $f\tau = 0.1$ being kept constant. According to Eq. (4.3), the transcription rate predicted by mean-field approximation is constant at fixed $f\tau$ and is represented by green curves. However, according to the numerical simulations, we distinguish three different dynamical regimes: the short pause (small τ) regime where mean-field approximation works well, intermediate pause (intermediate τ and long pause (large τ) regimes in which pausing reduces more or less severely transcription rate. Biologically relevant parameter values are found in the intermediate regime.

particles become important and significantly affect the transcription rate. The current is decreased rapidly. In the extreme case where τ is very large (long-tau regime), the dynamics of system is mainly driven by pausing, and the system seems to reach to another special regime, as the current does not depend on the pause duration and becomes likely constant again.

So according to the dependence of transcription rate on the pause duration, we classify the dynamical behavior of the system into three different regimes: short-pause, intermediate-pause and long-pause regimes. The mean-field approximation works only for the short-pause regime. An important question is to which regime the range of biological parameters does correspond. Since pause frequency and duration are normalized as $f = 0.001, \tau = 100$, they are clearly in the intermediate-tau regime. However, transcription and translation modeled by TASEP systems are usually theoretically studied via the mean-field approach, which works only for short pauses. Hence the study of the intermediate-pause and long-pause regimes offers particular interest despite of their mathematical difficulties. In the next section, we will propose a simple statistical approach to compute the transcription rate as a function of pause duration.

4.3 Probability of configurations associated with pause number

In order to compute the transcription rate curves of Fig 4.2, we have considered a finite-size TASEP model (number of particles and sites are both finite) because of intuitive consideration and mathematical simplicity. However, commonly encountered genes feature usually thousands or even millions of nucleotides, so that it is necessary to consider also the infinite-size limit (where the numbers of particles and sites both tend to infinity), not only for its physical interest, but also to take into account the fact that dynamical behavior may be dramatically changed during the transition from finite size to infinite size. We explain the statistical approach first for finite-size TASEP system and then in the infinite-size limit.

In a finite-size TASEP system with periodic boundary condition where C_0 particles move forwards along N discrete sites, we can classify configurations according to the number n of paused particles. To be specific, if there are no paused particles on the TASEP in ring, we define it as a zero-pause configuration; if all particles are paused, we define it as C_0-pause configuration. So there are in total $C_0 + 1$ pause configurations. As the system evolves in time, it will undergo transitions between different configurations, spending a certain amount of time in each configuration. In each configuration, there will be a contribution to current. For example, the contribution to current of the zero-pause configuration is $\rho(1-\rho)$, like the TASEP model without pauses. Note that there is a contribution to current, so long as there are active particles with empty sites in front of them so that they can advance. Even though paused particles occupy some sites, current may be still contributed so long as there are active particles whose next sites are empty (Fig 4.3).

Therefore, not only the zero-pause configuration contributes to current but also all configurations containing paused particles except of course the C_0-pause configuration, where there is no active particle. However, the contributions to current of different configurations are generally very different. We assume that the total current of the finite-size TASEP can be expressed for any value of τ as a sum over configurations with specific number of pauses:

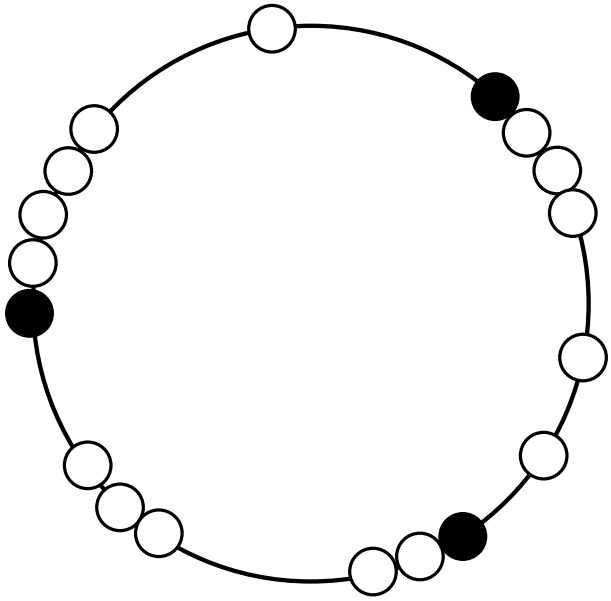

Figure 4.3: **Current contribution of configurations with constant number of paused particles**. In TASEP model of finite size with periodic boundary condition, C_0 particles are distributed in a ring of N discrete sites. Even though paused particles occupy some sites and block the movement of some particles behind them, there are still contributions to current so long as there are empty sites in front of active particles.

$$J_T = \sum_{n=0}^{C_0} P_n J_n \tag{4.4}$$

where P_n is the probability of n-pause configuration, and J_n is its contribution to current. Two extreme cases are already known: $J_0 = \rho(1-\rho)$ and $J_{C_0} = 0$. Eq. (4.4) involves two quantities, the probability P_n of n-pause configuration and the associated contribution to current J_n. Interestingly, P_n is independent of configuration of occupied sites, in other words, P_n is simply equivalent to the probability of having n paused particles in TASEP model. Next we analyze how to compute the probability of n-pause configuration in the finite-size system.

4.3.1 Probability of configurations with pauses in the finite-size system

Transitions between different configurations occur with some fixed rates so that we can expect that there is a well-defined probability distribution of observing a given configuration at some time. Here we write the master equation [189] that describes the time evolution of probabilities for these configurations as follows:

$$\frac{dP_0}{dt} = \frac{1}{\tau}P_1 - fC_0P_0 \tag{4.5a}$$
$$\frac{dP_1}{dt} = fC_0P_0 - \frac{1}{\tau}P_1 + \frac{2}{\tau}P_2 - f(C_0-1)P_1 \tag{4.5b}$$
$$... = ... \tag{4.5c}$$
$$\frac{dP_n}{dt} = f(C_0-(n-1))P_{n-1} - \frac{n}{\tau}P_n + \frac{n+1}{\tau}P_{n+1} - f(C_0-n)P_n \tag{4.5d}$$
$$... = ... \tag{4.5e}$$
$$\frac{dP_{C_0}}{dt} = fP_{C_0-1} - \frac{C_0}{\tau}P_{C_0} \tag{4.5f}$$

Eqs. 4.5 are basic rate equations which describe the time evolution of probability P_n depending on the probabilities P_{n-1} and P_{n+1} to be in neighboring configurations and on the transition rates between these configurations. The steady state of Eqs. 4.5 gives the stationary probabilities. Note that equations defining the steady state give the same solution as the principle of detailed balance would give (the probability of transition between n and $n+1$ pauses is the same as for the reverse transition). By computing the fixed point, we find the following expressions for the probabilities P_n.

$$P_0 = \frac{1}{(1+f\tau)^{C0}} \tag{4.6a}$$
$$P_n = (f\tau)^n C_{C0}^n P_0 \tag{4.6b}$$

where $C_{C_0}^n$ is the binomial coefficient. Note that the probability P_n of having n pauses is independent of the pause duration τ and depends only on the parameter product $f\tau$. This expression can also be recovered as follows. There are C_0 particles, and for each of those, the probability of being in pause is $\frac{f\tau}{1+f\tau}$. The probability to have n paused particles is given by the binomial distribution $B(C_0, \frac{f\tau}{1+f\tau})$, which is indeed the same solution as in Eqs. (4.6).

So in the finite-size TASEP system, the probability of having n pauses is simply the Binomial distribution for all three regimes. We next study the pause number probability distribution in the infinite-size limit.

4.3.2 Probability of configurations with pauses in the infinite-size limit

For TASEP system in the infinite-size limit, the total number of particles is infinite as well as the number of sites, and we must use different parameters to describe the system. We define an intensive continuous parameter, which is the fraction $\tilde{n} \in (0,1)$ of paused particles:

$$\tilde{n} = \frac{n}{C_0} \tag{4.7}$$

In this limit, the configuration is defined according to the fraction of paused particles. We analyze now how to find the probability distribution function (PDF) for these configurations. In the finite-size TASEP system, the probability of the n-pause configuration is simply the Binomial distribution written as $B(C_0, p)$ with $p = \frac{f\tau}{1+f\tau}$. When the total number of particles becomes sufficiently large, an excellent approximation to the Binomial distribution is given by

the continuous Normal distribution $N(C_0p, C_0p(1-p))$ with $p = \frac{f\tau}{1+f\tau}$. Based on Eq. (4.4), the expression of the total current for the case where number of particles is large can be written:

$$\begin{aligned} J_T &= \int_0^{C_0} P(n)J(n)dn \\ &= \int_0^{C_0} \frac{1}{\sqrt{2\pi}\sqrt{C_0p(1-p)}} e^{-\frac{(n-C_0p)^2}{2C_0p(1-p)}} J(n)dn \\ &= \int_0^1 \frac{1}{\sqrt{2\pi}\sqrt{\frac{p(1-p)}{C_0}}} e^{-\frac{(\tilde{n}-p)^2}{2\frac{p(1-p)}{C_0}}} J(\tilde{n})d\tilde{n} \\ &= \int_0^1 P(\tilde{n})J(\tilde{n})d\tilde{n} \end{aligned} \tag{4.8}$$

where

$$P(\tilde{n}) = \frac{1}{\sqrt{2\pi}\sqrt{\frac{p(1-p)}{C_0}}} e^{-\frac{(\tilde{n}-p)^2}{2\frac{p(1-p)}{C_0}}} \tag{4.9}$$

with $p = f\tau/(1+f\tau)$ is the probability of having the configuration for which the fraction of paused particles is between $\tilde{n}$ and $\tilde{n}+d\tilde{n}$, and $J(\tilde{n})$ is the contribution to the current associated with that configuration. The distribution (4.9) possesses an average of p and a variance of $p(1-p)/C_0$ (Fig 4.4). When increasing gradually the total number of particles, the shape of PDF becomes narrower and narrower, and the maximum centered at the average becomes higher and higher. In the infinite-size limit where the total number of particles C_0 is infinite, the variance is zero and the probability density function is the Dirac function centered at the average of the fraction of paused particles $\langle\tilde{n}\rangle = f\tau/(1+f\tau)$. Thus, the expression of total current in the infinite-size limit becomes:

$$\begin{aligned} J_T &= J(\langle\tilde{n}\rangle) \\ &= J(\frac{f\tau}{1+f\tau}) \end{aligned} \tag{4.10}$$

Therefore, we find that in the infinite-size limit, the total current is generated by the configuration with a fraction $f\tau/(1+f\tau)$ of paused particles. This represents a dramatical change for the PDF of configurations as well as for the expression of total current compared with the finite-size case.

In this section, we have established the probability of configurations corresponding to a given number of fraction of paused particles. To understand the variation of transcription rate in the three regimes, determining the contribution to current of different configurations will be our aim in the following sections. Our strategy is first to study the short-pause and long-pause regimes for both finite-size and infinite-size systems, and then to study the intermediate-pause case which is actually the transition between the two above limit cases.

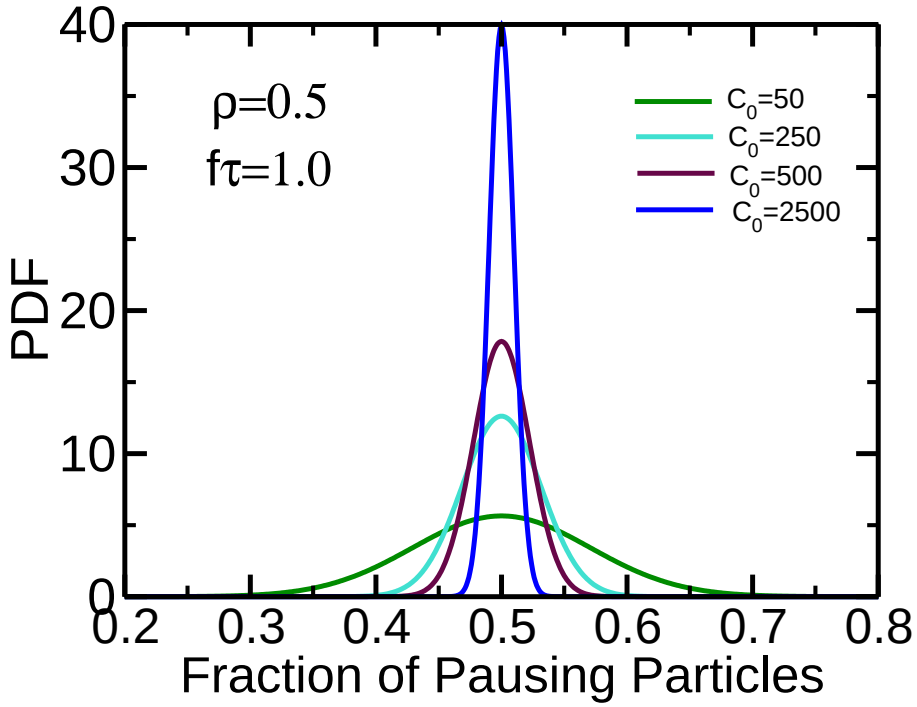

Figure 4.4: **Probability density function (PDF)** $P(\tilde{n})$ **of configuration with the fraction of paused particle between between** $\tilde{n}$ **and** $\tilde{n} + d\tilde{n}$. When the size of TASEP model increases gradually, this Normal distribution becomes narrower and narrower and the maximum value of PDF is approaching to infinite. In the limit of infinite size, PDF turns to be the Dirac function.

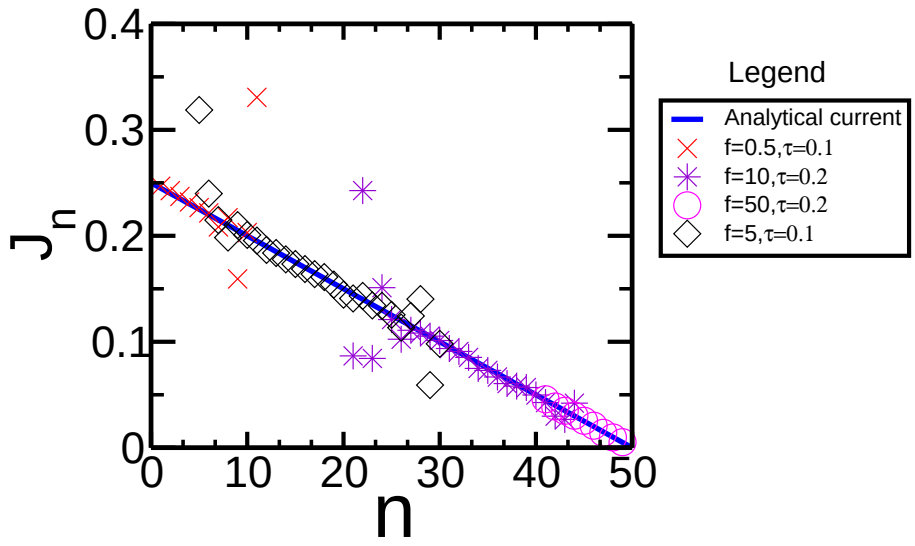

Figure 4.5: **Contribution to current as a function of the number of paused particles in the finite-size TASEP model**. The blue solid curve is obtained by Eq. (4.11). Different colored symbols represent numerical estimations for fixed parameters values $N = 100, \rho = 0.5$. Certain estimations deviate from the curve because of statistical fluctuations, since these configurations have lower probabilities. Therefore, the expression (4.11) predicts well the contribution to current of configurations.

4.4 Transcription rate in short pause limit

4.4.1 Finite size system

We first seek to compute the contribution to current J_n of n-pause configuration for the finite size system in the short pause limit, where the mean-field approximation still works. As the pause duration is sufficiently short, particles are independent of each other so that the active particles are not able to sense the presence of paused particles. The movement of active particles generates the current, so intuitively the current contribution J_n is proportional to the number of active particles in the configuration, with $J_0 = \rho(1-\rho)$. The expression of J_n is written as follows:

$$J_n = \frac{C_0 - n}{C_0}\rho(1-\rho) \tag{4.11}$$

where the term $\frac{C_0-n}{C_0}$ is the probability for a particle to be active (which is actually the fraction of active particles), so that $\frac{C_0-n}{C_0}\rho$ is the density of active particles for one site; $1-\rho$ is the probability that the next site is not occupied by another particle. Eq. (4.11) first confirms two extreme cases: $J_0 = \rho(1-\rho)$ and $J_{C_0} = 0$. In addition, J_n as a function of number of paused particle n, does not depend on pause duration τ and not on the parameter product $f\tau$, either. As Fig 4.5 shows, the expression of the contribution to current of configurations given by (4.11) agrees well with numerical simulations.

Combining our expression for J_n deduced in short pause limit with the probability P_n of the n-pause configuration in the general expression (4.4) yields the following expression of total current:

$$\begin{aligned} J_T &= \sum_{n=0}^{C_0} P_n J_n \\ &= \sum_{n=0}^{C_0} \frac{(f\tau)^n C_{C0}^n}{(1+f\tau)^{C0}} \rho(1-\rho) \frac{C_0 - n}{C_0} \\ &= \frac{\rho(1-\rho)}{1+f\tau} \end{aligned} \tag{4.12}$$

Thus we recover the expression of total current derived from mean-field approximation (Eq. 4.3), regardless of the size of TASEP model. Next, we will search for the total current in short pause regime for the infinite-size limit.

4.4.2 Infinite size limit

In the infinite size limit, $J(\tilde{n})$, the contribution to current of the configuration with a fraction $\tilde{n}$ of paused particles can be obtained from the finite-size expression (4.11):

$$J(\tilde{n}) = (1-\tilde{n})\rho(1-\rho) \tag{4.13}$$

where $(1-\tilde{n})$ is the fraction of particles which are active. According to the relation (4.10), only the configuration with the fraction of paused particles given by $\tilde{n}\frac{f\tau}{1+f\tau}$ contributes to the total current. By substituting $\tilde{n} = \frac{f\tau}{1+f\tau}$ in Eq. (4.13), we obtain:

$$J_T = \frac{\rho(1-\rho)}{1+f\tau} \tag{4.14}$$

In conclusion, we have studied the short pause regime and found the same expression of the current for the finite-size and infinite-size TASEPs, the one predicted by mean-field approximation. Thus, the transcription rate does not depend on the size of system. In the next section, we will investigate another particular limit, that of very long pauses.

4.5 Transcription rate in the long pause limit

4.5.1 Typical dynamical behavior

To investigate the behavior of the TASEP model with pauses under periodic boundary conditions in the long pause limit, we again first consider the finite-size case, for the sake of mathematical simplicity. We first discuss qualitatively the typical dynamical behavior of TASEP model in this regime, as it will give us some inspiration to derive the expression of current. When the average pause duration is sufficiently large, a paused RNAP may block several RNAPs and force them to wait for a certain account of time, during which they will probably enter into paused state themselves. We define the ensemble of a paused particle in head and its following particles, forming a contiguous block, as a cluster. In the long pause limit, clusters of large size can easily be formed. In consequence, the particles are highly correlated. As particles confined in clusters cannot produce current by moving forwards until the moment when

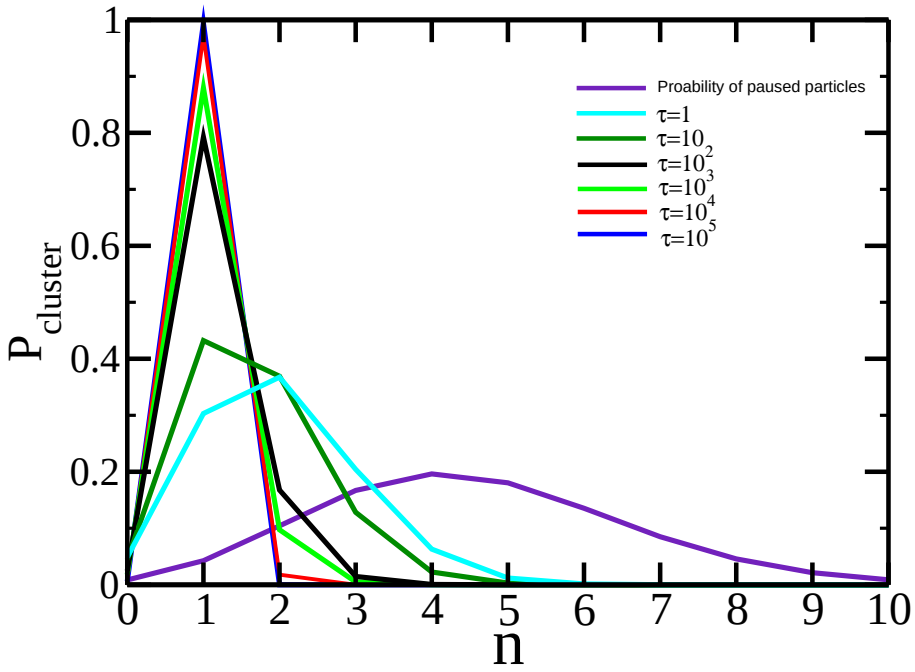

Figure 4.6: **Typical dynamical behavior in the long pause limit**. The probability distributions of the number of clusters are shown for different values of the pause duration τ, other parameters being kept constant at $N = 100$, $\rho = 0.5$ and $f\tau = 0.1$. A cluster is defined as a collection of contiguous particles with a paused particle in head. In the long pause limit, the most probable configuration is a single cluster.

clusters collapse or particles escape from them, the sizes and numbers of clusters significantly determine the current.

We have computed the number of clusters in our TASEP model when pause duration is gradually increased (Fig 4.6). When the pause duration τ is small or intermediate, several clusters coexist, whereas in the long pause regime, only one huge cluster is found along the TASEP ring. This implies that all paused particles are confined inside of this single cluster (Fig 4.7). For most of the time, all particles, active or paused, are confined in this cluster, and no contribution to current is observed. When the paused particle in head becomes active, it is released from the cluster and moves forwards without any exclusion, because all sites before it are empty. It is quickly followed by the active particles behind it which can also move rapidly. When all active particles at the head have "evaporated", another paused particle becomes the head of the cluster, particles behind the new head remaining immobile. As the round-trip time of moving particles is much shorter than the pause duration of the new head particle, released particles go around the ring and join the cluster again at its other end. After this transient flow of particles, we are back to the single immobile cluster configuration. Therefore, the only contributions to current come from active particles released from the single cluster after the paused particle in head returns to active state.

So the typical dynamical behavior for the finite-size system in the long pause limit is one-cluster dominating, which provides us an important clue to derive the current contribution of n-pause configuration.

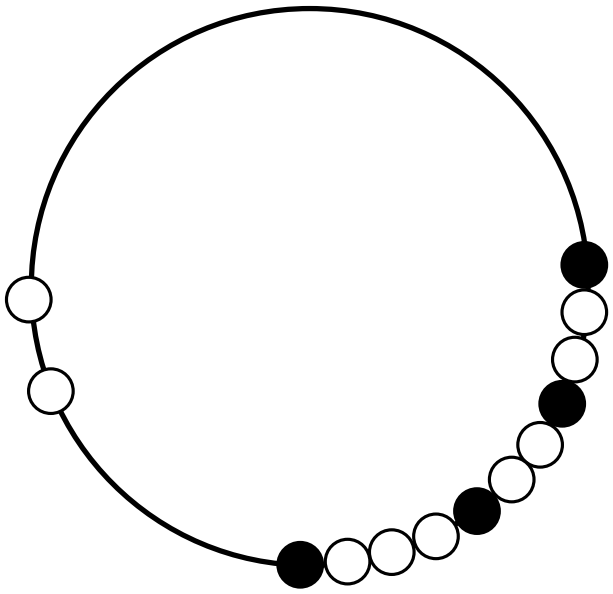

Figure 4.7: **Contribution to current in the single-cluster regime of the long pause limit**. Single huge cluster containing all particles is found for most of the time and no current is produced. When the paused particle at the head returns to active state, a number of active particles released from the cluster move forwards along the chain and contribute to current. As the round-trip time of active particles is much shorter than the pause duration, they rejoin the single cluster again before the another paused particle in head becomes active.

4.5.2 Current in an artificial model with fixed number of paused particles

As a first step toward computing the current for the TASEP in the long-pause limit, we consider an artificial but simpler model where the number n of paused particles is fixed. This will allow us to estimate the contribution to current of the n-pause configurations.

In this artificial model, we have C_0 particles on a one-dimensional lattice with N sites, of which n particles are paused with an average pause duration of τ. Active particles hop from one site to the next one if it is not occupied, which contributes to current as in the TASEP model. However, active particles do not spontaneously enter into the paused state. Each time a paused particle becomes active again, we randomly choose one and only one active particle and turn it into paused state in order to keep the number of paused particles constant.

In the long pause limit, single-cluster behavior is also found, if only because the pause duration is much longer than the round-trip time of active particles. In the same way as for the TASEP model, only active particles released from the single cluster contribute to the current after the paused particle in head has become active. Since the number of paused particles is fixed, the head of cluster becomes active again after an average duration of τ and a certain number of particles can escape from the cluster. The current originates from the movement of these particles. A key point is that these particles cannot make a complete round-trip because of the sites occupied by the cluster. The expression of the current for the model with a fixed number n of paused particles is the following:

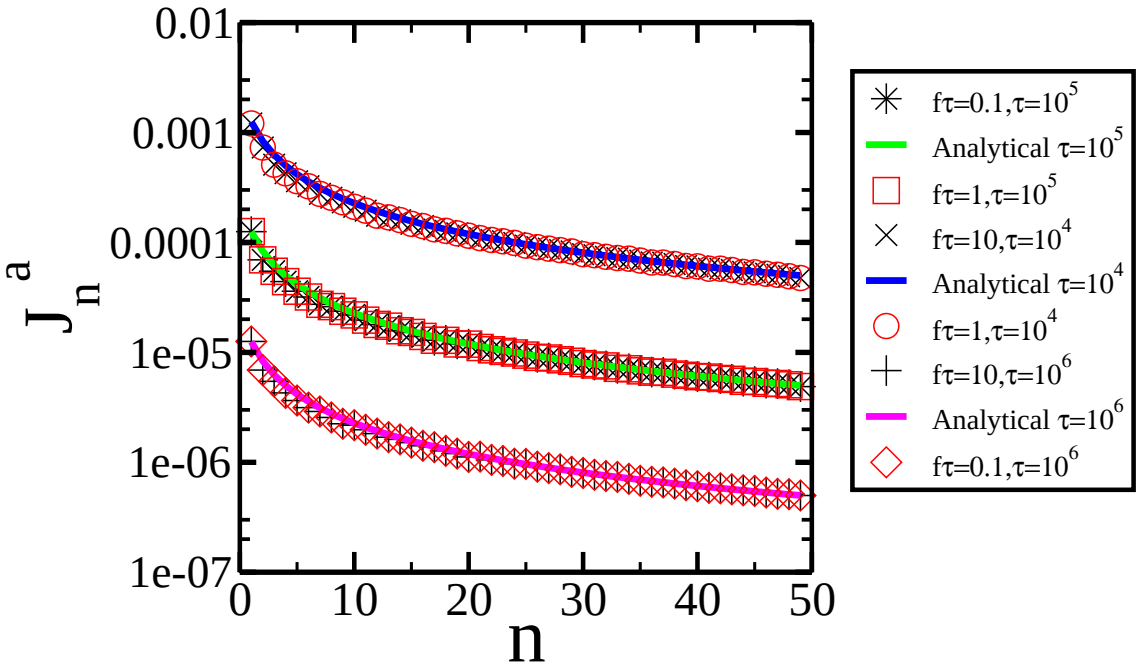

Figure 4.8: **Contribution to current J_n^a as a function of the number of pauses for different parameter sets in the artificial model with a fixed number of paused particles**. Analytical and numerical results are computed in the finite size system with parameters $N = 100$ and $\rho = 0.5$. Solid curves are computed from Eq. (4.15) for different values of pause duration and agree extremely well with the simulations (represented by various symbols), independently of the value of $f\tau$.

$$J_n^a = \frac{\frac{C_0}{n+1}}{\tau} * (1-\rho) = \frac{N\rho(1-\rho)}{(n+1)\tau} \tag{4.15}$$

where $C_0/(n+1)$ is the number of active particles released from the cluster when paused head becomes active. Note that this factor involves $n+1$, not n, because we put systematically another active particle into paused state before the head particle in pause becomes active so that number of paused particle actually is transiently $n+1$ but returns to n immediately thereafter. $1-\rho$ is the fraction of sites not occupied by the cluster. The expression (4.15), which agrees well with numerical simulation for different ranges of parameter product $f\tau$ (Fig 4.8), involves the product $\rho(1-\rho)$ so that the contribution to current is still a symmetric function of particle density. The contribution to current of n-pause configuration in this particular model depends on the size of system N, and it is inversely proportional to the pause duration τ, which is not surprising given that an average time interval of τ must elapse between two releases of active particles from the cluster.

Fig 4.8 shows that the expression (4.15) for J_n^a captures extremely well the contribution to current of the n-pause configuration J_n^a in the artificial model, which is extremely important to find the J_n in our original TASEP model. Even though the number of paused particles is not

fixed in the latter, the typical dynamical behavior and transitions between configurations are quite similar.

4.5.3 Current expression for the finite-size TASEP model

Now that we have established that the typical dynamical behavior in the long pause limit is the single-cluster configuration, and that we have derived the contribution to current of n-pause configuration in an artificial model with fixed number of paused particles, the contribution to current J_n of n-pause configurations in the TASEP model can be obtained.

Since the average pause duration of paused particles at the head of the single cluster is much larger than the round-trip time of active particles, there is no current most of the time. When the head particle becomes active, the number of paused particles is reduced by one. After that transition, a current is produced by active particles released, which travel around the ring until they hit the other side of the cluster. The dynamics just after a transition between two configurations is therefore the key point for the contribution to current, both in the artificial model and TASEP system.

Before the fixed pause-number expression (4.15) can be used in the context of a normal TASEP, a subtle point has to be made. We recall that the current associated to a certain number of pauses can be defined as being proportional to the average number of particles making a round trip in this configuration divided by the time spent in this configuration. This means that whenever a particle is advancing in a numerical simulation, we determine the number n of paused particles in the TASEP at that time, and we associate this contribution to the current with n-pause configuration.

In the artificial model with a fixed number of paused particles, the waiting time τ and the number $\frac{C_0}{n+1}$ of released particles are both counted for the n-pause configuration. However, in the TASEP model where the number of pauses varies in time, we need to be very careful when assigning a waiting time and a contribution to current to the configuration of a given-number paused particles. In particular, assume that just before the paused particle at the head of cluster becomes active, there are $n+1$ paused particles. When the head become active, the $(n+1)$-pause configuration becomes the n-pause configuration. Then $\frac{C_0}{n+1}$ active particles are released from the cluster and contribute to the current. Therefore, the waiting time τ should be included in statistics for $(n+1)$-pause configuration rather than for n-pause one, while the advance of the $\frac{C_0}{n+1}$ particles released contributes to the current for n-pause configuration. To account for this fact, a simple prescription is to renormalize the contribution to current J_n by a factor T_{n+1}/T_n, where T_n is the time spent so far in n-pause configurations, so that the number of advances occurring at times when there are exactly n pauses is divided by the correct amount of time. Fortunately, we know this ratio, because T_n is by definition proportional to the probability P_n of having n paused particles. Therefore, the expression of contribution to current J_n reads:

$$\begin{aligned} J_n &= \frac{N\rho(1-\rho)}{(n+1)\tau}\frac{T_{n+1}}{T_n} \\ &= \frac{N\rho(1-\rho)}{(n+1)\tau}\frac{P_{n+1}}{P_n} \\ &= \frac{N\rho(1-\rho)}{(n+1)} f \frac{C_0-n}{n+1} \end{aligned} \tag{4.16}$$

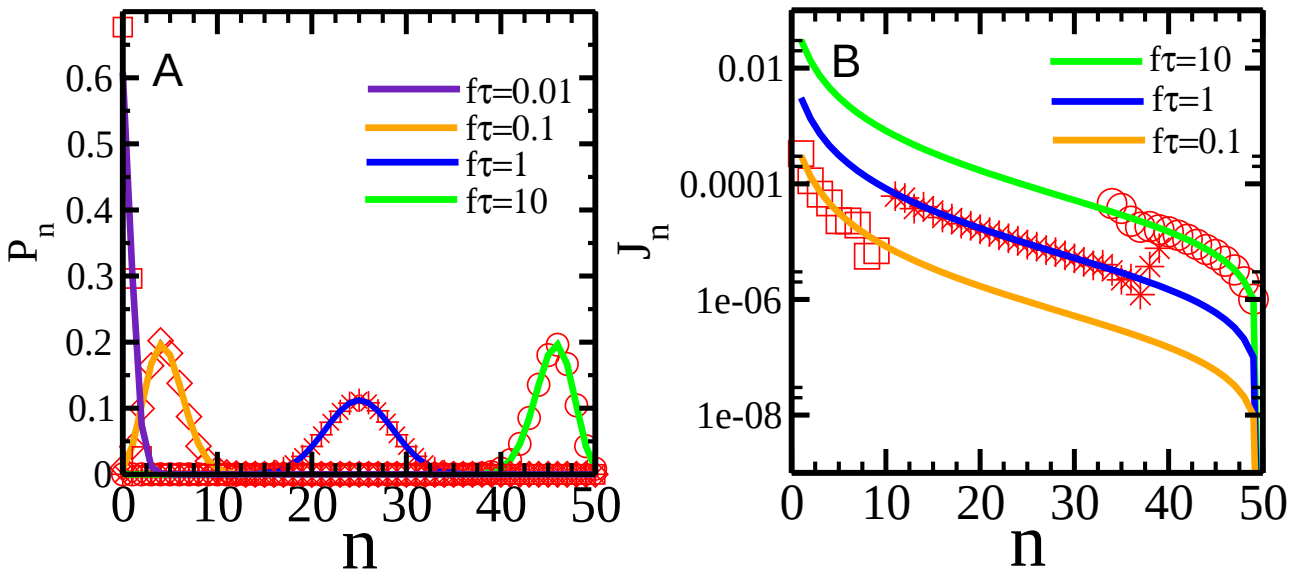

Figure 4.9: **Probability P_n and contribution to current J_n of n-pause configuration**. Curves for different values of $f\tau$ are obtained according to Eq. (4.6) and Eq. (4.16). They agree well with results from numerical simulations, shown by symbols of different shapes and colors.

Note that this expression contains the term $\rho(1-\rho)$ and also N and C_0. It is easy to see that the contribution to current is reduced if we increase the number of paused particles while keeping their total number constant, as expected. As shown in Fig 4.9, the expression (4.16) predicts perfectly the numerical estimations of the J_n for different ranges of parameters.

We have discussed the fact that the contribution to current (4.16) is generated by the transition from a $(n+1)$-pause configuration to a n-pause one. The general expression of J_n predicts a contribution of zero-pause configuration, which we will denote J_0^*, corresponding to the transition from a 1-pause configuration to a 0-pause configuration. The question then arises whether the contribution J_0^* should be taken into account in addition to the normal term $J_0 = \rho(1-\rho)$. We tested which of the two expressions $J_0 = \rho(1-\rho)$ or $J_0 + J_0^* = \rho(1-\rho)(1+fC_0N)$ agreed better with numerical simulations, and found that the first one was clearly more accurate. In the following we will therefore only consider the contribution of the J_n given by Eq. (4.16) for $n > 1$.

Given the contribution to current J_n of the n-pause configuration, and the probability P_n of having n-pause configuration, we can construct the expression of total current according to relation (4.4):

$$J_T = \frac{\rho(1-\rho)}{(1+f\tau)^{C_0}} + \sum_{n=1}^{C_0} \frac{C_{C_0}^n (f\tau)^n}{(1+f\tau)^{C_0}} \frac{N\rho(1-\rho)}{(n+1)} f \frac{C_0 - n}{n+1} \tag{4.17}$$

where the $C_{C_0}^n$ are binomial coefficients. The expression (4.17) of the current for the finite-size system in the long pause limit is relatively complex, but predicts extremely well the total current (Fig 4.10). Note that for very long pause durations, the total current tends to a finite constant $\frac{\rho(1-\rho)}{(1+f\tau)^{C_0}}$ which is not negligible when the size of the TASEP remains small. This is due to the fact that the probability of having no particles in pause is not negligible.

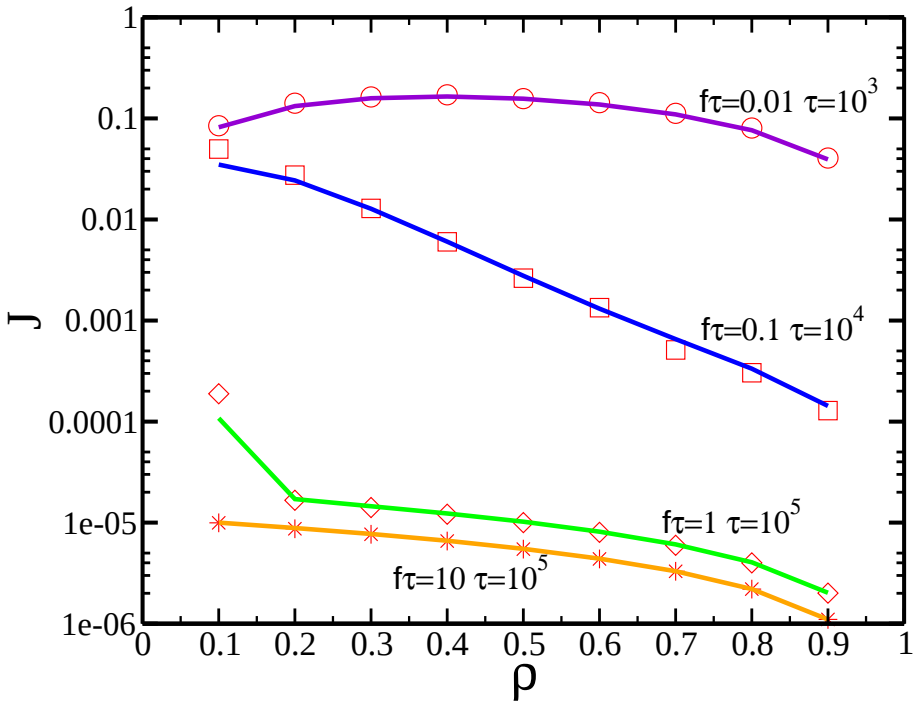

Figure 4.10: **Current as a function of RNAP density in the long pause regime**. Values of the total current computed by numerical simulations for different values of $f\tau$ with $N = 100$ in the long pause regime (circles, squares, diamonds and stars). They agree very well with the predictions of formula (4.17) which are represented by solid curves.

It is interesting to consider more closely in Fig 4.10 how the $J_T(\rho)$ curve evolves as a function of the product $f\tau$, which determines the fraction of paused particles. When there is a small fraction of paused particles, for example when $f\tau = 0.01$, the dependence of total current on density remains symmetric with a maximum for the half-occupied ring, as in the short-pause regime. In contrast, when the fraction of paused particles becomes larger, the curve becomes more and more asymmetric, and eventually decreases monotonously when the particle density increases. In the latter case, the highest transcription rate is observed at small density. In addition, the total current is globally reduced when $f\tau$ is increased. Thus RNAP pausing in the long pause limit qualitatively changes the dynamical behavior compared with the short pause case.

We can now make a summary of the behavior of the finite-size TASEP in the short-pause and long pause limits. Transcription rate (or current) displays three different regimes according to pause duration (Fig 4.11). In the short pause limit, particles are uncorrelated, thus the current does not depend on pause duration and is constant. In the long pause limit, particles most probably remain in one huge cluster and are highly correlated. The current is also approaching to a constant value when pause duration is infinitely large. Our expression of total current in both limits agrees well with numerical simulations. Even near the end of the intermediate pause regime, where the single-cluster assumption fails, is quite well approached by our expressions. Note that both expressions overestimate the current when applied in the intermediate pause case.

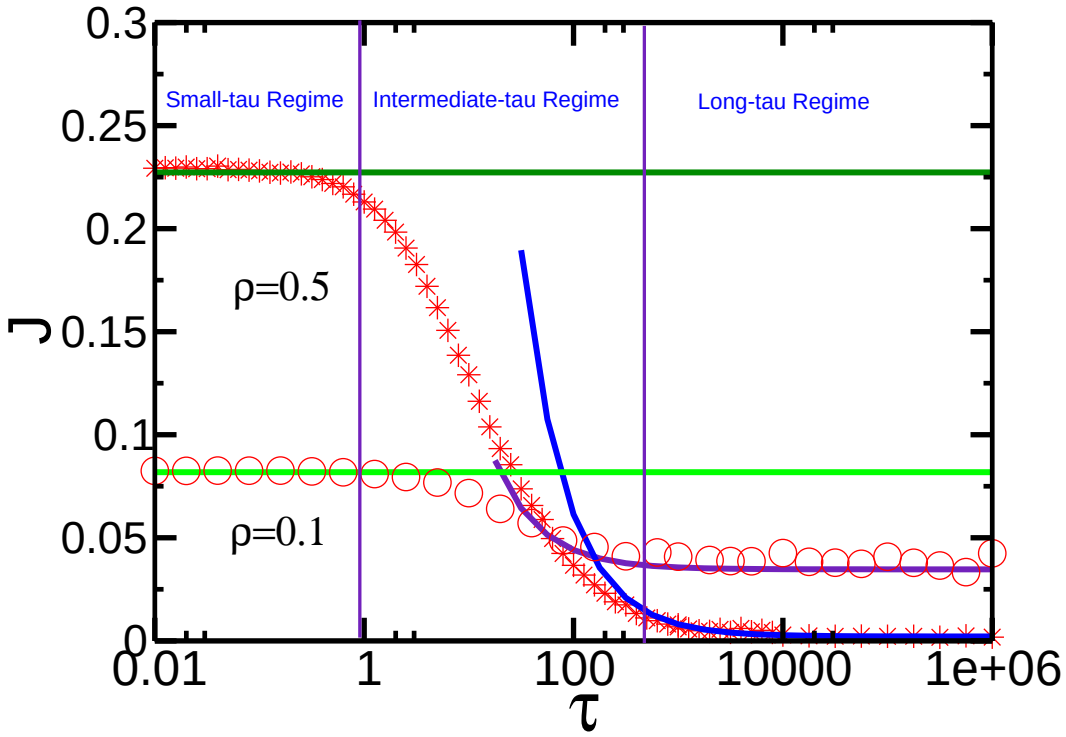

Figure 4.11: **Transcription rate of the finite-size TASEP system in the short pause and long pause regimes**. The mean-field value of current (4.12) valid for small values of the pause duration τ, shown by a dark green curve for $\rho = 0.5$ and by a green curve for $\rho = 0.1$, is obtained under mean-field approximation. The value of the expression (4.17) of the current in the large-τ regime, obtained in the single-cluster approximation, is plotted as a blue curve for $\rho = 0.5$ and an indigo curve for $\rho = 0.1$. Both of them agree with numerical simulations in their respective domains of validity.

4.5.4 Current expression for the infinite-size TASEP model

In previous sections, we found that there is no difference in the short pause limit between finite-size and infinite-size systems. Here we will establish the expression of total current for the infinite-size TASEP model in the long pause limit, and compare it with that in the finite-size TASEP.

We begin by evaluating the contribution to current $J(\tilde{n})$ in the infinite-size limit, where $\tilde{n}$ is, defined as before, the fraction of paused particles. Obviously, it is impossible in the infinite-size limit to have a single cluster. Rather, there are infinitely many clusters, which should however have approximately the same size distribution as in the case of a finite size. As before, the current only occurs when paused particles at the head of a cluster become active. We therefore expect that the result obtained in Eq. (4.16) also holds here and thus that configurations with a fraction $\tilde{n}$ of active particles contribute a current:

$$J(\tilde{n}) = \frac{f(1-\rho)}{\tilde{n}}(\frac{1}{\tilde{n}} - 1) \tag{4.18}$$

This expression is much simpler than that in the finite-size TASEP model. The total current is given by the contribution to current of the most probable fraction of particle in pause $\langle\tilde{n}\rangle =$

$f\tau/(1+f\tau)$:

$$\begin{aligned} J_T &= J(\langle\tilde{n}\rangle) \\ &= J(\frac{f\tau}{1+f\tau}) \\ &= \rho(1-\rho)\frac{1+f\tau}{f\tau\,\rho\tau} \end{aligned} \tag{4.19}$$

where we purposely keep the term $\rho(1-\rho)$ as a factor to conveniently compare our result with expressions of current found in other limits. The total current is inversely proportional to the pause duration τ, so it always decreases when pause duration is increased and it goes to zero

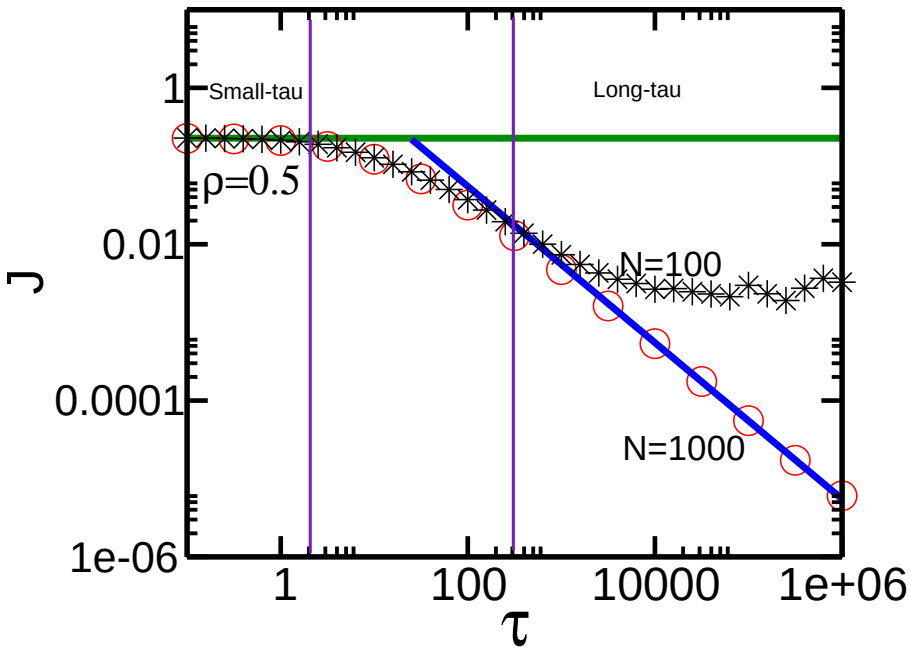

Figure 4.12: **Comparison of currents for the finite-size and the infinite-size TASEPs in the short, intermediate and long pause regimes**. Numerical estimations of current are carried out with parameters $\rho = 0.5$. $N = 100$ (black stars) considered to be representative of a finite-size system, and $N = 1000$ (red cycles) which is already approaching the infinite-size limit. It can be seen that short pause and intermediate pause regimes are not influenced by system sizes. In contrast, finite-size and infinite-size TASEPs display dramatical differences in the long pause limit. Current tends to a constant for finite size, but decreases indefinitely in the infinite-size case and is asymptotically proportional to $1/\tau$. Expressions obtained for the short pause (4.14) and long pause limits (4.19) predict well the variation of current.

The variations of the total current with pause duration in the infinite-size and finite-size TASEPs are shown in Fig 4.12. Note that the short pause and intermediate pause regimes of these two cases are perfectly superimposed, so that these two regimes are independent of system size. The only difference occurs in the long pause regime. In a finite-size system, the

total current is approaching to an asymptotic constant, in contrast, for infinite-size limit, it is just a continuation of intermediate pause case and there is no clear boundary between them. The current decreases with pause duration to zero. Most importantly, the expression of current given by Eq. (4.19) predicts perfectly this long pause regime for infinite-size limit.

In this section, we have discussed the dynamical influence of transcriptional pausing in the long pause regime in both finite-size and infinite-size systems. In the finite-size system, single-cluster behavior was found. This allowed us to deduce the contribution to current of each configuration, as well as the total transcription rate (current). The expression found for both finite-size and infinite-size systems predicts correctly the transcription rate. Next, we will discuss how the transcription rate depends on the pause duration in the intermediate pause case, for which the ranges of biological parameters are found.

4.6 Phenomenological description and transcription rate for the intermediate pause case

In previous sections, transcription interrupted by RNAP pausing was modeled by TASEP systems with periodic boundary conditions. It is found that the transcription rate displays three dynamical regimes, which we termed the short, intermediate and long pause regimes. We have studied dynamical behaviors of the short and long pause regimes for which expressions of the transcription rate were found.

Two processes control the dynamical behavior and affect transcription rate. One is the movement of active particles which is determined by elongation rate and particle density, and the other is the blocking of particles by other particles in pause that leads to clusters defined as an ensemble of particles with a paused particle in head, as discussed before.

In the short pause limit, the time scale of particle movement is larger than that of pausing so that a particle is only affected by the active or paused nature of its own states, not sensing that of its neighbors. Particles are not correlated. Occupancy of each site is independent of that of others. The transcription rate is well predicted by the mean-field approximation. In contrast, in the long pause limit, clusters dominate system behavior. Particles spend most fraction of time in clusters, otherwise, they become active moving forwards between clusters. In particular, in finite-size system, there is only one cluster on the TASEP ring.

In the intermediate pause case, the typical dynamical behavior is the coexistence of active particle movement and of clusters. A particle may move forwards without being affected by pausing, or it may be confined in a cluster. Sizes of clusters are smaller than those in long pause regimes, however, there are much more of them. Moreover, these small-size clusters are quickly created once a particle switches to the paused state, and can easily collapse when paused particles return to the active state. The intriguing dynamics of clusters can highly reduce the transcription rate.

An approach which seems promising is that of studying the statistics of residence time. During its motion around the TASEP ring, a particle will spend different amounts of time on different sites. A question is then how these amounts of time are distributed. We may further divide the residence time into different contributions depending on whether the particle was active or paused. There is a direct connection between residence times and the current which is in fact the inverse of the average residence time.

The simplest and most naive way to combine the formulas we have obtained for the short

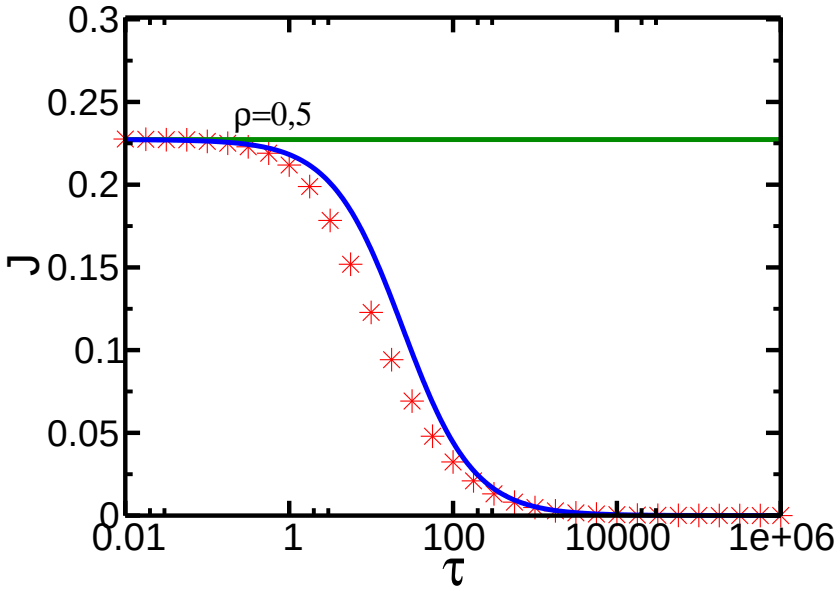

Figure 4.13: **Transcription rate (current) for three regimes**. Numerical simulation is carried out with parameters: $N = 1000$, $f\tau = 0.1$, $\rho = 0.5$. The theoretical expression (4.21) shown by blue curve predicts the current in all ranges of pause duration τ, including the intermediate pause regime with certain deviation if compared to numerical simulations.

pause and long pause regimes into an expression working in the entire range of pause durations is to assume that the average residence time t_r is the sum of the mean-field residence time t_r^{MF} and of the mean residence time t_r^{LP} we have obtained in the long pause regime:

$$t_r = t_r^{MF} + t_r^{LP} = t_r^{MF} + \mathcal{P}\tau, \quad t_r^{MF} = \frac{1+f\tau}{(1-\rho)}, \quad \mathcal{P} = \frac{\rho f\tau}{(1+f\tau)(1-\rho)} \tag{4.20}$$

where it has been expressed that the long-pause residence time is proportional to pause duration τ so that the coefficient $\mathcal{P}$ can be viewed as the average number of pauses encountered when moving over one site (in other words, $N\mathcal{P}$ is the average number of pauses encountered over one turn). We have also assumed that $\epsilon = 1$ (i.e., time is renormalized to the mean elongation time). Figure 4.13 compares the current predicted by formula (4.21) with the result of numerical simulations.

It can be seen in Fig 4.13 that expression (4.21) interpolates relatively well between the short pause and the long pause limits but that there is a notable difference with the output of numerical simulations in the intermediate pause regime. In this regime, indeed, the dynamics is quite complex. Permutations of pausing between particles are rapid but each pausing is long enough to create clusters of small or intermediate size blocking active particles. These clusters of all sizes form and break down continuously.

This line of research is still being developed in our group and I will not specify its details here, except that some interesting results have been obtained. For example, we have found that in the intermediate regime, particles are blocked by significantly more pauses than expected during one round trip around the TASEP. An explanation is that during the time in which a particle is blocked by other particles queued behind a paused particle, waiting for the latter to become active, one of the particles in front of it may itself enter into paused state, thereby masking the original paused particle. As the particle waits for this new pause to finish, a new

pause may appear, and so on. It seems that an effective expression for the mean residence time of the form

$$t_r = t_r^{MF} + \gamma(\tau)\mathcal{P}\tau, \tag{4.21}$$

agrees quite well with numerical simulations if $\gamma(\tau)$ is a function varying typically between 2 for small τ and 1 for large τ (so that the long pause limit expression is recovered), with a maximum around 2.6. It remains how to explain with a simple argument the origin of the correcting factor $\gamma(\tau)$.

4.7 Conclusion

In this chapter, we have studied the dynamical effects of RNA pausing on transcription using a classical model of out-equilibrium statistical physics, the so-called TASEP model with periodic boundary condition. We found that depending on pause duration, the transcription rate dynamics displays three regimes. The first is the short pause regime where particles are uncorrelated, so that the mean-field approximation works well and TASEP model can be mapped to the model without pausing. The intermediate pause regime is the second one. Pause duration is larger than the advance time of active particles, so that pausing can block active particles, thereby forming clusters, which reduces significantly the transcription rate. The third regime is the case where pause duration is infinitely large. Clusters dominate the dynamical behavior. We understood the characteristic dynamical behavior of all three regimes and found the expression of the transcription rate in finite-size and infinite-size systems for short pause and long pause limits which agree well with numerical simulations. We also discussed briefly the strategy to obtain an expression of the transcription rate in the intermediate regime and directions for future research. Our results emphasize the significant influences of RNAPs pausing on the transcription dynamics. If the experimental technologies in synthetic biology allow to tune the RNAPs pause duration, it will be very helpful to confirm our conclusions. In addition, comparing the different behaviors of synthetic genes with RNAPs pausing and of that without pausing will provide insights into the dynamical role of RNAP pausing in the transcriptional bursting.

General conclusion and perspectives

Recently the interaction of biology with physics has favored significant advances. Physics not only brings new experimental technologies to measure and manipulate molecular and cellular events with high resolution, but also theoretical modeling, which provides powerful predictions and can be further tested by quantitative experiments. In particular, modeling enables us to reveal the design principles of various biological phenomena and to quantitatively understand cellular processes. The focus of this thesis has been to theoretically elucidate the influence of some key ingredients on the dynamics of genetic networks, such as the time delay and fluctuation in cellular processes, as well as RNA polymerase (RNAP) pausing.

This thesis is divided into three studies. In the first part, we analyzed the combination of various and multiple time delays and how it affects oscillatory behavior in a minimal genetic network comprising a self-repressing gene. The second part developed a cumulant expansion of the master equation which revealed the interplay between fluctuations and nonlinearities of genetic networks, and it showed how fluctuations can induce biological oscillations. In the third part, we investigated the dynamical effect of RNAP pausing on transcription in the context of a classical system of out-of-equilibrium statistical physics: the Total Asymmetric Simple Exclusion Process (often referred to as TASEP). Here we will make a brief conclusion and perspective for each of three studies.

Oscillation arising from combination of various and multiple time delays with non-linearities in a self-repressing gene

The ubiquity and importance of oscillations have been highlighted in various biological systems, such as circadian rhythms [53, 75, 160, 40, 3, 32], cell cycles [129], immune response [66], cell growth/death [93] and embryo development [2, 60]. Intensive studies have focused on the design principles of biological oscillations [100, 127, 135, 145]. Several key ingredients of oscillations have been revealed, such as negative feedback (a necessary condition), delay and nonlinear degradation. A typical feature of biological systems is that various and multiple time delays are derived from different sources, in particular are originated in transcription, translation, molecule transport, phosphorylation, etc. Moreover, nonlinear degradation is also expected to be an important source of delay [135]. The combination of these various and multiple delays may be the key to explain the experimentally observed robustness of oscillations [135, 175]. However, the combination principle of delays had not yet quantitatively studied. There are two strategies of modelling delays commonly used in literature: explicit delay and so-called reaction delay, whose differences remained unclear.

Using a simple self-repressing gene circuit involving two delays due to dynamics of gene activity and to protein transport, analytical criteria for oscillations were found. They not only

quantitatively unify key ingredients previously discussed, but also demonstrate the trade-off between delay and nonlinear degradation. There is a time scale of delay at which the system is most destabilized. As to the combination principle of various delays, explicit delays are simply combined depending on their sum, whereas reaction delays interact in a non-trivial way so that how reaction delays are distributed among different biological steps also influences oscillations. For small delays or saturated degradation, explicit and reaction delays display the same dynamical influence. In the general case, however, their influences are dramatically different. An explicit delay always promotes oscillation by destabilizing the steady state of the system. In contrast, a reaction delay can destabilize or even stabilize the system.

We hope that our analytical comparisons between explicit and reaction delays will guide people to choose how to model delays in theoretical studies and also experimental data analysis. In addition, our conclusion that the non-trivial combination of various delays can launch more easily oscillations than a single delay equal to their sum may explain the robustness of oscillations and need to be further tested experimentally.

Influence of stochastic fluctuations on the oscillation of a self-repressing gene

Experimental and theoretical investigations [38, 105, 179, 183, 33, 151] reveal that genetic networks are subject to large fluctuations (or noise), as most biochemical reactions in cell involve low copy numbers of reacting molecules. It was theoretically demonstrated that fluctuation is unavoidable in biological systems [125]. Fluctuation is not always a nuisance, but can be a useful mechanism responsible for phenotypic heterogeneity [157, 120], coordinated expression of a large set of genes and probabilistic differentiation strategies [195, 35, 5, 77]. However, the influence of fluctuation on the dynamics of genetic networks is not fully understood, and theoretical approaches that describe molecular fluctuations and are easily solved are needed.

To address these questions in a simple setting, we have focused on the minimal genetic network consisting of a self-repressing gene. In this circuit, sources of fluctuations are not only the low copy numbers of mRNA and protein molecules, but also the dynamics of gene activity which switches between the "ON" and "OFF" states. In order to study the dynamical influence of fluctuations, we performed a cumulant expansion of the master equation which allows us to appropriately describe the binary gene states. Using this cumulant expansion, we reveal that the interplay between nonlinearities of the genetic circuit and stochastic fluctuations dramatically shifts the steady state predicted by deterministic rate equations and induces oscillations without requiring nonlinear degradation. The oscillation zone predicted by cumulant expansion coincides with parameter region with most regular spiking behavior characterized by the Fano factor.

This study suggested an interesting approach to take into account fluctuations in genetic networks. This theoretical approach is easily solved owing to its deterministic form that comprises the average quantities and other-order cumulants representing fluctuations. It can also applied for other biological systems that are subject to fluctuations.

Dynamical effects of RNAP pausing on transcription

Transcription is one of the most complex and tightly regulated processes in gene expression [106, 192, 76, 18, 178]. During the past decade, the intriguing dynamics of transcription in which the macromolecular machine RNA Polymerase (RNAP) plays a critical role caused much interest. Moreover, RNAPs display some unusual behaviors during the process of producing mature RNA molecules, such as stochastic pausing which affects RNAP traffic and probably contributes to transcriptional bursting. According to single-molecule experiments, there are two main types of pauses, elemental short pauses [56, 107, 119, 55, 57] and prolonged backtracking pauses [55, 57, 4]. The majority of pauses are elemental short pauses with weak sequence dependence and are not affected by trailing RNAPs.

In order to investigate the dynamical influence of elemental pauses on transcription, we have considered the classical statistical TASEP system with periodic boundary conditions to model the transcription incorporating pausing. In the limit case where pause duration is short, we can construct a mean-field model to analyze the transcription rate and site occupation. In the general case where the mean-field approach no longer applies, we have used a statistical approach to study this model. We have obtained a good understanding of the various mechanisms driving the transcription dynamics over the entire range of pause duration, and in particular a theoretical expression of transcription rate agreeing well with numerical simulations.

This study shows the critical role of RNAP pausing in transcription dynamics (which possibly underlies the transcriptional bursting) and provides theoretical analysis of how pausing affects transcription rate in the parameter range where mean-field approach fails. In this region, the dynamics is governed by the formation of clusters, where a number of particles remain blocked behind a paused particle for some time and then are released. Future studies should be focused on the TASEP model with open boundary conditions, which resembles more closely the transcription process, as well as on the dynamical influence of backtracking, which is another specific behavior of RNAPs.

Appendix A

Appendix A: Linear stability analysis of the ODE and DDE systems

A.1 Normalization and analytical criterion of the basic ODE model

In the model proposed by Morant *et al.* [138], The kinetic of the self-repressing gene is described by the following equations:

$$\frac{dG}{dT} = \theta_0(1-G) - \alpha_0 PG \tag{A.1a}$$

$$\frac{dP}{dT} = \frac{dG}{dt} + \beta_0 M - \delta_P F(P) \tag{A.1b}$$

$$\frac{dM}{dT} = \mu_0 + \lambda_0 G - \delta_M H(M) \tag{A.1c}$$

where G, P and M represent respectively gene activity, protein and RNA copy numbers. Then the time, variables and parameters of Eqs. (A.1) are normalized as follows:

$$\begin{aligned} &T = \frac{t}{\delta_m}, \;\; G = g, \;\; P = pP_0, \;\; M = mM_0 \\ &P_0 = \frac{\theta_0}{\alpha_0}, \;\; M_0 = \frac{\delta_P P_0}{\beta_0}, \;\; \theta = \frac{\theta_0}{\delta_M}, \\ &\alpha = \frac{\theta_0}{P_0\delta_M}, \;\; \delta = \frac{\delta_P}{\delta_M}, \;\; \lambda = \frac{\lambda_0}{M_0\delta_M} \\ &\mu = \frac{\mu_0}{M_0\delta_M}, \;\; f(p) = \frac{F(P)}{P_0}, \;\; h(m) = \frac{H(M)}{M_0} \end{aligned} \tag{A.2}$$

Thus Eqs. (A.1) can be rewritten in dimensionless form:

$$\frac{dg}{dt} = \theta\,(1 - g(1+p)) \tag{A.3a}$$

$$\frac{dp}{dt} = \alpha\,(1 - g(1+p)) + \delta(m - f(p)) \tag{A.3b}$$

$$\frac{dm}{dt} = \mu + \lambda\, g - h(m) \tag{A.3c}$$

Eqs. (A.3) have a single steady state (g_*, p_*, m_*). The behavior of the degradation of proteins and mRNAs in the neighborhood of the steady state is described by the slopes:

$$s = \frac{df(p)}{dp}\Big|_{p=p_*}, u = \frac{dh(m)}{dm}\Big|_{m=m_*} \tag{A.4}$$

For the sake of simplicity, here we assume perfect repression when the gene is bound by a protein ($\mu = 0$) and a large threshold ($P_0 \gg 1$) leading to $\alpha \sim 0$. Under this approximation, the Routh-Hurwitz stability criterion [89] indicating that the Hopf bifurcation occurs is specified as:

$$\mathcal{H} = (\delta\, s + u)(\delta\, s\tau_g + 1)(u\tau_g + 1) - g_*^2 \delta\lambda\tau_g < 0 \tag{A.5}$$

where $\tau_g = g_*/\theta$ denotes the time needed for gene response to a sudden variation of proteins. $\delta\, s$ and u characterize respectively the degradation rate of proteins and mRNAs. In order to simplify this criterion, both degradation rates are replaced by their sum $\sigma = \delta\, s + u$ and product $\gamma = \delta\, su$. Therefore, the criterion (A.5) becomes:

$$\mathcal{H} = \sigma + (\sigma^2 - g_*^2 \delta\, \lambda)\, \tau + \sigma\, \gamma\, \tau^2 < 0 \tag{A.6}$$

The criterion (A.6) is renormalized as follows:

$$\sigma_c = g_* \sqrt{\delta\, \lambda},\ \sigma = \sigma_c\, \Sigma,\ \gamma = \frac{\epsilon^2\, \Sigma^2\, \sigma_c^2}{4},\ \tau = \frac{T}{\sigma_c} \tag{A.7}$$

So the following expression of the Routh-Hurwitz criterion is obtained:

$$\mathcal{H}_\epsilon(\Sigma,\, T) = \Sigma \left(\frac{\epsilon^2\, \Sigma^2}{4} T^2 + (\Sigma - \frac{1}{\Sigma})T + 1 \right) < 0 \tag{A.8}$$

A.2 Linearization of the DDE system

The model with explicit delays describing the simple negative feedback loop of a self-repressing gene is expressed as follows:

$$\frac{dg}{dt} = \theta\ (1 - g(1 + p(t - \tau_1))) \tag{A.9a}$$

$$\frac{dm}{dt} = \lambda\, g - h(m) \tag{A.9b}$$

$$\frac{dp}{dt} = \delta\, (m(t - \tau_2) - f(p)) \tag{A.9c}$$

where τ_1 and τ_2 are two explicit delays due to protein and mRNA transports, respectively. We first search for the steady state (g_*, m_*, p_*), which is the same as for the model (2.2) without transport delays. We also find that the steady state does not depend on the parameters τ_1 and τ_2. We use the linearization in the neighborhood of steady state, so variables are rewritten as: $(g = g_* + \Delta\, g, p = p_* + \Delta\, p, m = m_* + \Delta\, m)$ where $\Delta\, g$, $\Delta\, p$ and $\Delta\, m$ are respectively small perturbations of variables around the their steady state. In keeping the first order of Taylor expansion around the steady state, these perturbations are expressed as:

$$\Delta\, g(t) = \Delta g_0\, e^{\xi\, t},\ \Delta\, p(t) = \Delta p_0\, e^{\xi\, t},\ \Delta\, m(t) = \Delta m_0\, e^{\xi\, t}, \tag{A.10a}$$

$$\Delta\, g(t - \tau_2) = \Delta g_0\, e^{\xi\, (t - \tau_2)},\ \Delta\, m(t - \tau_1) = \Delta m_0\, e^{\xi\, (t - \tau_1)} \tag{A.10b}$$

Then we inject these variables in the neighborhood of steady state, we obtain the Jacobian matrix:

$$\vec{0} = \begin{pmatrix} -\theta/g - \xi & -\theta\, g & 0 \\ 0 & -\delta\, s - \xi & \delta\, e^{-\xi\,\tau_1} \\ \lambda\, e^{-\xi\,\tau_2} & 0 & -u - \xi \end{pmatrix} \times \begin{pmatrix} \Delta\, g_0 \\ \Delta\, p_0 \\ \Delta\, m_0 \end{pmatrix} \tag{A.11}$$

where ξ is the eigenvalue. Therefore, the characteristic equation of eigenvalue is:

$$\xi^3 + \frac{1 + g_*\tau_g(\delta\, s + u)}{g_*\tau_g}\xi^2 + \frac{\delta\, s + u + g_*\tau_g\delta\, s\, u}{g_*\tau_g}\xi + \frac{\delta\, s\, u + \delta\, \lambda\, g_*^2\, e^{-(\tau_1+\tau_2)\xi}}{g_*\tau_g} = 0 \tag{A.12}$$

In this characteristic equation, only the sum of explicit delays appears, so we suppose $\tau_e = \tau_1 + \tau_2$.

A.3 Stability analysis of DDE system

A.3.1 Stability analysis for a general case

In order to show the approach for the stability analysis of DDEs, we start with a general system whose characteristic equation is written as:

$$P_1(\xi) + P_2(\xi)e^{-\xi\,\tau_e} = 0 \tag{A.13}$$

where ξ is the eigenvalue and τ_e is an explicit delay. $P_1(\xi)$ and $P_2(\xi)$ are arbitrary functions of eigenvalues. For the purpose of understanding stability of steady state, it is crucial to determine the value of parameter τ_e at which the characteristic equation has a pair of conjugate pure imaginary roots. Therefore, we suppose that $\xi = i\omega$ and Eq. (A.13) becomes:

$$P_1(i\omega) + P_2(i\omega)e^{-i\omega\tau} = 0 \tag{A.14}$$

We express functions $P_1(\xi)$ and $P_2(\xi)$ in the real and imaginary parts:

$$R_1(\omega) + iQ_1(\omega) + (R_2(\omega) + iQ_2(\omega)(\cos(\omega\tau) - i\sin(\omega\tau)) = 0 \tag{A.15}$$

where $R_1(\omega)$ and $Q_1(\omega)$ are respectively the real and imaginary parts of $P_1(\xi)$, and $(R_2(\omega)$ and $Q_2(\omega)$ correspond to real and imaginary parts of $P_2(\xi)$. In separating real and imaginary parts of Eq. (A.15), we obtain two equations:

$$\begin{aligned} R_1(\omega) + R_2(\omega)\cos(\omega\tau) + Q_2(\omega)\sin(\omega\tau) = 0 \\ Q_1(\omega) - R_2(\omega)\sin(\omega\tau) + Q_2(\omega)\cos(\omega\tau) = 0 \end{aligned} \tag{A.16}$$

By squaring both sides of Eqs. (A.16) and adding them, we eliminate the $\cos(\omega\tau)$ and $\sin(\omega\tau)$ and get an equation for the frequency of eigenvalue corresponding to the frequency of periodic solution.

$$F(\omega) = R_1(\omega)^2 + Q_1(\omega)^2 - R_2(\omega)^2 - Q_2(\omega)^2 = 0 \tag{A.17}$$

We have focused on the critical value of parameters at which the eigenvalues become a pair of conjugate pure imaginary roots. Another important question is in which direction eigenvalues cross the imaginary axis. We determinate the direction by the following expression:

$$S = sign[\frac{d(Re\,\xi(\tau))}{d\tau}\Big|_{\xi=i\omega}] \tag{A.18}$$

We can deduce it as:

$$S = sign(\frac{dF(\omega)}{d\omega}) \tag{A.19}$$

So the sign of derivate of the equation for eigenvalue frequency crossing the imaginary axis measures the transition of system. If S is positive, eigenvalues cross the imaginary axis from left to right, which means the steady state becomes unstable. Likewise, the steady state become stable.

A.3.2 Stability analysis for our model with molecular transport delay

In applying the approach above into our model, we obtain:

$$R_1(\omega) = -\frac{(1 + g_*\tau_e(\delta s + u))\omega^2}{g_*\tau_e} + \frac{\delta s u}{g_*\tau_e} \tag{A.20a}$$

$$Q_1(\omega) = -\omega^3 + \frac{(\delta s + u + g_*\tau_e\delta s u)\omega}{g_*\tau_e} \tag{A.20b}$$

$$R_2(\omega) = \frac{\delta g_* \lambda}{g_*\tau_e} \tag{A.20c}$$

$$Q_2(\omega) = 0 \tag{A.20d}$$

The criterion for oscillations is expressed as follows:

$$\cos(\omega\,\tau_e) = \frac{[1 + g_*\tau_g(\delta\, s + u)]\omega^2 - \delta\, s\, u}{g_*^2\delta\,\lambda} \tag{A.21a}$$

$$\sin(\omega\,\tau_e) = \frac{-g_*\tau_g\,\omega^3 + (\delta\, s + u + g_*\tau_g\,\delta\, s\, u)\omega}{g_*^2\delta\,\lambda} \tag{A.21b}$$

The equation for the frequency of eigenvalue crossing the imaginary axis is:

$$F(\Psi) = \Psi^3 + [\frac{1}{T^2} + (1 - \frac{\epsilon^2}{2})\,\Sigma^2]\Psi^2 + [\frac{\epsilon^4\,\Sigma^4}{16} + (1 - \frac{\epsilon^2}{2})\,\frac{\Sigma^2}{T^2}]\,\Psi + (\frac{\epsilon^4\,\Sigma^4}{16T^2} - \frac{1}{T^2}) = 0$$

where $\Psi = \omega^2$.

Appendix B

Appendix B: Derivation of the cumulant expansion of the master equation

B.1 Master equation

In this section, we will derive the cumulant expression of the master equation for the simple genetic circuit of a self-repressing gene. In order to decrease the nonlinearities of genetic network, we suppose both linear degradations for mRNA and proteins.

$$\begin{aligned}
\frac{dP_{0,m,p}}{dt} &= \alpha(p+1)P_{1,m,p+1} - \theta\, P_{0,m,p} \\
&\quad +\delta_m(m+1)P_{0,m+1,p} - \delta_m m P_{0,m,p} \\
&\quad \beta\, m P_{0,m,p-1} - \beta\, m P_{0,m,p} \\
&\quad +\delta_p(p+1)P_{0,m,p+1} - \delta_p\, p P_{0,m,p} \\
\frac{dP_{1,m,p}}{dt} &= \theta\, P_{0,m,p-1} - \alpha\, p P_{1,m,p} \\
&\quad \lambda\, P_{1,m-1,p} - \lambda\, P_{1,m,p} \\
&\quad \delta_m(m+1)P_{1,m+1,p} - \delta_m m P_{1,m,p} \\
&\quad \beta\, m P_{1,m,p-1} - \beta\, m P_{1,m,p} \\
&\quad \delta_p\,(p+1)P_{1,m,p+1} - \delta_p\, p P_{1,m,p}
\end{aligned}$$

where $P(g,m,p)$ is the probability of find the gene activity characterized by g, m copies number of mRNA and p copies number of proteins. We first compute moments of variables, g, m and p and products of variable, gp and gp. Owing to the binary values of g, the moment of g^2 is equal to that of g.

$$\begin{aligned}
\langle g\rangle &= \sum_{g,m,p} g P_{g,m,p} = \sum_{m,p} 0 * P_{0,m,p} + \sum_{m,p} 1 * P_{1,m,p} = \sum_{m,p} P_{1,m,p} \\
\langle m\rangle &= \sum_{g,m,p} m P_{g,m,p} = \sum_{m,p} m P_{0,m,p} + \sum_{m,p} m P_{1,m,p} =< m >_0 + < m >_1 \\
\langle p\rangle &= \sum_{g,m,p} p P_{g,m,p} = \sum_{m,p} p P_{0,m,p} + \sum_{m,p} p P_{1,m,p} =< p >_0 + < p >_1 \\
\langle gp\rangle &=< p >_1 \\
\langle gm\rangle &=< m >_1
\end{aligned} \tag{B.2}$$

B.2 Equations for joint cumulants

According to the definition of moments, we deduce first equations for first-order and second-order moments:

$$\frac{d\langle g\rangle}{dt} = \theta(1-\langle g\rangle) - \alpha\langle gp\rangle \tag{B.3a}$$
$$\frac{d\langle m\rangle}{dt} = \lambda\langle g\rangle - \langle m\rangle \tag{B.3b}$$
$$\frac{d\langle p\rangle}{dt} = \theta(1-\langle g\rangle) - \alpha\langle gp\rangle + \beta\langle m\rangle - \delta\langle p\rangle \tag{B.3c}$$
$$\frac{d\langle m^2\rangle}{dt} = \lambda\langle g\rangle + 2\lambda\langle gm\rangle + \langle m\rangle - 2\langle m^2\rangle \tag{B.3d}$$
$$\frac{d\langle p^2\rangle}{dt} = \theta(2\langle p\rangle - \langle gp\rangle + 1 - \langle g\rangle) - \alpha(\langle gp\rangle - 2\langle gp^2\rangle) \tag{B.3e}$$
$$+\beta(2\langle mp\rangle + \langle m\rangle) + \delta\langle p\rangle - 2\delta\langle p^2\rangle \tag{B.3f}$$
$$\frac{d\langle gp\rangle}{dt} = \theta(1 + \langle p\rangle - \langle gp\rangle - \langle g\rangle) - \alpha\langle gp^2\rangle + \beta\langle gm\rangle - \delta\langle gp\rangle \tag{B.3g}$$
$$\frac{d\langle gm\rangle}{dt} = \theta(\langle m\rangle - \langle gm\rangle) - \alpha\langle gmp\rangle + \lambda\langle g\rangle - \langle gm\rangle \tag{B.3h}$$
$$\frac{d\langle mp\rangle}{dt} = -\alpha\langle gmp\rangle + \theta(\langle m\rangle - \langle gm\rangle) - \langle mp\rangle + \lambda\langle gp\rangle \tag{B.3i}$$
$$+\beta\langle m^2\rangle - \delta\langle mp\rangle \tag{B.3j}$$

As joint cumulants are defined as a special combination of moments, we are searching for the equations for joint cumulants with third-order cumulants vanishing.

$$\frac{d\langle g\rangle}{dt} = \theta(1-\langle g\rangle) - \alpha\langle gp\rangle \tag{B.4a}$$
$$\frac{d\langle m\rangle}{dt} = \lambda\langle g\rangle - \langle m\rangle \tag{B.4b}$$
$$\frac{d\langle p\rangle}{dt} = \theta(1-\langle g\rangle) - \alpha\langle gp\rangle + \beta\langle m\rangle - \delta\langle p\rangle \tag{B.4c}$$
$$\frac{d\langle \Delta_{gp}\rangle}{dt} = \theta(1-\langle g\rangle - \Delta_{gg} - \Delta_{gp}) + \beta\Delta_{gm} - \delta\,\Delta_{gp} \tag{B.4d}$$
$$-\alpha(\langle g\rangle\Delta_{pp} + \langle p\rangle\Delta_{gp} - \langle g\rangle\Delta_{gp} - \langle g\rangle^2\langle p\rangle) \tag{B.4e}$$
$$\frac{d\langle \Delta_{mp}\rangle}{dt} = -\theta\Delta_{gm} + \lambda\Delta_{gp} + \beta\Delta_{mm} - (\delta+1)\Delta_{mp} \tag{B.4f}$$
$$-\alpha(\langle g\rangle\Delta_{mp} + \langle p\rangle\Delta_{gm}) \tag{B.4g}$$
$$\frac{d\langle \Delta_{gm}\rangle}{dt} = \lambda\Delta_{gg} - (\theta+1)\Delta_{gm} - \alpha(\langle g\rangle\Delta_{mp} + \langle p\rangle\Delta_{gm}) \tag{B.4h}$$
$$\frac{d\langle \Delta_{pp}\rangle}{dt} = \theta(1-\langle g\rangle - 2\Delta_{gp}) + \beta(2\Delta_{mp} + \langle m\rangle) + \delta(\langle p\rangle - 2\Delta_{pp}) \tag{B.4i}$$
$$-\alpha(2\langle g\rangle\Delta_{pp} - \langle g\rangle\langle p\rangle - (1-2\langle p\rangle)\Delta_{gp}) \tag{B.4j}$$
$$\frac{d\langle \Delta_{mm}\rangle}{dt} = \lambda(\langle g\rangle + 2\Delta_{gm}) - 2\Delta_{mm} + \langle m\rangle \tag{B.4k}$$

By renormalizing time, variables and parameters according to:

$$\begin{array}{c} t \to T\delta_m \quad \delta \to \dfrac{\delta_p}{\delta_m} \quad < g > \to G \quad < p > \to \dfrac{\theta}{\alpha} P \quad < m > \to \dfrac{\delta\theta}{\beta\alpha} M \\ \Delta_{gg} \to \Delta_{GG} \quad \Delta_{pp} \to (\dfrac{\theta}{\alpha})^2 \Delta_{PP} \quad \Delta_{mm} \to (\dfrac{\delta\theta}{\beta\alpha})^2 \Delta_{MM} \quad \Theta \to \dfrac{\theta}{\delta_m} \\ \Delta_{gp} \to \dfrac{\theta}{\alpha} \Delta_{GP} \quad \Delta_{gm} \to \dfrac{\delta\theta}{\beta\alpha} \Delta_{GM} \quad \Delta_{mp} \to \dfrac{\delta}{\beta} (\dfrac{\theta}{\alpha})^2 \Delta_{MP} \\ \dfrac{\lambda\beta\alpha}{\delta\theta} \to \Lambda \quad \dfrac{\theta}{\alpha} \to P_0 \end{array} \tag{B.5}$$

Eqs. (B.4) can be rewritten in dimensionless form:

$$\frac{dG}{dt} = \Theta(1 - G - GP - \Delta_{GP}) \tag{B.6a}$$

$$\frac{dM}{dt} = \Lambda\, G - M \tag{B.6b}$$

$$\frac{dP}{dt} = \alpha(1 - G - GP - \Delta_{GP}) + \delta(M - P) \tag{B.6c}$$

$$\frac{d\Delta_{GP}}{dt} = \alpha[1 - 2G + G^2 - + G^2 P + G\Delta_{GP} - P_0(G\Delta_{PP} + \Delta_{GP} + P\Delta_{GP})] \tag{B.6d}$$

$$+\delta(\Delta_{GM} - \Delta_{GP}) \tag{B.6e}$$

$$\frac{d\Delta_{GM}}{dt} = \lambda G(1 - G) - \Delta_{GM} - \alpha\, P_0(\Delta_{GM} + g\Delta_{MP} + P\Delta_{GM}) \tag{B.6f}$$

$$\frac{d\Delta_{MP}}{dt} = -\alpha(\Delta_{GM} + G\Delta_{MP} + P\Delta_{GM}) + \Lambda\Delta_{GP} - (\delta + 1)\Delta_{MP} + \delta\Delta_{MM} \tag{B.6g}$$

$$\frac{d\Delta_{MM}}{dt} = \Lambda(\mu\, G + 2\Delta_{GM}) - 2\Delta_{MM} + \mu\, M \tag{B.6h}$$

$$\frac{d\Delta_{PP}}{dt} = \alpha[\frac{1}{P_0}(1 - G + \Delta_{GP} + GP) - 2(\Delta_{GP} + G\Delta_{PP} + P\Delta_{GP})] \tag{B.6i}$$

$$+\delta(2\Delta_{MP} - 2\Delta_{PP} + \frac{1}{P_0}(M + P)) \tag{B.6j}$$

Note that $P_0 = \theta/\alpha$ is defined as the number of proteins required to reduce the transcription rate by half. We assume this threshold is large ($P_0 >> 1$) which is equivalent to $\alpha \simeq 0$. Therefore, the time evolution equations for the first-order cumulants (i.e., averages) and second-

order cumulants (i.e., covariances) are given by:

$$\frac{dG}{dT} = \Theta(1 - G - GP - \Delta_{GP}) \tag{B.7a}$$

$$\frac{dM}{dT} = \Lambda\, G - M \tag{B.7b}$$

$$\frac{dP}{dT} = \delta(M - P) \tag{B.7c}$$

$$\frac{d\Delta_{GP}}{dT} = \delta(\Delta_{GM} - \Delta_{GP}) + G\Delta_{PP} + \Delta_{GP} + P\Delta_{GP}) \tag{B.7d}$$

$$\frac{d\Delta_{GM}}{dT} = \Lambda\, G(1 - G) - \Delta_{GM} + \Delta_{GM} + G\Delta_{MP} + P\Delta_{GM}) \tag{B.7e}$$

$$\frac{d\Delta_{MP}}{dT} = \Lambda\Delta_{GP} - (\delta + 1)\Delta_{MP} + \delta\Delta_{MM} \tag{B.7f}$$

$$\frac{d\Delta_{MM}}{dT} = 2\Lambda\Delta_{GM}) - 2\Delta_{MM} \tag{B.7g}$$

$$\frac{d\Delta_{PP}}{dT} = 2\delta(\Delta_{MP} - \Delta_{PP}) \tag{B.7h}$$

Bibliography

[1] I. Artsimovitch and R. Landick. Pausing by bacteria rna polymerase is mediated by mechanistically distinct signals. *Proc. Nat. Acad. Sci.*, 97:7090–7095, 2000.

[2] A. Aulehla and B. G. Herrmann. Segmentation in vertebrates: clock and gradient finally joined. *Genes Dev.*, 18:2060–2067, 2004.

[3] J. ay C. Dunlap. Molecular bases for circadian clocks. *Cell*, 96:271–290, 1999.

[4] L. Bai, A. Shundrovsky, and M. D. Wang. Sequence-dependent kinetic model for transcription elongation by rna polymerase. *J. Mol. Biol.*, 344:335–349, 2004.

[5] G. Balazsi, A. van Oudenaarden, and J. J. Collins. Cellular decision making and biological noise: from microbes to mammals. *Cell*, 144:910–925, 2011.

[6] R. L. Bar-Or, R. Maya, L. A. Segel, U. Alon, A. J. Levine, and M. Oren. Generation of oscillations by the p53-mdm2 feedback loop: A theoretical and experimental study. *Proc. Natl. Acad. Sci. USA*, 97:11250–11255, 2000.

[7] B. Barzel and O. Biham. Stochastic analysis of dimerization systems. *Phys. Rev. E*, 80:031117, 2009.

[8] B. Barzel and O. Biham. Binomial moment equations for stochastic reaction systems. *Phys. Rev. Let.*, 106:150602, 2011.

[9] A. Becskei and L. Serrano. Engineering stability in gene networks by autoregulation. *Nature*, 405:590–593, 2000.

[10] T. P. Bender, C. B. Thompson, and W. M. Kuehl. Differential expression of c-myb mrna in murine b lymphomas by a block to transcription elongation. *Science*, 237:1473–1476, 1987.

[11] By S. Bernard, B. Cajavec, L. Pujo-Menjouet, M. C.Mackey, and H. Herzel. Modelling transcriptional feedback loops: the role of gro/tle1 in hes1 oscillations. *Philos. Trans. Roy. Soc.*, 364:1155–1170, 2006.

[12] Y. Bessho, H. Hirata, Y. Masamizu, and R. Kageyama. Periodic repression by the bhlh factor hes7 is an essential mechanism for the somite segmentation clock. *Genes. Dev*, 17:1451–1456, 2003.

[13] R. D. Bliss, P. R. Painter, and A. G. Marr. Role of feedback inhibition in stabilizing the classical operon. *J. Theor. Biol.*, 97:177–193, 1982.

[14] S. Boireau. The transcriptional cycle of hiv-1 in real-time and live cells. *J. Cell. Biol*, 179:291–304, 2007.

[15] S. Borukhov, V. Sagitov, and A. Goldfarb. Transcript cleavage factors from escherichia coli. *Cell*, 72:459–466, 1993.

[16] N. E. Bouchler, U. Gerland, and T. Hwa. Nonlinear protein degradation and the function of genetic circuits. *Proc. Nat. Acad. Sci.*, 102:9559–9564, 2005.

[17] C. A Brackley, M. C. Romano, and M. Thiel. The dynamics of supply and demand in mrna translation. *PLos. Comput. Biol.*, 7:e1002203, 2011.

[18] D. F. Browning and S. J. W. Busby. The regulation of bacterial transcription initiation. *Nat. Rev.*, 2:1–9, 2004.

[19] L. Cai, C. K. Dalal, and M. B. Elowitz. Frequency-modulated nuclear localization bursts coordinate gene regulation. *Nature*, 455:485–491, 2008.

[20] L. Cai, N. Friedman, and X. S. Xie. Stochastic protein expression in individual cells at the single molecule level. *Nature*, 440:358–362, 2006.

[21] R. Cheong and A. Levchenko. Oscillatory signaling processes: the how, the why and the where. *Curr. Opin. Gene. Dev.*, 20:665–669, 2010.

[22] J. R. Chubb, T. Trcek, S. M. Shenoy, and R H. Singer. Transcriptional pulsing of a developmental gene. *Curr. Biol.*, 16:1018–1025, 2006.

[23] L. Ciandrini, I. Stansfield, and M. C. Romano. Role of the particle's stepping cycle in an asymmetric exclusion process: a model of mrna translation. *Phys. Rev. E*, 81:051904, 2010.

[24] L. J. Cook, R. K. P. Zia, and B. Schmittmann. Competition between multiple totally asymmetric simple exclusion process for a finite pool of resources. *Phys. Rev. E*, 80:031142, 2009.

[25] X. Darzacq, Y. Shav-Tal, V. de Turris, Y. Brody, S. M Shenoy, R. D. Phair, and R. H. Singer. In vivo dynamics of rna polymerase ii transcription. *Nat. Struct. Mol. Biol.*, 14:796–806, 2007.

[26] R. J. Davenport, G. J. L. Wuite, R. Landick, and C. Bustamante. Single-molecule study of transcriptional pausing and arrest by e.col rna polymerase. *Science*, 287:2497–2500, 2000.

[27] J. de Gier and F. H. L. Essler. Bethe ansatz solution of the asymmetric exclusion process with open boundaries. *Phys. Rev. Let*, 95:240601, 2005.

[28] B. Derrida. An exactly soluble non-equilibrium system: the asymmetric simple exclusion process. *Physics Reports.*, 301:65–83, 1998.

[29] B. Derrida, E. Domany, and D. Mukamel. An exact solution of a one-dimensional asymmetric exclusion model with open boundaries. *J. Stat.Phys.*, 69:667–687, 1992.

[30] B. Derrida, J. L. Lebowitz, and E. R. Speer. Exact free energy functional for a driven diffusive open stationary nonequilibrium system. *Phys. Rev. Let.*, 89:030601, 2002.

[31] M. Dobrzynski and F. J. Bruggeman. Elongation dynamics shape bursty transcription and translation. *Proc. Nat. Acad. Sci.*, 106:2583–2588, 2009.

[32] G. Dong and S. S. Golden. How a cyanobacterium tells time. *Curr. Opin. Microbiol.*, 11:541–546, 2008.

[33] Y. Dublanche, K. Michalodimitrakis, N. Kummerer, M. Foglierini, and L. Serrano. Noise in transcription negative feedback loops: simulation and experimental analysis. *Mol. Syst. Biol.*, page doi:10.1038/msb4100081, 2006.

[34] M. Dudzinski and G. M. Schutz. Relaxation spectrum of the asymmetric exclusion process with open boundaries. *J. Phys. A: Math.Gen.*, 33:8351–8363, 2000.

[35] A. Eldar and M. B. Elowitz. Functional roles for noise in genetic circuits. *Nat. Rev.*, 467:167–173, 2010.

[36] J. Elf and M. Ehrenberg. Fast evolution of fluctuations in biochemical networks with the linear noise approximation. *Genom. Res.*, 13:2475–2484, 2010.

[37] M. B. Elowitz and S. Leibler. A synthetic oscillatory network of transcriptional regulators. *Nature*, 403:335–338, 2000.

[38] M. B. Elowitz, A. J. Levine, E. D. Siggia, and P. S. Swain. Stochastic gene expression in a single cell. *Science*, 297:1183–1186, 2002.

[39] V. Epshtein and E. Nudler. Cooperation between rna polymerase molecules in transcription elongation. *Science*, 300:801–805, 2003.

[40] Historical Perspective Essay. Plant circadian rhythms. *The Plant Cell*, 18:792–803, 2006.

[41] A. Hoffmann et al. The ikb-nf-kb signaling module: temporal control and selective gene activation. *Science*, 298:1241–1245, 2002.

[42] A. Pare et al. Visualization of individual scr mrnas during drosophila embryogenesis yields evidence for transcriptional bursting. *Curr. Biol.*, 19:2037–2042, 2009.

[43] B. Sclavi et al. Real-time characterization of intermediates in the pathway to open complex formation by e. coli rna polymerase at the t7a1 promoter. *Proc. Nat. Acad. Sci.*, 102:4706–4711, 2005.

[44] C. V. Harper et al. Dynamic analysis of stochastic transcription cycle. *Plos. Biol.*, 9:e1000607, 2011.

[45] D. A. Gilchrist et al. Pausing of rna polymerase ii disrupts dna-specified nucleosome organization to enable precise gene regulation. *Cell*, 143:540–551, 2010.

[46] D. Bratsun et al. Delay-induced stochastic oscillations in gene expression. *Proc. Nat. Acad. Sci.*, 102:14593–14598, 2005.

[47] D. W. Austin et al. Gene network shaping of inherent noise spectra. *Nature*, 439:608–611, 2006.

[48] E. A. Abbondanzieri et al. Direct observation of base-pair stepping by rna polymerase. *Nature*, 438:460–465, 2005.

[49] E. Nudler et al. The rna-dna hybrid maintains the register of transcription by preventing backtracking of rna polymerase. *Cell*, 89:33–41, 1997.

[50] F. Toulme et al. In vivo evidence for back and forth oscillations of the transcription elongation complex. *EMBO J.*, 18:5052–5060, 1999.

[51] G. A. Perdrizet II et al. Transcriptional pausing coordinates folding of the aptamer domain and the expression platform of a ribswitch. *Proc. Nat. Acad. Sci.*, 109:3323–3328, 2012.

[52] G. Lahav et al. Dynamics of the p53-mdm2 feedback in individual cells. *Nat. Genet.*, 36:147–150, 2004.

[53] J. S. Takahashi et al. The genetics of mammalian circadian order and disorder: implication for physiology and disease. *Nat. Rev. Gene.*, 9:764–775, 2008.

[54] J. Zhou et al. Applied force provides insight into transcriptional pausing and its modulation by transcription factor nusa. *Mol. Cel.*, 44:635–646, 2011.

[55] K. Adelman et al. Single molecule analysis of rna polymerase elongation reveals uniform kinetic behavior. *Proc. Nat. Acad. Sci.*, 99:13538–13543, 2002.

[56] K. C. Neuman et al. Ubiquitous transcriptional pausing is independent of rna polymerase backtracking. *Cell*, 115:437–447, 2003.

[57] K. M. Herbert et al. Sequence-resolved detection of pausing by single rna polymerase molecules. *Cell*, 125:1083–1094, 2006.

[58] L. Zhu et al. Ca^{2+} oscillation frequency regulates agonist-stimulated gene expression in vascular endothelial cells. *J. Cell Sci.*, 121:2511–2518, 2008.

[59] M. Dundr et al. A kinetic framework for a mammalian rna polymerase in vivo. *Science*, 298:1623–1626, 2002.

[60] M. L. Dequeant et al. A complex oscillating network of signaling genes underlies the mouse segmentation clock. *Science*, 314:1595–1598, 2006.

[61] M. L. Kireeva et al. Nature of the nucleosome barrier to rna polymerase ii. *Mol. Cell*, 18:97–108, 2005.

[62] M. Voliotis et al. Fluctuations, pauses, and backtracking in dna transcription. *Biophys. J.*, 94:334–348, 2008.

[63] M. Voliotis et al. Backtracking and proofreading in dna transcription. *Phys. Rev. Let.*, 102:258101, 2009.

[64] M. Yoda et al. Roles of noise in single and coupled multiple genetic oscillators. *J. Chem. Phys.*, 126:115101, 2007.

[65] N. R. Forde et al. Using mechanical force to probe the mechanism of pausing and arrest during continuous elongation by escherichia coli rna polymerase. *Proc. Nat. Acad. Sci.*, 99:11682–11687, 2002.

[66] S. Bartfeld et al. High-throughput and single-cell imaging of nf-kb oscillations using monoclonal cell lines. *BMC. Cell Biol*, 11:21, 2010.

[67] T. Rajala et al. Effects of transcriptional pausing on gene expression dynamics. *Plos. Comput. Biol.*, 6:e1000704, 2010.

[68] T. Ushikubo et al. Testing the transition state theory in stochastic dynamics of a genetic switch. *Chem. Phys. Lett.*, 430:139–143, 2006.

[69] V. Elgart et al. Connecting protein and mrna burst distribution for stochastic models of gene expression. *Phys. Biol.*, 8:046001, 2011.

[70] Y. Drabsch et al. Mechanism of and requirement for estrogen-regulated myb expression in estrogen-receptor-positive breast cancer cells. *Proc. Nat. Acad. Sci.*, 104:13762–13767, 2007.

[71] Y. Taniguchi et al. Quantifying e. coli proteome and transcriptome with single-molecule sensitivity in single cells. *Science*, 329:533–538, 2010.

[72] N. Fedoroff and W. Fontana. Small numbers of big molecules. *Science*, 297:1129–1131, 2002.

[73] J. Forde and P. Nelson. Applications of strum sequences to bifurcation analysis of delay differential equation models. *J. Math. Anal. App.*, 300:273–284, 2004.

[74] H. B. Fraser, A. E. Hirsh, G. Giaever, J. Kumm, and M. B. Eisen. Noise minimization in eukaryotic gene expression. *Plos. Biol.*, 2:e137, 2004.

[75] L. Fu and C. Lee. The circadian clock: pacemaker and tumor suppressor. *Nat. Rev. Cancer*, 3:350–361, 2003.

[76] N. J. Fuda, M. B. Ardehali, and J. T. Lis. Defining mechanisms that regulate rna polymerase ii transcription in vivo. *Nat. Rev.*, 464:186–192, 2009.

[77] T. Gagatay, M. Turcotte, M. B. Elowitz, J. Garcia-Ojalvo, and G. M. Suel. Architecture-dependent noise discriminates functionally analogous differentiation circuits. *Cell*, 139:512–522, 2009.

[78] D. C. Galburt. Backtracking determines the force sensitivity of rnap ii in a factor-dependent manner. *Nature*, 446:820–823, 2007.

[79] T. S. Gardner, C. R. Cantor, and J. Collins. Construction of a genetic toggle switch in escherichia coli. *Nature*, 403:339–342, 2000.

[80] N. Geva-Zatorsky, N. Rosenfeld, S. Itzkovitz, R. Milo, A. Sigal, E. Dekel1, T. Yarnitzky, Y. Liron, P. Polak, G. Lahav, and U. Alon. Oscillations and variability in the p53 system. *Mol. Syst. Biol.*, page doi:10.1038/msb4100068, 2006.

[81] M. A. Gibson and J. Bruck. Efficient exact stochastic simulation of chemical systems with many species and many channels. *J. Phys. Chem. A*, 104:1876–1889, 2000.

[82] D. T. Gillespie. Exact stochastic simulation of coupled chemical reactions. *J. Phys. Chem.*, 81:2340–2361, 1977.

[83] A. Goldbeter. A model for circadian oscillations in the drosophila period protein (per). *Proc. R. Soc. Lond. B*, 261:319–324, 1995.

[84] A. Goldbeter. *Biochemical Oscillations and Cellular Rhythms*. Cambridge University Press, Cambridge, 1996.

[85] I. Golding, J. Paulsson, S. M. Zawilski, and E. C. Cox. Real-time kinetics of gene activity in individual bacteria. *Cell*, 123:1025–1036, 2005.

[86] C. A. Gomez-Uribe and G. C. Verghese. Mass fluctuation kinetics: Capturing stochastic effects in systems of chemical reactions through coupled mean-variance computations. *J. Chem. Phys.*, 126:024109, 2007.

[87] B. C. Goodwin. Oscillatory behavior of enzymatic control processes. *Adv. Enzyme Regul.*, 3:425–439, 1965.

[88] J. L. Gouze. Positive and negative circuits in dynamical systems. *J. Biol. Syst.*, 6:11–15, 1998.

[89] I. S. Gradshteyn and I. M. Ryzhik. *Tables of Integrals, Series, and Products*. Academic Press, San Diego, 2000.

[90] S. J. Greive and P. H. von Hippel. Thinking quantitatively about transcriptional regulation. *Nat. Rev.|Mol. Cel. Biol.*, 6:221–232, 2005.

[91] J. S. Griffith. Mathematics of cellular control processes i. negative feedback to one gene. *J. Theor. Biol.*, 20:202–208, 1968.

[92] R. Guantes and J. F. Poyatos. Dynamical principles of two-component genetic oscillators. *PLoS Comput. Biol.*, 2(3):e30, 2006.

[93] D. A. Hamstra. Real-time evaluation of p53 oscillatory behavior in vivo using bioluminescent imaging. *Cancer. Res.*, 66:7482–7489, 2006.

[94] L. H. Hartwell, J. J. Hopfield, S. Leibler, and A. W. Murray. From molecular to modular cell biology. *Nature*, 402:C47–C51, 1999.

[95] H. Hirata, S. Yoshiura, T. Ohtsuka, Y. Bessho, T. Harada, K. Yoshikawa, and R. Kageyama. Oscillatory expression of the bHLH factor Hes1 regulated by a negative feedback loop. *Science*, 298:840–843, 2002.

[96] S. Hockfiled. The next innovation revolution. *Science*, 323:1147, 2009.

[97] F. C. P. Holstege, U. Fiedler, and H. Th. M. Timmer. Three transitions in the rna polymerase ii transcription complex during initiation. *EMBO J.*, 16:7468–7480, 1997.

[98] J. E. M. Hornos, D. Schultz, G. C. P. Innocentini, J. Wang, A. M. Walczak, J. N. Onuchic, and P. G. Wolynes. Self-regulating gene: an exact solution. *Phys. Rev. E.*, 72:051907, 2005.

[99] G. C. P. Innocentini and J. E. M. Hornos. Modeling stochastic gene expression under repression. *J. Math. Biol.*, 55:413–431, 2007.

[100] M. H. Jensen, K. Sneppen, and G. Tiana. Sustained oscillations and time delays in gene expression of protein hes1. *FEBS Lett.*, 541:176–177, 2003.

[101] J. J.Tyson, K. C.Chen, and B. Novak. Sniffers, buzzers, toggles and blinkers: dynamics of regulatory and signaling pathways in the cell. *Curr. Opin. Cel. Biol.*, 15:221–231, 2003.

[102] J. J.Tyson, K. Chen, and B. Novak. Network dynamics and cell physiology. *Nature*, 2:908–916, 2001.

[103] John J.Tyson. *Computational cell biology*, volume 210. Spring, third edition, 2005.

[104] R. Karmakar and I. Bose. Stochastic model of transcription factor-regulated gene expression. *Phys. Biol*, 3:200–208, 2006.

[105] T. B. Kepler and T. C. Elston. Stochasticity in transcriptional regulation: origins, consequences, and mathematical representations. *Biophys. J.*, 81:3116–3136, 2001.

[106] H. Kimura, K. Sugaya, and P. R. Cook. The transcription cycle of rna polymerase ii in living cell. *J. Cell. Biol.*, 159:777–782, 2002.

[107] M. L. Kireeva and M. Kashlev. Mechanism of sequence-specific pausing of bacteria rna polymerase. *Proc. Nat. Acad. Sci.*, 106:8900–8905, 2009.

[108] S. Klumpp. Pausing and backtracking in transcription under dense traffic conditions. *J. Start. Phys.*, pages Doi 10.1007/s10955–011–0120–3, 2011.

[109] S. Klumpp and T. Hwa. Stochasticity and traffic jams in the transcription of ribosomal rna: intriguing role of termination and antitermination. *Proc. Nat. Acad. Sci.*, 105:18159–18164, 2008.

[110] M. Koern, T. C. Elston, W. J. Blake, and J. J. Collins. Stochasticity in gene expression: from theories to phenotypes. *Nat. Rev.*, 6:451–464, 2005.

[111] K. W. Kohn. Molecular interaction map of the mammalian cell cycle control and dna repair systems. *Mol. Biol. Cell*, 10:2703–2734, 1999.

[112] A. B. Kolomeisky, G. M. Schutz, G. M. Kolomeisky, and J. P. Straley. Phase diagram of one-dimensional driven lattice gases with open boundary. *J. Phys. A. Math. Gen.*, 31:6911–6919, 1998.

[113] N. Komissarova and M. Kashlev. Rna polymerase switches between inactivated and activated states by translocating back and forth along the dna and the rna. *J. Biol. Chem.*, 272:15329–15338, 1997.

[114] S. Krishna, M. H. Jensen, and K. Sneppen. Minimal model of spiky oscillations in nf-kappab signaling. *Proc. Nat. Acad. Sci.*, 103:10840–10845, 2006.

[115] J. Krug. Boundary-induced phase transition in driven diffusive systems. *Phys. Rev. Lett.*, 67:1882–1885, 1991.

[116] Y. Kuang. *Delay differential equations with applications in population dynamics*, volume 191. Academic Press, 1993.

[117] G. Lakatos and T. Chou. Totally asymmetric exclusion process with particles of arbitrary size. *J. Phys. A. Math. Gen*, 36:2027–2041, 2003.

[118] R. Landick. The regulatory roles and mechanism of transcriptional pausing. *Biochem. Soc. Trans.*, 34:1062–1066, 2006.

[119] R. Landick. Transcriptional pausing without backtracking. *Proc. Nat. Acad. Sci.*, 106:8797–8798, 2009.

[120] D. R. Larson, R. H. Singer, and D. Zenklusen. A single molecule view of gene expression. *Trends. Cel. Biol.*, 19:630–637, 2009.

[121] D. R. Larson, D. Zenklusen, B. Wu, J. A. Chao, and R. H. Singer. Real-time observation of transcription initiation and elongation on an endogenous yeast gene. *Science*, 332:475–478, 2011.

[122] A. Lederhendler and O. Biham. Validity of rate equation results for reaction rates in reaction networks with fluctuations. *Phys. Rev. E*, 78:041105, 2008.

[123] J. C. Leloup, D. Gonze, and A. Goldbeter. Limit cycle models for circadian rhythms based on transcriptional regulation in drosophila and neurospora. *J. Biol. Rhyth.*, 14:433–448, 1999.

[124] D. Lepzelter, H. Feng, and J. Wang. Oscillation, cooperativity, and intermediates in the self-repressing gene. *Chem. Phys. Lett.*, 490:216, 2010.

[125] I. Lestas, G. Vinnicombe, and J. Paulsson. Fundamental limits on the suppression of molecular fluctuations. *Nature*, 467:174–178, 2010.

[126] M. Levine. Paused rna polymerase ii as a developmental checkpoint. *Cell Review*, 145:502–511, 2011.

[127] J. Lewis. Autoinhibition with transcriptional delay: a simple mechanism for the zebrafish somitogenesis oscillator. *Curr. Biol.*, 13:1398–1408, 2003.

[128] M. E. Lidstrom and M. C. Konopka. The role of physiological heterogeneity in microbial population behavior. *Nat. Chem. Biol.*, 6:705–712, 2010.

[129] Y. Lu and F. R. Cross. Periodic cyclin-cdk activity entrains an autonomous cdc14 release oscillator. *Cell*, 141:268–279, 2010.

[130] M. C. Mackey and L. Glass. Oscillation and chaos in physiological control systems. *Science*, 4300:287–289, 1977.

[131] N. J. Marianayagam, M. Sunde, and J. M. Matthews. The power of two: protein dimerization in biology. *Trends in Biochemical Sciences*, 29, 2004.

[132] S. Martin and A. Pombo. Transcription factories: quantitative studies of nanostructures in the mammalian nucleus. *Chromosome Research*, 11:461–470, 2003.

[133] W. Mather, M. R. Bennett, J. Hasty, and L. S.Tsimring. Delay-induced degradate-and-fire oscillations in small genetic circuits. *Phys. Rev. Lett.*, 102:068105, 2009.

[134] D. A. McQuarrie. Stochastic approach to chemical kinetics. *J. Appl. Probab.*, 4:413–478, 1967.

[135] B. Mengel, A. Hunziker, L. Pedersen, A. Trusina, M. H. Jensen, and S. Krishna. Modelling oscillatory control in nf-kb, p53, and wnt signaling. *Curr. Opin. Gen. Dev.*, 20:656–664, 2010.

[136] N. Mitarai, I. B. Dodd, M. T. Crooks, and K. Sneppen. The generation of promoter-mediated transcriptional noise in bacteria. *PLoS Comput. Biol.*, 4:e1000109, 2008.

[137] N. A. M. Monk. Oscillatory expression of hes1, p53 and nk-κb driven by transcriptional time delays. *Curr. Biol.*, 13:1409–1413, 2003.

[138] P. E. Morant, Q. Thommen, F. Lemaire, C. Vandermoere, B. Parent, and M. Lefranc. Oscillations in the expression of a self-repressed gene induced by a slow transcriptional dynamics. *Phys. Rev. Lett.*, 102:068104–1–068104–4, 2009.

[139] L. G. Morelli and F. Jülicher. Precision of genetic oscillators and clocks. *Phys. Rev. Lett.*, 98:228101, 2007.

[140] B. Munsky, G. Neuert, and A. van Oudenaarden. Using gene expression noise to understand gene regulation. *Science*, 336:183–187, 2012.

[141] J.D. Murray. *Mathematical Biology*, volume 17. Springer, third edition, 2001.

[142] R. Murugan. Stochastic transcription initiation: Time dependent transcription rates. *Biophys. Chem.*, 121:51–56, 2006.

[143] G. Nair and A. Raj. Time-lapse transcription. *Science*, 332:431–432, 2011.

[144] D. E. Nelson, A. E. C. Ihekwaba, M. Elliott, J. R. Johnason, C. A. Gibney, B. E. Foreman, and G. Nelson. Oscillations in nf-kb signaling control the dynamics of gene expression. *Science*, 306:704–708, 2004.

[145] B. Novak and J. J.Tyson. Design principles of biochemical oscillators. *Nat. Rev.*, 9:981–991, 2008.

[146] J. Ohakubo. Approximation scheme based on effective interaction for stochastic gene regulation. *Phys. Rev. E*, 83:041915, 2011.

[147] Y. Okabe, Y. Yagi, and M. Sasai. Effects of the dna state fluctuation on single-cell dynamics of self-regulating gene. *J. Chem. Phys.*, 127:105107, 2007.

[148] P. Paszek, D. A. Jackson, and M. RH. White. Oscillatory control of signalling molecules. *Curr. Opin. Gen. and Dev.*, 20:670–676, 2010.

[149] H. Paul. Photon antibunching. *Rev. Mod. Phys.*, 54:1061–1102, 1982.

[150] J. Paulsson. Summing up the noise in gene networks. *Nature*, 427:415–418, 2004.

[151] J. Paulsson. Models of stochastic gene expression. *Phys. Life Rev.*, 2:157–175, 2005.

[152] J. M. Pedraza and J. Paulsson. Effects of molecular memory and bursting on fluctuations in gene expression. *Science*, 319:339–343, 2008.

[153] P. Pierobon, A. Parmeggiani, F. von Oppen, and E. Frey. Dynamics correlation functions and boltzmann-langevin approach for driven one-dimensional lattice gas. *Phys. Rev. E*, 72:036123, 2005.

[154] M. Ptashne. *A genetic switch*. COLD SPRING HARBOR LABORATORY PRESS, third edition, 2004.

[155] A. Raj, C. S. Peskin, D. Tranchina, D. Y. Vargas, and S. Tyagi. Stochastic mrna synthesis in mammalian cells. *PLos. Biol.*, 4:1707–1719, 2006.

[156] A. Raj, S. A. Rifkin, E. Andersen, and A. van Oudenaarden. Variability in gene expression underlies incomplete penetrance. *Nature*, 463:913–918, 2010.

[157] A. Raj and A. van Oudenaarden. Nature, nurture, or chance: stochastic gene expression and its consequences. *Cell*, 135:216–226, 2008.

[158] A. F. Ramos, G. C. P. Innocentini, and J. E. M. Hornos. Exact time-dependent solution for a self-regulating gene. *Phys. Rev. E.*, 83:062902, 2011.

[159] J. M. Raser and E. K. O'Shea. Noise in gene expression: origins, consequences and control. *Science Rev.*, 309:2010–2013, 2005.

[160] T. Roenneberg and M. Merrow. Circadian clocks– the fall and rise of physiology. *Nat. Rev. Mol. Cel. Biol.*, 6:965–971, 2005.

[161] M. C. Romano, M. Thiel, I. Stansfield, and C. Grebogi. Queueing phase transition: theory of translation. *Phys. Rev. Let.*, 102:198104, 2009.

[162] N. Rosenfeld, M. Elowitz, and U. Alon. Negative autoregulation speeds the response times of transcription networks. *J. Mol. Biol.*, 323:785–793, 2002.

[163] W. Ross and R. L. Gourse. Analysis of rna polymerase-promoter complex formation. *Methods*, 47:13–24, 2009.

[164] M. R. Roussel and R. Zhu. Stochastic kinetic description of a simple transcription model. *Bull. Math. Biol.*, 68:1681–1713, 2006.

[165] A. Sanchez and J. Kondev. Transcriptional control of noise in gene expression. *Proc. Nat. Acad. Sci.*, 105:5081–5086, 2008.

[166] M. Sasai and P. G. Wolynes. Stochastic gene expression as a many-body problem. *Proc. Nat. Acad. Sci.*, 100:2374–2379, 2003.

[167] A. Saunders, L. J. Core, and J. T. Lis. Breaking barriers to transcription elongation. *Nat. Rev.*, 7:557–567, 2006.

[168] M. Schnarr, P. Oertel-Buchheit, M. Kazmaier, and M. Granger-Schnarr. Dna binding properties of the lexa repressor. *Biochimie*, 73:423–431, 1991.

[169] D. Schultz, J. N. Onuchic, and P. G Wolynes. Understanding stochastic simulations of the smallest genetic networks. *J. Chem. Phys.*, 126:245102, 2007.

[170] M. Scott, T. Hwa, and B. Ingalls. Deterministic characterization of stochastic genetic circuits. *Proc. Nat. Acad. Sci.*, 104:7402–7407, 2007.

[171] J. W Shaevitz, E. A. Abbondazierl, R. Landick, and S. M. Block. Backtracking by single rna polymerase molecules observed at near-base-pair resolution. *Nature*, 426:684–687, 2003.

[172] L. B. Shaw, R. K. P. Zia, and K. H. Lee. Totally asymmetric exclusion process with extended objects: a model for protein synthesis. *Phys. Rev. E*, 68:021910, 2003.

[173] B. Snijder and L. Pelkmans. Origins of regulated cell-to-cell variability. *Nat. Rev.||Mol. Cel. Biol.*, 12:119–125, 2011.

[174] E. H. Snoussi. Necessary condition for multistationarity and stable periodicity. *J. Biol. Syst.*, 6:3–9, 1998.

[175] J. Stricker, S. Cookson, M. R. Bennett, W. H. Mather, L. S. Tsimring, and J. Hasty. A fast, robust and tunable synthetic gene oscillator. *Nature*, 456:516–520, 2008.

[176] D. M. Suter, N. Molina, D. Gatfield, K. Schneider, U. Schibler, and F. Naef. Mammalian genes are transcribed with widely different bursting kinetics. *Science*, 332:472–474, 2011.

[177] D. M. Suter, N. Molina, F. Naef, and U. Schibler. Origins and consequences of transcriptional discontinuity. *Curr. Opin. Cel. Biol.*, 23:657–662, 2011.

[178] H. Sutherland and W. Bickmore. Transcription factories: gene expression in unions. *Nat. Rev.*, 10:457–465, 2009.

[179] P. S. Swain, M. B. Elowitz, and E. D Siggia. Intrinsic and extrinsic contributions to stochasticity in gene expression. *Proc. Nat. Acad. Sci.*, 99:12795–12800, 2002.

[180] Y. Takashima, T. Ohtsuka, A. Gonzalez, H. Miyachi, and R. Kageyama. Intronic delay is essential for oscillatory expression in the segmentation clock. *Proc. Nat. Acad. Sci.*, 108:3300–3305, 2011.

[181] R. Z. Tan and A. van Oudenaarden. Transcript counting in single cells reveals dynamics of rdna transcription. *Mol. Syst. Biol.*, 6:358, 2010.

[182] G. Tang, R. Roy, R. P. Bandwar, T. Ha, and S. Patel. Real-time observation of the transition from transcription initiation to elongation of the rna polymerase. *Proc. Nat. Acad. Sci.*, 52:22175–22180, 2009.

[183] M. Thattai and A. van Oudenaarden. Intrinsic noise in gene regulatory networks. *Proc. Nat. Acad. Sci.*, 98:8614–8619, 2001.

[184] G. Tiana, S. Krishna, S. Pigolotti, M. H. Jensen, and K. Sneppen. Oscillations and temporal signalling in cells. *Phys. Biol.*, 4:R1–R17, 2007.

[185] Tsz-Leung To and N. Maheshri. Noise can induce bimodality in positive transcriptional feedback loops without bistability. *Science*, 327:1142–1145, 2010.

[186] T. Tripathi and D. Chowdhury. Interacting rna polymerase motors on a dna track: effects of traffic congestion and intrinsic noise on rna synthesis. *Phys. Rev. E*, 77:011921, 2008.

[187] T. Tripathi and D. Chowdhury. Transcriptional bursts: a unified model of machines and mechanisms. *The Frontiers of Physics*, 84:68004, 2008.

[188] J. J. Tyson, C. I. Hong, C. D.Thron, and B. Novak. A simple model of circadian rhythms based on dimerization and proteolysis of per and tim. *Biophys. J.*, 77:2411–2417, 1999.

[189] N. G. van Kampen. *Stochastic processes in physics and chemistry*. Elsevier, 2007.

[190] D. G. Vassylyev. Elongation by rna polymerase: a race though roadblocks. *Curr. Opin. Struct. Biol.*, 19:691–700, 2009.

[191] U. Vogel and K. F. Jensen. The rna chain elongation rate in escherichia coli depends on the growth rate. *J. Bacteriol.*, 176:2806–2813, 1994.

[192] P. H. von Hippel. An integrated model of the transcription complex in elongation, termination and editing. *Science*, 281:660–665, 1998.

[193] A. M. Walczak, J. N. Onuchic, and P. G. Wolynes. Absolute rate theories of epigenetic stability. *Proc. Nat. Acad. Sci.*, 102:18926–18931, 2005.

[194] A. M. Walczak and P. G. Wolynes. Gene-gene cooperativity in small networks. *J. Biophys.*, 96:4525–4541, 2009.

[195] Z. Wang and J. Zhang. Impact of gene expression noise on organismal fitness and the efficiency of nature selection. *Proc. Nat. Acad. Sci.*, page pnas.1100059108, 2011.

[196] D. J. Wilkinson. Stochastic modelling for quantitative description of heterogeneous biological systems. *Nat. Rev.|Gen.*, 10:122–133, 2009.

[197] P. Xie. Dynamics of backtracking long pauses of rna polymerase. *Biochimica et Biophysica Acta*, 1789:212–219, 2008.

[198] D. Zenklusen, D. R. Larson, and R. H. Singer. Single-rna counting reveals alternative modes of gene expression in yeast. *Nat. Struct. Mol. Biol.*, 15:1263–1271, 2008.

[199] V. P. Zhdanov. Model of gene transcription including the return of a rna polymerase to the beginning of a transcription cycle. *Phys. Rev. E*, 80:051925, 2009.

Abstract

Living cells can be viewed as dynamical systems which receive and process information from highly changing environments, then make appropriate decisions for survival and reproduction. Cellular dynamical properties and physiological functions essentially reply on nonlinear complex networks, called genetic networks in which macromolecules such as DNAs, mRNAs, proteins interact via biochemical reactions.

Motivated by the important roles and one striking dynamical behavior, oscillation, of genetic networks, this thesis work elucidates dynamical effects of time delay, fluctuation and transcriptional pausing on genetic networks, and thus mainly includes three studies.

The first study is about delay that is one of key ingredients of biological oscillation. In mathematical modeling, delay usually appears in an explicit manner (explicit delay) or is originated from a reaction step (reaction delay). By studying a minimal genetic network, a self-repressing gene involving various delays due to different biological processes, our results analytically derived from deterministic models reveal the combination principle of various and multiple delays. In particular, we find that reaction delays interact in a non-trivial way. Dynamical influences of two types of delay on oscillations are also compared. Explicit and reaction delays have the same influences only if delays are small or degradations are saturated. In the general cases, they display dramatical differences.

Genetic networks are usually subject to large fluctuations due to small copy numbers of reacting molecules present in cells. In order to investigate their dynamical influences on the oscillatory behavior, in the second study we then develop a cumulant expansion of the master equation and apply it to the self-repressing gene circuit. We find that fluctuations shift significantly the averages of molecular quantities predicted by deterministic models and induce oscillations without requiring the nonlinear degradation.

Transcriptional pausing is a typical unusual behavior of RNA Polymerase enhancing stochasticity and modulating transcription rate. In the third study, we investigate the dynamical effects of pausing on transcription in using a classical system in out-of-equilibrium statistical physics, so-called TASEP model with periodic boundary condition. In the limit case where pause duration is short, we can still construct a mean-field model to analyze the transcription rate and site occupation. In the general case where mean-field approach no long applies, we obtain a good understanding of various mechanisms driving the transcription dynamics over the entire range of pause duration. Importantly, by using a statistical approach, we find the theoretical prediction of transcription rate that agrees well with numerical simulations.

Résumé

Les cellules vivantes peuvent être considérées comme des systèmes dynamiques qui reçoivent et traitent les informations d'un environnement complexe et changeant, puis prennent des décisions appropriées pour survivre et proliférer. Les propriétés dynamiques et les fonctions physiologiques des cellules sont essentiellement régulées par des réseaux d'interaction appelés réseaux génétiques au sein desquels des macromolécules telles que les ADNs, ARNs, protéines interagissent via des réactions biochimiques.

Motivé par l'existence de comportements dynamiques récurrents des réseaux génétiques tels que les oscillations, ce travail de thèse étudie les effets des délais, des fluctuations et des pauses transcriptionnelles sur la dynamique des réseaux génétiques, et s'articule donc autour de trois études.

La première étude concerne le rôle des délais qui sont des ingrédients clés des oscillations biologiques. Dans la modélisation déterministe, le délai apparaît généralement de manière explicite (délai explicite) ou est issue d'une étape de réaction (délai réactionnel). En étudiant un réseau génétique comprenant un gène auto-réprimé qui contient divers délais, nos résultats montrent analytiquement le principe de combinaison des différents délais. En particulier, les délais réactionnels interagissent d'une manière non triviale. Les influences des délais explicites et réactionnels sur les oscillations sont également comparées. Les délais explicites et réactionnels ont les mêmes influences que si les délais sont petits ou les dégradations sont saturées.

La seconde étude s'intéresse à l'impact des fluctuations dues aux petits nombres des molécules présentes dans les cellules. Afin d'étudier leur influences dynamiques sur les oscillations, nous proposons un développement de cumulants de l'équation maîtresse et l'appliquons au circuit de gène auto-réprimé. Nous trouvons que les fluctuations modifient significativement les moyennes des quantités moléculaires prévues par les modèles déterministes, et induisent des oscillations sans avoir recours à la dégradation non-linéaire.

Enfin, la troisième étude concerne le processus de pause transcriptionnelle qui est un comportement typique de l'ARN Polymérase. Nous analysons les effets dynamiques de la pause sur la transcription à l'aide d'un système classique en physique statistique : le modèle TASEP avec condition aux limites périodique. Pour des durées des pauses intermédiaires et longues pour lesquelles l'approche de champ moyen n'est pas validée, nous parvenons néanmoins à une bonne compréhension des différents mécanismes qui contrôlent la dynamique de transcription. Une approche statistique permet d'obtenir une description quantitative du taux de transcription en bon accord avec les simulations numériques.