Politecnico di Torino

Beyond Steady-State: First-Passage Times in microRNA-mediated Gene Regulation

Supervisors:
Prof. Carla Bosia
Dr. Silvia Grigolon

Candidate:
Raffaella Renzulli
s330822

Tutor:
Mathéo Aksil

A.A. 2024-2025

Contents

Introduction

Every cell within an organism must execute a multitude of tasks essential for survival, such as cellular development and differentiation. These fundamental processes, despite being driven by intrinsically stochastic biochemical reactions, must reliably achieve their intended outcomes. To ensure this reliability, besides precise molecular levels, the cell must maintain a precise temporal ordering of events by means of various regulatory mechanisms acting on gene expression.

In order to study event timing, a threshold-crossing framework is typically established. In this context, a biological process is initiated when the concentration of a specific chemical species reaches a predetermined threshold. However, the probabilistic nature of gene expression introduces significant fluctuations in the molecule counts. This variability, or noise, is a critical factor and it can be categorized into intrinsic or extrinsic noise. The latter stems out from fluctuations in global cellular conditions or external factors such as cell size, number of ribosomes, environmental conditions, etc., while intrisic noise arises from the inherent stochasticity of biochemical reactions themselves (e.g., the random timing of transcription and translation events) and therefore cannot be neglected.
This total noise directly translates into variability in the threshold-crossing time. To ensure proper event timing, the cell must employ strategies to control this variability and manage noise.
One crucial control mechanism operates by regulating the availability of mRNA molecules. As the direct precursors to proteins, which are the primary effectors of cellular tasks, mRNA levels are fundamental. A key class of regulators in this context is microRNAs (also called miRNAs). These are small non-coding RNA molecules which post-transcriptionally regulate gene expression by binding to target mRNAs, typically leading to their degradation or more simply to the inhibition of their translation.

The aim of this thesis is to investigate the distribution of threshold-crossing times, known as the *First Passage Time (FPT) distribution*, of mRNA molecules in the context of a miRNA-based genetic circuit. This focus on the transient dynamics, specifically "how long" it takes to reach a determined state, represents a critical departure from many traditional analysis, which have focused solely on the steady-state properties of gene expression.

Particularly, this thesis is divided into three Parts.

The first one provides all biological information necessary to understand the gene expression mechanism and its post-transcriptional regulation.
Then, in the second one, the mathematical models—known as genetic circuits—used to analyze biochemical processes are presented. Both a deterministic and a stochastic analysis is carried out, providing the equations describing the evolution of the system under study.
As these circuits may be very complex, the corresponding mathematical equations describing their evolution in time may be hard-to-solve, therefore approximation methods are usually implemented. The one carried out in this study is the Van Kampen Expansion, a powerful method that allows to gain an approximation for the probability of having a number of molecules at a certain time.
In this context, the First-Passage Time definition is introduced as the initiation time of a biological process. Also, the more formal definition of the FPT distribution is given, connecting it to the probability of having a given number of molecules at a certain time, therefore highlighting the importance and the need for an approximation method such as the Van Kampen one.

All concepts presented in Part II are introductory as they provide fundamental insights on the mathematical information necessary to analyze First-Passage Time distribution related to miRNA-regulated gene expression mechanism. To carry out this analysis on microRNA post-transcriptional regulation, a specific miRNA-mediated genetic circuit is introduced and both deterministic and stochastic approaches are here applied to recover the equations describing its dynamics.

The third Part of the work focuses solely on the study of the First-Passage Time distribution related to the miRNA-mediated circuit. Both numerical and analytical approaches are illustrated with the purpose of studying the First-Passage Time distribution of mRNA molecules involved in the system.
The numerical simulations—based on a fundamental method for simulating stochastic processes, i.e. the Gillespie Algorithm—allow for the direct numerical construction of the First-Passage Time distribution for the circuit and, therefore, the analysis of its properties, such as mean, variance, etc., both in the regulated and unregulated case.
The analytical approach is then carried out with the aim of obtaining an explicit expression of the First-Passage Time distribution. The regulated and unregulated cases are treated separately due to the complexity of the regulated system

which includes the non-linear interactions between mRNA and miRNA. An approximation is therefore necessary to obtain the analytical FPT distribution, as briefly stated in Part II; to this end the Van Kampen expansion, also known as the Linear Noise Approximation, is developed for the miRNA-mediated genetic circuit of interest.
After the analytical calculation, the numerical and analytical results are illustrated and commented.

In general, this analysis of the First-Passage Time distribution, unlike traditional steady-state analysis, offers a powerful framework to further understand miRNA regulation and its role in gene expression. Focusing on the dynamics of threshold-crossing processes, this work reveals that the miRNA-mediated regulation can significantly lower the mean FPT while maintaining small noise. This finding highlights the possibility of achieving a faster response—lower FPT—when regulation is optimal, thereby confirming the function of miRNAs as powerful noise buffers in addition to their known role in fine-tuning.

Part I
Biological context

This study is at the edge of physics and biology, as mathematical tools will be employed to comprehend and model a biological mechanism characterizing all cells of all living organisms. To this end, this First Part will provide a biological context necessary to understand the problem setting.

Particularly, the cell acts as the fundamental unit of biological organization, operating as a sophisticated machine capable of self-replication and adaptation. While inside a cell there is the complete library of instructions necessary for an organism, the mere presence of these instructions is insufficient to sustain life. The cell must actively select, decode, and execute specific segments of this genetic code to produce the molecular machinery required for survival.
This fundamental biological imperative—translating the encoded instructions of the genotype into the observable functional traits of the phenotype—is governed by the mechanism of *gene expression*.
However, the complexity of life requires more than a production line. It demands precise control. In fact, the linear flow of genetic information is overlaid by intricate regulatory networks, which may operate at multiple levels.

The following Chapters provide a comprehensive overview of this biological framework. First the Central Dogma of molecular biology will be analysed, explaining where information is kept and how it can be retrieved and utilized by the cell. Subsequently, the critical role of gene regulation is introduced, with specific focus on microRNA molecules—small non-coding molecules that have emerged as fundamental ingredients of post-transcriptional control.

1 The mechanism of gene expression

The ability of cells to store and translate the genetic instructions required to sustain a living organism is fundamental to life. These instructions are stored in all living cells in the same way, as they are all contained in a nucleic acid, the DNA (deoxyribonucleic acid). [2]
In order to read and use it, DNA must be translated into macromolecules that may actually carry out the relevant biological functions needed by each cell. The most popular macromolecules of this type are proteins, as they play different functions in several processes, such as transport of molecules or structural support.

Particularly, the process allowing for protein production from DNA, known as *gene expression*, can be coarsely summarized into two main steps: i) transcription and ii) translation.

1.1 What is DNA?

DNA is the fundamental macromolecule containing all genetic information. A DNA molecule is formed by many subsequent building blocks, called *nucleotides*, made of a phosphate group, a sugar and a base. The sugar characterizing this nucleic acid is deoxyribose and the base could be any of the following four: adenine, thymine, cytosine and guanine. [1]
Nucleotides are not symmetric, therefore they can be covalently linked together forming a long poly-nucleotide chain with a sugar-phosphate backbone and exposed basis. Since the way in which the nucleotides are attached gives to each DNA strand chemical polarity, a DNA molecule has a very specific 3d conformation, where two antiparallel strands are linked through hydrogen bonds between paired bases, namely a double helix. The complementary base-pairing, also referred to as the Watson-Crick base-pairing [26], leads to the lowest energy configuration, in fact hydrogen bonds form efficiently without deforming the double helix only between adenine and thymine (with 2 hydrogen bonds) and between guanine and cytosine (with 3 hydrogen bonds).

All information contained in DNA, known as *genome*, must pass on from a cell to its daughter, therefore DNA must be accurately copied to ensure the genome transmission, which, as briefly noted before, primarily consists of instructions for making proteins. These macromolecules are fundamental for cellular life, as they are involved in most of its functions, from catalyzing chemical reactions to gene regulation. Their production process is schematically repre-

sented in Figure 1, and it is referred to as *central dogma of biology*. It consists into transcription and translation of DNA, as fully explained in the following.

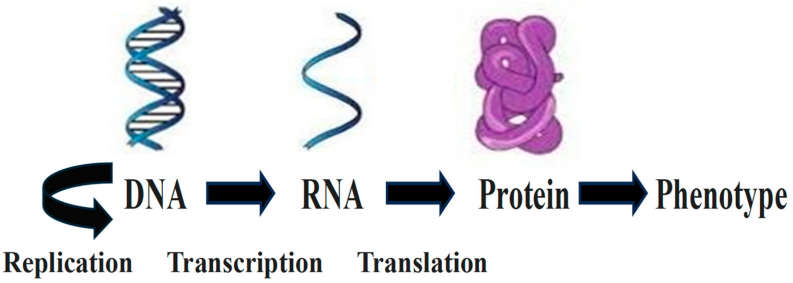


Figure 1: Schematic representation of gene expression, adapted from the work of Haseltine & Patarca [13].

1.2 Transcription: from DNA to RNA

Transcription is the first step of gene expression and it is a mechanism that allows the creation of a nucleic acid very similar to DNA, the RNA (ribonucleic acid), used as an information carrier throughout the cell. [1]
The differences between these nucleic acids are mainly two: the type of sugar, which is changed to ribose, and a base pair, as thymine is substituted by uracil in RNA. The fact that RNA has a different sugar with respect to DNA has a huge impact on its conformation, in fact, it is usually single-stranded, often containing only stacks of canonical base pairs.
The process of transcription is very similar to DNA duplication, as it begins with the unwinding and opening of a DNA molecule in order to expose the bases. Then, one of the two DNA strands serves as a template, and RNA nucleotides are added one by one to form a chain that is complementary to this template strand. In contrast to a newly formed DNA strand, the RNA chain is displaced just beyond the region where new nucleotides are added and DNA double-helix is restored.
All these steps are performed by enzymes called RNA polymerases, which catalyze the link between different nucleotides in the RNA chain. The process differs between bacteria and eukaryotes, as in the former RNA polymerase produces RNA molecules, starting and stopping at specific spots on the DNA, respectively called promoter and terminator. Instead, in eukaryotes there is more than one RNA polymerase, each of which performs a different function, and there are many ingredients, called *general transcription factors* (TFs), that are involved in the positioning of the polymerase on the promoter.
There exist many types of RNA molecules, such as messenger RNA (mRNA),

transfer RNA (tRNA), micro-RNA (miRNA), but only the messenger RNA is protein coding, meaning it carries genetic information. All other types of non-coding RNAs may take part in the process of protein production fulfilling distinct functions, such as transport or catalytic. In order to produce mRNA molecules in eukaryotes, transcription is only the first of several steps needed, such as covalent modification of the ends of mRNA (capping, cleavage and polyadenylation) and splicing, which consists into the removal of all non-coding regions form the mRNA molecule just transcribed. [12]

Furthermore, since transcription involves only a limited region of DNA, particularly only genes necessary to perform a given task are read, RNA molecules are typically much shorter than DNA ones: DNA molecules in human can have up to 250 million nucleotide base-pairs, while RNA ones are no more than a few thousand nucleotides long [1].

1.3 Translation: from RNA to protein

The mechanism to go from a nucleotide sequence of a mRNA to protein is referred to as *translation*. Proteins are indeed made of amino-acids, not nucleotides, therefore a whole process of association of amino-acids to nucleotides is needed.

The rules through which a sequence of mRNA is translated into a sequence of amino-acids are known as the *genetic code* and the main concept is that the sequence of nucleotides is read in consecutive groups of three, ensuring 64 possible combinations (AAA, AUG, etc.). This code is evidently redundant, as only twenty amino-acids are commonly found in proteins, and therefore some amino-acids are specified by more than one triplet. This is also known as the *coding problem*.[4]

Translation takes place in the cytoplasm, on a complex catalytic machine, the ribosome, made by different proteins and several RNA molecules (rRNAs). The ribosome is characterized by two subunits, only joined together around a mRNA molecule when a protein is being synthesized. Translation is carried out thanks to an adaptor that recognizes and binds to both the triplet of nucleotides, called *codon*, and the corresponding amino-acid. The molecule responsible for this process is the transfer RNA (tRNA).
It all starts when initiator tRNA—which always carries a particular amino-acid

(methionine)—binds to the small subunit of the ribosome, so that this complex begins the reading of the mRNA molecule. As soon as the start codon (AUG) is matched with initiator, the large ribosomal subunit binds to the small one and translation begins with the first bond between methionine and the subsequent amino-acid. The process ends when one of the stop codons (UAA,UAG, UGA) is recognized, as they do not represent any amino-acid, and the protein is immediately released.
Usually proteins are made on poly-ribosomes, namely large assemblies of ribosomes, so that multiple initiations take place on each mRNA molecule.

During the translation, long chains of amino acids are created, called polypeptides since the peptide bond is the one connecting two amino acids. As a polypeptide starts to exit the ribosome, so during its production, it folds with the aid of molecular chaperons. [12]
Proteins are present in almost all essential biological processes for the cell, covering a multitude of possible functions, such as transport, motility, catalysis, storage. etc.

2 Gene regulation

Transcription and translation are the main mechanisms through which cells express their genetic instructions. From a single gene, many identical RNA molecules can be produced and similarly from each RNA molecule many identical proteins are synthetized, allowing the cell to generate a large number of several different proteins. Furthermore, the efficiency of transcription of each gene can be modulated, so that cells are able to produce large quantities of some proteins and small quantities of others. In this way, regulation of gene expression is tailored to the cell needs at any given moment.
In order to get a better insight into this concept, let us consider the cells of a multicellular organism: they share the same genome but differ dramatically in both structure and function. Typically, a cell expresses only a fraction of its genes and can also change its expression pattern in response to environmental change.

As noted before, gene expression is a multi-step process, so its regulation may occur at any stage of the pathway from DNA to RNA to protein. [1]
Therefore, a cell may manage the amount of protein produced by controlling when and how a given gene is transcribed. This mechanism is known as *transcriptional control* and it allows the enhancement or the repression of a gene transcription through regulatory proteins called *transcription factors* (TFs). These factors bind to specific regulatory regions of DNA, as gene regulatory proteins contain motifs that recognize DNA regulatory sequences without breaking the double helix structure. Once TFs are bound to DNA, the corresponding gene is either transcribed or repressed, ensuring that the respective genetic information is either used or kept for future use.
However, there exist other types of gene regulation managed by RNA molecules directly. Cells may indeed select which mRNAs are to be exported outside the nucleus or they may destabilize some mRNAs directly in the cytoplasm. A cell can also perform protein activity control, therefore activating or degrading proteins after their synthesis. [1]

2.1 mRNA degradation by post-transcriptional control

It has been known for long time that the whole genome does not contain only coding regions, i.e., regions of nucleotides which do not encode proteins. Before the 90s, these regions where called *junk DNA*. Yet, we now know that inside cells there are several types of non-coding RNA molecules performing many

functions besides being intermediate carriers of genetic information, such as regulating the flow of information from DNA to protein.

In particular, one type of short RNA molecule, called microRNA or miRNA, plays a crucial role in regulation of gene expression in plants and animals at a post-transcriptional level. MicroRNAs are small (near 22 nucleotides) non-coding RNAs that may bind to specific mRNAs with the purpose of degrading them or just inhibiting their translation. [1]
Different outcomes are possible as they depend on how extensive the base-pairing between miRNA and mRNA is. On the one hand, with low complementarity, translation is repressed and eventually the mRNA molecule is destabilized; on the other hand, with high complementarity, miRNA induces rapid mRNA degradation. These microRNAs play the fundamental role of post-transcriptional regulators and even a single miRNA can regulate a set of different mRNAs and, viceversa, many miRNAs can also bind to the same mRNA molecule leading to further reductions in its translation [17].
This mechanism has a huge impact on the output of gene expression, by either switching it off completely or by non-trivially modulating the final protein distribution [8].

2.2 Discovery of miRNAs

MicroRNAs were first discovered in 1993 [14] in the nematode worm *C.elegans*, where the gene *lin-4*, until then characterized as one of the genes controlling the developmental timing of the worm, was shown to produce a small RNA that regulated the expression of another gene, *lin-14*, by binding its mRNA. This marked the first known example of gene regulation by small non-coding RNAs in animals and, initially, it was thought to be a nematode-specific mechanism. Thanks to the discovery of another gene, *let-7*, which performs the same role as *lin-4*, the previous hypothesis was rejected as *let-7* gene is not only present in the nematode worm but in many other species in a nearly identical way, including humans, mice and flies [19]. This discovery confirmed that microRNAs are a widespread, evolutionarily conserved class of post-transcriptional gene regulators. There are many types of miRNAs: humans alone express more than 400 different types, and new miRNAs are still being discovered.

2.3 Biogenesis of miRNAs

The canonical pathway of miRNA biogenesis is the most common way for miRNA production in a cell. It starts with the transcription from DNA of primary miRNAs or pri-miRNAs, long transcripts that contain one or more hairpin structures (i.e., secondary structures that form when the RNA strand folds back on itself due to complementary base-pairing). Still in the nucleus, these are then processed with the aid of the microprocessor complex made by a RNA binding protein (DGCR8) and a ribonuclease III enzyme (Drosha) into precursor miRNAs or pre-miRNAs. During the processing from pri-miRNAs to pre-miRNAs all flanking sequences are cleaved out by the complex, only the hairpin structures of primary-transcript are retained.

Once pre-miRNAs are formed, they are exported into the cytoplasm and processed once again by a ribonuclease III enzyme (Dicer) so that the terminal loop of the hairpin is removed, generating a 22-nucleotide long mature miRNA duplex. One strand of the duplex, referred to as the *guide strand*, is loaded into an Argonaute (AGO) protein, while the other strand, known as the *passenger strand*, is unwound from the complex. The RNA-induced silencing complex or RISC is formed and it will seek out mRNAs with complementary nucleotides sequences. [17]

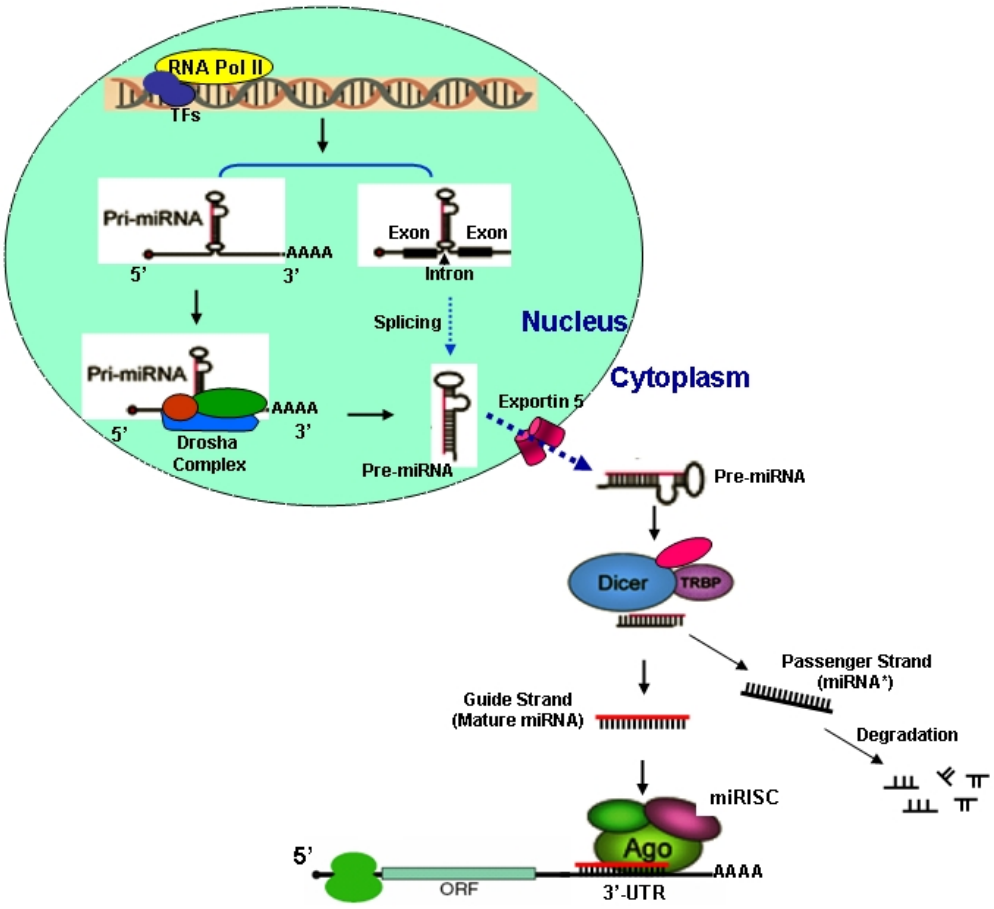


Figure 2: Schematic representation of miRNA biogenesis adapted from Zhu et al. [27].

It was found that there are many other ways in which miRNA molecules are produced, which are referred to as non canonical miRNA biogenesis pathways. These can be divided into two main categories: *Drosha-independent pathways*, which do not employ the microprocessor complex, and *Dicer-independent pathways*, in which Drosha's products are directly processed by AGO proteins instead of Dicer. These alternative pathways, while using different combinations of canonical proteins, can produce functional miRNA molecules in a cell. [17]

In this first Part, the fundamental mechanism of gene expression is outlined, specific emphasis placed on gene regulation.
The comprehensive examination of microRNAs—from their discovery and biogenesis to their function in mRNA degradation—illustrates a crucial biologial concept: the components of gene expression do not act in isolation but form a vast, intricate, interconnected network.
A qualitative description is insufficient to fully understand how miRNA *post-transcriptional control* modulates the final protein output, therefore to rigorously quantify these dynamics, we must move beyond descriptive biology and employ the analytical tools of systems biology and mathematical modeling [15].

To this end, the following Part of this study introduces the mathematical framework used to describe these biological processes.
To fully understand how the modeling of gene expression is carried out, first the foundational system—the central dogma of biology (Figure 1)—is analyzed. This, in fact, will serve as the baseline for understanding more complex systems where also regulatory activity is included.

Part II
Genetic circuits analysis

In the previous Chapters, the fundamental molecular mechanisms of gene expression, namely transcription and translation, were described. In living organisms, these processes do not occur in isolation. They are part of extremely complex biochemical networks, in which a vast number of components interact, such as genes, RNAs, proteins and regulatory molecules.[12]

To quantitatively study the dynamical and stationary properties of gene expression and its regulation, it is necessary to simplify this complexity and employ quantitative mathematical methods. This may be achieved by building mathematical models that focus only on key "ingredients" and their interactions. These simplified representations, which capture the fundamental logic of the system of interest, are known as *genetic circuits*. The analysis of these circuits can vary in complexity, ranging from the simplest model—a single gene expressed without regulation describing the flow of information from DNA to protein—to more complex circuits that include control and feedback mechanisms.

Part II of this work introduces the framework for analyzing these systems in a more quantitative way. Accordingly, Chapter 3 examines the simple single-gene system, a circuit where no control activity is modeled for simplicity.
Firstly, a deterministic study will be carried out, followed by a stochastic analysis, which, as will be discussed, proves to be more suitable for the description of genetic circuits. This stochastic description represents the system using the probability of it being in a given state—specifically in our case this is the probability of having a given number of molecules—at a certain time, which will satisfy the so-called Master Equation. This is usually a hard-to-solve equation for a generic circuit, but thanks to some mathematical tools, at least the moments of the distribution are accessible; these are recovered explicitly for the single-gene circuit.

Then, the main topic of this study is introduced, i.e.: the *First-Passage Time (FPT) Distribution*. This is the distribution of times at which the molecule number of a given chemical species crosses a predetermined threshold. It is fundamental to comprehend how this distribution changes, as many biological processes inside a cell are modeled to start whenever a given chemical species is abundant enough. A small variability in the crossing time may ac-

tually have huge biological consequences, therefore cells must somehow ensure the correct event timing. Particularly, the First-Passage Time distribution is mathematically connected to the probability of being in a state at a given time, highlighting the importance of finding a solution to the Master Equation in order to gain an analytical estimate for the distribution of interest.
When a solution is not achievable, approximation methods may be employed to recover an analytical expression for the probability of being in a state at a certain time. In this study a particular approximation will be implemented, i.e. the Van Kampen expansion, and it is in the following introduced.

Chapter 4 proceeds then into the description and analysis of a more complex genetic circuit, where regulation is included. Particularly, the circuit considered models the post-transcriptional control activity mediated by microRNA molecules.
This type of circuit and its analysis are crucial to understand the role of microRNA molecules in regulation of gene expression and were abundantly previously analyzed. This study though marks a clear departure from previous works, as the main focus are not the steady-state properties of the system but its dynamical ones.

3 Single-gene circuit analysis

3.1 Genetic circuits as deterministic dynamical systems

Genetic circuit models are mathematically formulated by means of differential equations, which describe the change over time of a species concentration according to the mass-balance conservation law: $change = flux_{in} - flux_{out}$.

For the simplest case, a single unregulated gene, four fundamental reactions can be considered: i) mRNA production, ii) mRNA degradation, iii) protein production and iv) protein decay. In this way, the model describes effectively the main steps of gene expression described in Part I. The four processes are represented in the following as elementary chemical reactions:

- Transcription: $\emptyset \xrightarrow{k_R} R$
- mRNA Degradation: $R \xrightarrow{g_R} \emptyset$
- Translation: $R \xrightarrow{k_P} R + P$
- Protein Degradation: $P \xrightarrow{g_P} \emptyset$

where k_R is the transcription rate and g_R is the degradation rate of mRNA (R), while k_P is the translation rate and g_P is the degradation rate of the protein (P).

Applying the mass-balance law to this reaction scheme yields a set of differential equations for the concentration of the species involved, namely R(t) and P(t),

$$\begin{aligned} \frac{dR}{dt} &= k_R - g_R R, \\ \frac{dP}{dt} &= k_P R - g_P P, \end{aligned} \tag{1}$$

whose unique, thus stable, steady-state solution is given by:

$$R_{ss} = \frac{k_R}{g_R}; \;\; P_{ss} = R_{ss}\frac{k_P}{g_P} = \frac{k_R k_P}{g_R g_P}. \tag{2}$$

This deterministic ordinary differential equation (ODE) framework for modeling gene expression involves major several simplifications.

Firstly, the concentrations of reactants must evolve continuously and differentiably to allow the use of differential equations. In order to treat concentrations as continuous real-valued variables, the number of reactants must be large enough, fact that is not typically true in the cellular environments.

Furthermore, diffusion is assumed to be infinitely fast, allowing us to use concentrations (moles per volume) and ignore the physical location and movement of individual molecules within the cell. This amounts to assume the system is "well-mixed" or spatially homogeneous.

Moreover, this approach is deterministic, meaning that it completely neglects the intrinsic stochasticity or noise of all biochemical reactions.

In fact, biochemical events occur at random times because of the discreteness of molecules and because molecular interactions at the cellular scale are governed by stochastic collisions between diffusing molecules. The timing of these encounters depends on chance rather than determinism, especially when the number of reacting molecules is rather small as it happens in actual living cells. [10]

As a result, the occurrence of each reaction event is probabilistic, leading to intrinsic noise in biochemical networks. In addition, extrinsic noise arising from fluctuations in cellular conditions, such as changes in enzyme levels, temperature, energy availability, etc., may affect as well all reactions simultaneously. This is a major source of variability between different cells in a population.[7]

With the aim of correctly describing the mechanisms underlying gene expression, the deterministic approach is normally replaced in favor of a stochastic one.

3.2 Stochastic approach: the Master Equation

As briefly noted before, within the cell, the number of key reactants in gene expression tends to be of the order of 10-1000, so any reaction altering the molecule number by one or two actually generates a large relative change, much so that the concentration evolves step-wise.

With the purpose of correctly describing gene expression for small pools of reactants, the deterministic approach is left in favor of a stochastic one, where each reaction is considered as probabilistic. Instead of mass-balance, the conservation law now involved is *probability balance*. To this end, the probability $P(\vec{n}, t)$ to observe $\vec{n}$ molecules at time t is defined, where the vector $\vec{n}$ is a sort of inventory of mRNA molecule numbers transcribed from first gene (n_1), from second gene (n_2) and so on and protein numbers translated from mRNA from first gene (n_3), etc.

The aforementioned probability is actually a conditional probability, conditioned by surely being in state n_0 at initial time: $P(\vec{n}, 0) = \delta_{\vec{n},n_0}$. [22]

To determine how the probability $P(\vec{n}, t)$ changes over a small increment of time, one has to write the in-going and out-going fluxes as the product of the probability of being in a state $\vec{n}$ times the transition probability from that state to another $W_{\vec{n}\to\vec{n'}}$, i.e.:

$$\Delta P(\vec{n}, t) = \sum_{\vec{n'}} \left[W_{\vec{n'}\to\vec{n}} P(\vec{n'}, t) - W_{\vec{n}\to\vec{n'}} P(\vec{n}, t) \right] \Delta t. \tag{3}$$

Dividing both sides by Δt and taking the limit where this increment becomes very small, the so-called *Master Equation* is obtained:

$$\frac{\partial P(\vec{n}, t)}{\partial t} = \sum_{\vec{n'}} \left[W_{\vec{n'}\to\vec{n}} P(\vec{n'}, t) - W_{\vec{n}\to\vec{n'}} P(\vec{n}, t) \right]. \tag{4}$$

The transition probabilities W are assumed here to depend only on state at time t, so that the process is memory-less and therefore *Markovian*.
Considering only the transcription event in the single gene circuit already seen in the deterministic approach, the master equation reads

$$\begin{aligned} \frac{\partial P(R, t)}{\partial t} =& k_R \left[P(R-1, t) - P(R, t) \right] + \\ &+ g_R \left[(R+1)\, P(R+1, t) - R\, P(R, t) \right], \end{aligned} \tag{5}$$

where the negative terms are the loss ones, while the positive terms are the gain ones. This can be pictured diagrammatically as follows:

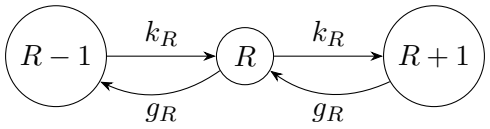


Figure 3: Transition scheme between states $R-1$, R e $R+1$.

3.2.1 Solving the Master Equation: the moment generating function

Typically, the master equation associated to a genetic circuit is hard to solve exactly, due to the complexity of the chemical reactions involved.

One way to circumvent this consists in analyzing directly the evolution of the moments of the probability distribution through the *Moment Generating Function*. Considering again the case where only mRNA is transcribed from

a single-gene, the one-dimensional moment generating function may be introduced as

$$M(z,t) = \sum_{R=0}^{\infty} z^R P(R,t), \tag{6}$$

with properties:

* $M(z=1,t) = 1$;
* $\frac{\partial M(z,t)}{\partial z}\Big|_{z=1} = \sum_{R=0}^{\infty} R\, z^{R-1} P(R,t)\Big|_{z=1} = \sum_{R=0}^{\infty} R\, P(R,t) = \langle R(t)\rangle$;
* $\frac{\partial^2 M(z,t)}{\partial z^2}\Big|_{z=1} = \langle R^2(t)\rangle - \langle R(t)\rangle$.

The master equation in this case will read:

$$\frac{\partial M(z,t)}{\partial t} = k_R(z-1)F(z,t) - g_R(z-1)\frac{\partial M(z,t)}{\partial z}. \tag{7}$$

The steady-state solution, $M^{ss}(z,t) = e^{\frac{k_R}{g_R}(z-1)}$, implies that the first moment is the same one obtained with the deterministic approach, i.e.:

$$\frac{\partial M^{ss}(z,t)}{\partial z}\Big|_{z=1} = \frac{k_R}{g_R} = \langle R\rangle^{ss}.$$

It is useful to compute the second moment, using the third property of the moment generating function, that reads

$$\frac{\partial^2 M^{ss}(z,t)}{\partial z^2}\Big|_{z=1} = \left(\frac{k_R}{g_R}\right)^2 = \langle R^2\rangle^{ss} - \langle R\rangle^{ss},$$

$$\sigma^2 = \langle R^2\rangle^{ss} - (\langle R\rangle^{ss})^2 = \langle R\rangle^{ss}.$$

For a simple system with transcription and degradation, also referred to as *birth and death process* in the literature of stochastic systems, the mean and the variance equal each other, which is the hallmark of a Poisson process with Fano Factor $\frac{\sigma^2}{\langle R\rangle} = 1$.

3.2.2 Gaussian approximation: the van Kampen approach

Only in rare cases it is possible to solve the Master Equation explicitly. When this equation cannot be solved, it is necessary to have a systematic approximation method, i.e.: the Van Kampen expansion, also known as Linear Noise Approximation. [24]

This method is based on the distinction between two scales, the one relative to the macroscopic behavior of the system and the one regarding the noise. As a matter of fact, fluctuation are caused by the discrete nature of matter: density of a gas fluctuates because the gas consists of molecules, chemical reaction fluctuation arise as they consists of individual relative collisions, etc. On the other hand, the macroscopic features are determined by all particles together, thus one expects the importance of fluctuations to be relatively small when the system is large. Therefore, the main goal of this approximation is to separate the deterministic behavior from the stochastic one.[24]

To this end, the Van Kampen expansion allows to re-write the state of the system as the sum of two terms, each of which is characterized by a different scaling with the size parameter of the system, in agreement with the previous consideration. Particularly, in the case of genetic circuits, the size parameter is the volume V of the cell and the state of the system, which is just the number of molecules n, may be re-written as

$$n = V\phi(t) + \sqrt{V}\xi, \tag{8}$$

where $\phi(t)$ is the deterministic concentration of molecules present in the system and ξ represents the noise. Notice that n, ϕ, ξ may be vectors when more than one molecular species is present in the analyzed genetic circuit, but for now are kept as scalar quantities for simplicity.

Then, also the probability satisfying the Master Equation can be written as a function of the noise variable just introduced,

$$P(n,t) = \Pi(V\phi + \sqrt{V}\xi, t) = \Pi(\xi, t). \tag{9}$$

Therefore, it is possible to systematically expand the Master Equation following this rule—all the details of the mathematical procedure carried out are presented in Part III for the specific miRNA-based system of interest.

Particularly, all terms of order $V^{-1/2}$ lead to the deterministic equations (such as equations 1) describing the evolution of the concentration of the molecular species involved. Then, all terms of order V^0 are considered and they lead to a linear Fokker-Planck equation for $\Pi(\xi, t)$, whose solution is a Gaussian distribution. Therefore, just the knowledge of first and second moment, which can be gained from the Fokker-Planck equation and the deterministic ones, is enough to specify this distribution. The *Gaussian Approximation* is the name

given to the Van Kampen expansion because, by ignoring the higher-order terms and retaining only the dominant terms (V^0), the system's stochastic behavior is approximated by a linear noise process, whose probability distribution is Gaussian.

This method allows to gain an approximation for probability of having n molecules at time t, simply using the connection between $P(n,t)$ and $\Pi(\xi,t)$ (equation 9). Of course, upon including higher orders of the expansion, correction to the gaussian behavior may be included (V^{-1}, order of a single molecule), but the procedure then gets mathematically much more complex.

The need for an analytic expression of this probability $P(n,t)$ is explained in the following, as the main focus of this study, the First-Passage Time distribution, is linked to it.

3.3 First-passage Time distribution

Normally, the analysis of the Master Equation of a given system focuses on the stationary solution, i.e., the stationary probability distribution, $P(\vec{n})$, and its moments. More complex is the study of the dynamics.
However, biologically, each cell in an organism has to perform many tasks to ensure its survival, such as cellular development or cell differentiation. All of these fundamental processes must occur correctly in time, as they serve specific purposes, and although they are driven by intrinsically stochastic biochemical reactions, they must reliably achieve their intended outcome. To accomplish this, the cell must maintain the correct time ordering of events and precise timing across all processes by means of various regulatory mechanisms acting on gene expression.

In order to study event timing, a threshold-crossing framework is typically established. In this context, when a given chemical species reaches a predetermined threshold level, it triggers the initiation of a biological process.
More precisely, when a gene is switched on, the average level of a generic chemical species approaches the steady-state according to the dynamics described by the deterministic equations associated to the system. However, since gene expression is an intrinsically stochastic process, individual trajectories fluctuate around this mean behavior, so that each of them reaches the threshold level at a different time.
The distribution of times crossing for the first time a fixed threshold is the *first-*

passage time (FPT) distribution and it represents the variability in reaching a certain level of expression. In fact, a gene may reach a target level of expression with a substantial cell-to-cell variability, even in genetically identical population of cells exposed to the same stimulus; this, as noted before, is due to the intrinsic stochasticity of gene expression.
With the aim of ensuring the correct timing for each event, the cell must somehow control this variability in the timing of reaching the target level.

3.3.1 First-passage times: mathematical tools

Mathematically, the timing of molecular events can be characterized through different probability measures.
The starting point is the already defined instantaneous probability $P(n,t)$ of observing n molecules at time t. However, This probability distribution does not reflect the history of the threshold crossings. To this end, it is convenient to introduce the so called *visit probability* $Q(n,t) = \mathbb{P}(n_\tau = n \text{ at time } t_0 \leq \tau \leq t)$, namely the probability that the molecule count has reached the given threshold up to time t [25].
The visit probability is the complement of the survival probability $S(t)$, which measures the probability that the threshold has never been reached by time t. If the threshold level is n^*, then the survival probability is defined as:

$$S(t) = \sum_{n=0}^{n^*-1} P(n,t). \tag{10}$$

The survival probability contains all the information about the process. In particular, it is directly related to the *first-passage time (FPT) distribution*, $F(t)$, through its time derivative as follows:

$$F(t) = \frac{\partial Q}{\partial t} = -\frac{\partial S}{\partial t}. \tag{11}$$

The First-Passage Time distribution, which is the aim of this study, quantifies the probability that the system crosses the threshold for the first time exactly at time t.

This formalism allows one to switch from the stochastic dynamics of molecular numbers in genetic circuits to predictions about event timing, essential for understanding the reliability of the dynamics of intra-cellular events.

4 miRNA-mediated genetic circuit

This study focuses on a particular miRNA-mediated circuit, which represents one of the simplest way to model the mRNA-miRNA interaction. This circuit involves three molecular species, namely the miRNA (S), its target mRNA (R) and the protein (P), translated from the mRNA R. In particular, for simplicity, we assume that mRNA and microRNA are transcribed from independent genes at independent rates and that all reaction rates are constant.
The circuit is represented in Figure 4.

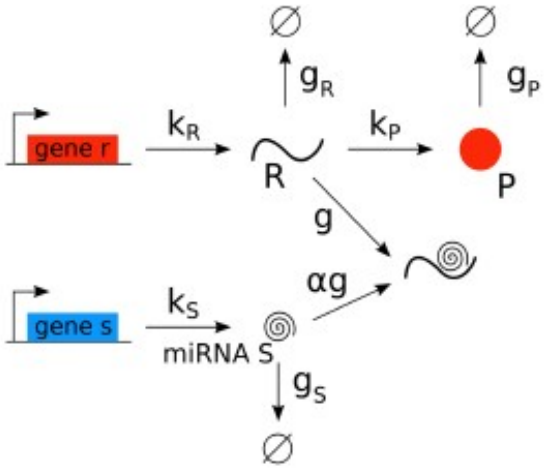


Figure 4: MiRNA-mediated genetic circuit, adapted from Del Giudice et al. [6]. Two different genes are individually transcribed into mRNA (R) and microRNA (S) molecules. They can either bind with one another or decay independently. If a mRNA molecule is not bound to a miRNA one, it can be translated into protein (P).

Therefore, the molecular reactions involving the molecular species R, S and P are

$$\emptyset \underset{g_R}{\overset{k_R}{\rightleftharpoons}} R \qquad \text{mRNA transcription and degradation,}$$

$$\emptyset \underset{g_S}{\overset{k_S}{\rightleftharpoons}} S \qquad \text{miRNA transcription and degradation,}$$

$$R + S \xrightarrow{g} RS \qquad \text{mRNA-miRNA interaction,}$$

$$RS \xrightarrow{1-\alpha} S \qquad \text{recycled miRNA,}$$

$$R \xrightarrow{k_P} R + P,\ P \xrightarrow{g_P} \emptyset \qquad \text{protein production and degradation.}$$

In this framework, the miRNA and mRNA are transcribed and degraded through independent processes with transcription rates k_S and k_R and degradation rates g_S and g_R respectively. To model the post-transcriptional role played by miRNA molecules, we consider a further reaction such that it can bind to target mRNA with rate g. This effective parameter amounts to the strength of the coupling, which biologically depends on the affinity of the two molecular species in terms of base-pairing. From a complex, we assume that mRNAs are always degraded and that a fraction α of miRNAs cannot be recycled from the RS complex. If a mRNA molecule is free, it can then be translated into protein with translation rate k_P. In turn, proteins can be degraded with rate g_P.

As noted in the previous Chapter, these molecular reactions can be rewritten into a set of differential equations, one for each molecular species involved, that govern the dynamics of the considered system, i.e.:

$$\begin{aligned} \frac{dR}{dt} &= k_R - g_R R - gRS, \\ \frac{dS}{dt} &= k_S - g_S S - g\alpha RS, \\ \frac{dP}{dt} &= k_P R - g_P P, \end{aligned} \tag{12}$$

where R, S and P are the concentrations, i.e., number of molecules over volume ratio, of the corresponding species.

Due to the intrinsic stochasticity of molecular reactions, a stochastic approach is more appropriate to fully describe gene expression.

To this end, the probability $P(\vec{n}, t)$ of observing $\vec{n} = (n_R, n_S, n_P)$ molecules at time t is defined, where the number of molecules of species X is related to the concentration ρ_X as $n_X = V_{cell}\rho_X$, where V_{cell} is the cell volume. Following the same scheme previously adopted, the dynamics can be described by the following Master Equation,

$$
\begin{aligned}
\frac{dP(\vec{n},t)}{dt} &= k_R[P(n_R-1,t)-P(n_R,t)]+\frac{g_R}{V_{cell}}[(n_R+1)P(n_R+1,t)+\\
&-n_RP(n_R,t)]+k_S[P(n_S-1,t)-P(n_S,t)]+\\
&+\frac{g_S}{V_{cell}}[(n_S+1)P(n_S+1,t)-n_SP(n_S,t)]+\\
&+\frac{k_Pn_R}{V_{cell}}[P(n_P-1,t)-P(n_P,t)]+\\
&+\frac{g_P}{V_{cell}}[(n_P+1)P(n_P+1,t)-n_PP(n_P,t)]+\\
&\frac{g\alpha}{V_{cell}^2}[(n_S+1)(n_R+1)P(n_R+1,n_S+1,t)-n_Sn_RP(n_R,n_S,t)]+\\
&+\frac{g(1-\alpha)n_S}{V_{cell}^2}[(n_R+1)P(n_R+1,t)+\\
&-n_RP(n_R,t)],
\end{aligned}
\tag{13}
$$

where $P(n_R+1,t)$ is the simplified notation for $P(n_R+1,n_S,n_P,t)$.

4.1 State-of-the-Art in miRNA-meadiated circuits: titration, noise buffering, and bimodal protein distributions

The analysis of molecular circuits like the one described has been crucial to understand how miRNAs regulate gene expression at the single-cell level.

A cornerstone in the analysis of regulatory network topologies is the miRNA-mediated incoherent feed-forward loop, identified by Osella et al. (2011) [18] as a statistically overrepresented motif. This system is shown to achieve a fine-tuned protein output that remains stable, i.e., maintains low noise, even in the presence of significant upstream regulatory fluctuations.

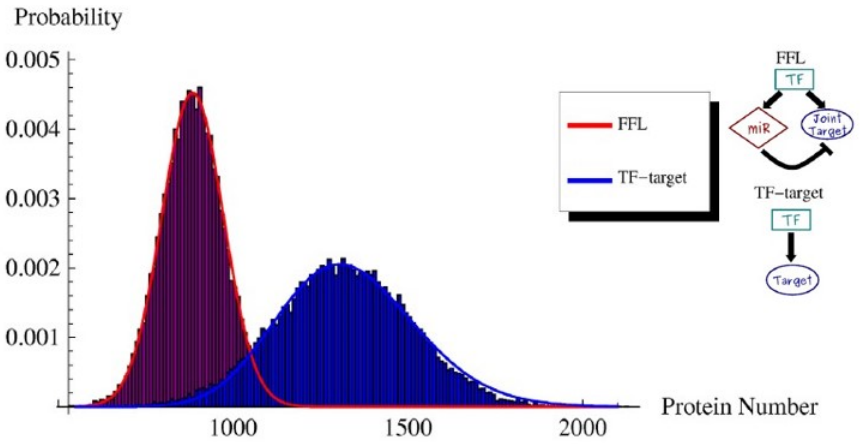
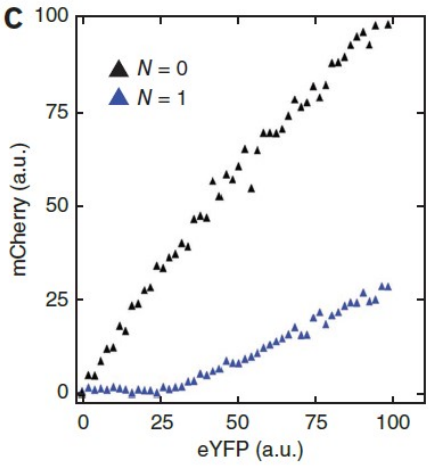


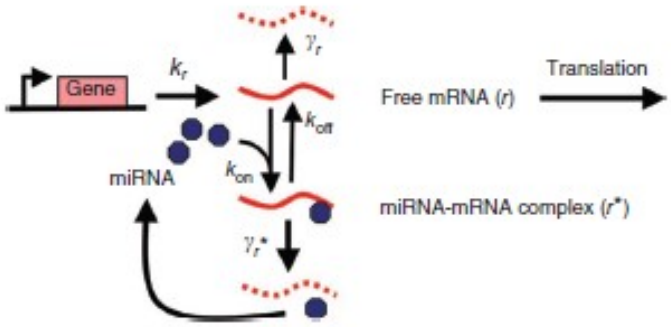
Figure 5: Steady-state distribution of target protein numbers, adapted from Osella et al. (2011) [18]. The comparison highlights the difference between simple regulation (blue) and the incoherent Feed-Forward Loop (red). The narrower shape of the red curve shows how the FFL structure reduces stochastic fluctuations (noise buffering), stabilizing the protein output.

Therefore, by buffering fluctuations, miRNA post-transcriptional regulation ensures stability and robustness in protein level (Figure 5), as confirmed also by the work of Siciliano et al. (2013) [23] . Both works by Osella and Siciliano show that miRNA acts as a fine-tuner of gene expression.

Other key insights came from experimental work, such as the one carried out by Mukherji et al. (2011) [16]. In this work, the authors were able to reconstruct the miRNA-mediated circuit *in vitro*. To this aim, they used a two-color fluorescent reporter system to simultaneously monitor a given target gene expression with and without miRNA regulation in individual cells. This experiment allowed them to observe that protein production is highly repressed below a specific threshold level of target mRNA but responds sensitively and increases rapidly once that threshold is crossed (Figure 6a).



(a) Experimental transfer function showing the threshold effect. Particularly, because mCherry and eYFP are two fluorescent protein that share a bidirectional promoter, eYFP fluorescence (x-axis) serves as a proxy for the transcriptional input, while mCherry (y-axis) represents the target protein output subject to miRNA regulation.



(b) Biochemical model of miRNA-mediated molecular titration.

Figure 6: Threshold mechanism in miRNA regulation, adapted from Mukherji et al. (2011)[16].
(a) Unregulated gene expression (black triangles) is compared with miRNA-regulated expression (blue triangles). The "hockey stick" shape of the blue curve indicates a molecular titration mechanism: the pool of available miRNAs sequesters and represses target mRNAs until the threshold is crossed, after which protein production resumes linearly.
(b) Schematic of the molecular titration model explaining this phenomenon: free miRNAs bind and sequester target mRNAs (r), preventing translation until the miRNA pool is saturated.

This behavior is consistent with a mathematical model of *molecular titration* (Figure 6b), through which miRNAs act as a threshold-dependent switch. In fact, in the case of low mRNA molecules (below the threshold), miRNAs

can titrate almost all target mRNAs, effectively acting as an on/off switch that is on the off state. On the contrary, when mRNA molecules cross the threshold, they saturate the pool of available miRNAs. This means that the system becomes very sensitive to the number of free mRNA molecules, allowing for a linear response in protein production rather than just turning it on or off.

This experimentally verified threshold mechanism is highly sensitive to the stochastic fluctuations inherent in gene expression. Gene expression noise is broadly classified into two types, as experimentally demonstrated by Elowitz et al. (2002), i.e.:

- *intrinsic noise* which is the stochasticity inherent to the chemical reactions themselves and to the fact that molecules are discrete entities;
- *extrinsic noise* which is given by fluctuations in cellular conditions, such as temperature, cell size, etc.

Intrinsic noise alone can induce bimodality [6], as the system can jump from a low-mRNA molecule state to a high-mRNA molecule state. This phenomenon requires a very steep threshold, which is only possible for a high interaction strength between miRNA and target mRNA molecules [16].

Expanding on this, Bosia et al. (2017) [3] demonstrated that this bimodal behavior is strongly influenced by the competition for shared miRNAs. This competition generates regulatory cross-talk, a mechanism where distinct RNA targets indirectly influence each other's expression; in fact, by sequestering a portion of the finite miRNA pool, an abundance of one target relieves the repression on others. This showed that increasing the number of binding sites enhances this cross-talk, thereby amplifying stochastic fluctuations and forcing the system to segregate into two distinct subpopulations (bimodality) characterized by high and low expression states.

It was also studied by Del Giudice et al. (2018) how extrinsic noise shapes bimodal gene distributions in the context of microRNA-mediated regulation—precisely the genetic circuit reported in Figure 4. They modeled the production rate of miRNA molecules k_S not as a constant, but as a fluctuating parameter drawn from a Gaussian distribution. In this context, it is found that a noisy environment can compensate for low miRNA-target interaction to obtain a bimodal distribution, therefore relaxing the requirements of the intrinsic noise

mechanism (Figure 7).

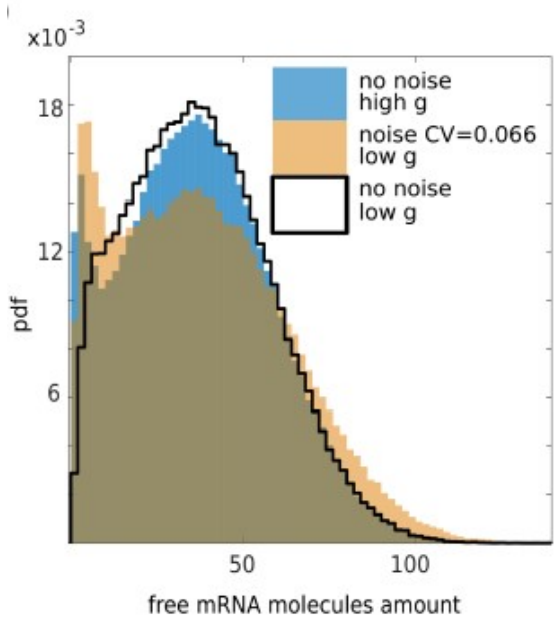


Figure 7: The graph illustrates how extrinsic noise influences the probability density function of free mRNA molecule counts, adapted from Del Giudice et al. (2018) [6].
The black line represents the scenario with low miRNA-target interaction strength (g) and only intrinsic noise, which results in a unimodal distribution. In contrast, the blue histogram shows that a high interaction strength (g)—still only intrinsic noise is included—generates a bimodal distribution. Then, the orange histogram demonstrates that introducing extrinsic noise to the low interaction strength scenario is sufficient to induce a bimodal distribution, indicating that extrinsic noise can compensate for weak molecular affinity.

All these previous studies, however, focus on the steady-state behavior of the system. A foundational study regarding the dynamics of genetic systems is the analysis carried out by Dal Co et al. (2017) [5]. This research focused on the stochastic timing of protein accumulation, investigating how intrinsic noise affects the time required for a gene product to reach a specific threshold.

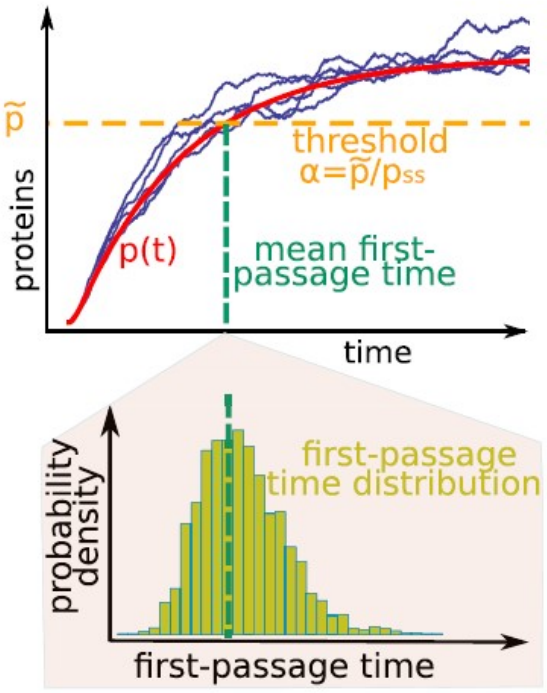


Figure 8: Protein First-Passage Time framework, adapted from Dal Co et al. (2017) [5].

Specifically, they were able to derive a geometric analytical approximation to estimate the noise of the First-Passage Time distribution (Figure 8). They showed that transcriptional auto-regulation is a key determinant of timing precision, with negative auto-regulation minimizing noise for rapid responses and positive auto-regulation optimizing timing for longer delays.

This laid the groundwork for understanding how transcriptional architecture—specifically simple gene activation and feedback loops—influences timing noise by analyzing properties of FPT distribution of proteins.

In contrast, this thesis shifts the focus to dynamical properties of miRNA-mediated regulation by investigating the First-Passage Time distributions of target mRNAs. Therefore, this thesis advances beyond moment estimation (mean and variance) by aiming to compute the full First-Passage Time distribution of mRNA molecules, both in the absence and presence of miRNA-mediated post-transcriptional regulation.

Also, being mRNA molecules the direct precursors of proteins and the direct targets of miRNAs, studying mRNA FPT distributions rather than protein ones

provides a higher resolution analysis on the timing dynamics under study.

Part III

First-Passage Time study of a miRNA-mediated circuit

The aim of this Part is to analyze the stochastic dynamics of the miRNA-mediated circuit (Figure 4) previously introduced, specifically focusing on the First-Passage Time (FPT) distribution for the system's mRNA molecules. This FPT analysis provides a powerful quantitative framework for understanding gene expression timing and variability.
To investigate this, both a numerical and an analytical approach will be carried out.

First, the numerical method is implemented: the Gillespie algorithm. This algorithm simulates the exact stochastic time evolution of the system. Therefore, by generating a large ensemble of individual trajectories, the FPT distribution can be numerically constructed.
Second, the analytical approach is developed, based on the Van Kampen expansion, also known as the Linear Noise Approximation (LNA). This LNA framework is particularly useful as it allows us to derive an analytical prediction of the FPT distribution of interest.

This provides deep insights into the system's timing, which can then be compared directly with the exact numerical results from the Gillespie simulations to test the validity and accuracy of the approximation. Particularly, the results for the unregulated circuit are presented, comparing the Gillespie simulations with the analytical calculation of the FPT distribution. Then, results regarding the miRNA-mediated circuit are shown, with particular focus on if and when the approximation implemented is valid.

5 Numerics: the Gillespie algorithm

In order to analyze genetic circuits or, more in general, stochastic chemical reactions, many approximation methods can be carried out.
A fundamental method that performs numerical stochastic simulations is going to be studied in the following: the Gillespie algorithm [9].
This algorithm provides a rigorous and probabilistic framework for simulating chemical reactions [10]. It does not attempt to solve the Master Equation associated to the system of interest (Figure 4). Instead, it provides an output trajectory that exactly conforms to the solution of the chemical Master Equation.

5.1 The algorithm

The Gillespie algorithm, or Stochastic Simulation Algorithm (SSA), is an exact simulation method, as it calculates the exact time and nature of the next discrete event, for us a single chemical reaction, that will occur [11].
The core logic of the algorithm, which allows this exact simulation, relies on treating the system as a probabilistic "race" between all competing reactions to determine which event will happen next.

First, the definition of the initial state—namely the discrete populations of all species X_j, with $j \in [1, N]$ index of the N species present in the system—and the set of all possible reactions M is needed.

The algorithm then relies on the propensity a_i, which represents the probability rate for each reaction M based on the current state. For example, for a reaction like $A + B \to C$, the propensity a_1 is calculated as

$$a_1 = X_A X_B k_1,$$

where X_A and X_B are the reactant populations and k_1 is the stochastic rate constant.
By computing then the total propensity

$$a_0 = \sum_{i=1}^{M} a_i,$$

the algorithm then determines the time of the next event. In fact, it generates a putative time τ by drawing from an exponential distribution based on the

total propensity, i.e.:

$$\tau = (-\frac{1}{a_0}) \ln\left(\frac{1}{r_1}\right),$$

where r_1 is a uniformly distributed random number between 0 and 1. If the system is highly reactive (high a_0), τ will be very small. If the system is in a state with few reactants (low a_0), τ will be large.

After the time step τ has been calculated, the algorithm must determine which specific reaction μ, from all M possible ones, will actually modify the current system state.

The probability that the i-th reaction is the next to occur is given by its relative share of the total propensity, i.e.:

$$Prob(\mu = i) = \frac{a_i}{a_0}.$$

The higher the propensity a_i, the higher the probability of reaction i being the next one. To implement this probabilistic choice, the algorithm employs an efficient linear search to find the reaction μ. Particularly, it generates a random target value by scaling a new uniformly random number r_2 by the total propensity: $r_2 a_0$. This value represents a random point along the interval $[0, a_0]$. The algorithm then searches for the reaction index $\mu \in [1, M]$ that corresponds to this point. Formally, it finds the smallest integer μ that satisfies the condition:

$$\sum_{i=1}^{\mu-1} a_i < \ r_2 a_0 \ \leq \sum_{i=1}^{\mu} a_i.$$

The first reaction μ whose propensity a_μ causes the partial sum to equal or exceed the $r_2 a_0$ target is selected as the event that will occur at time $t + \tau$. This process guarantees that reactions with higher propensities are proportionally more likely to be selected.

Finally, the system's populations are updated according to the stoichiometry of reaction μ, and the global time is advanced by τ, before the entire cycle repeats from the new state.

The algorithm iterates this procedure; in each step it re-computes the propensities and generates a new pair (τ', μ') to update the state of the system and the time of the evolution. It will stop when this time exceeds the given end time t_{end}.

The result is an exact stochastically generated trajectory of the state of the system, as will be shown in the following analysis.

5.2 Single gene simulations

The genetic circuit that will be studied in the following is the same miRNA-based one described in Chapter 4. In order to verify that the Gillespie algorithm implemented works as expected, at first the mRNA-miRNA interaction will be neglected, meaning that the parameter g encoding the strength of such coupling will be set to zero. As previously noted, this means that the circuit simplifies to a single-gene one, where there is just transcription and degradation of a given chemical species.

Furthermore, as discussed in Part II, for a birth and death process a Poisson process is recovered. This may be also numerically verified for each gene considered.

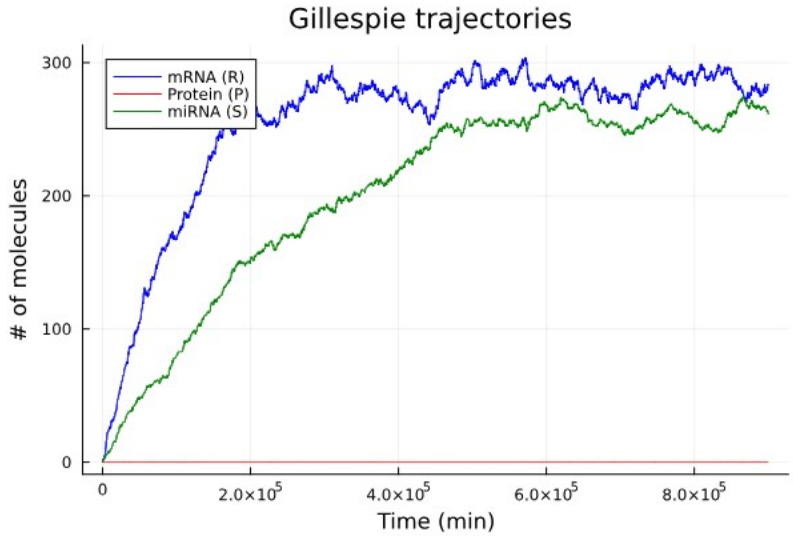
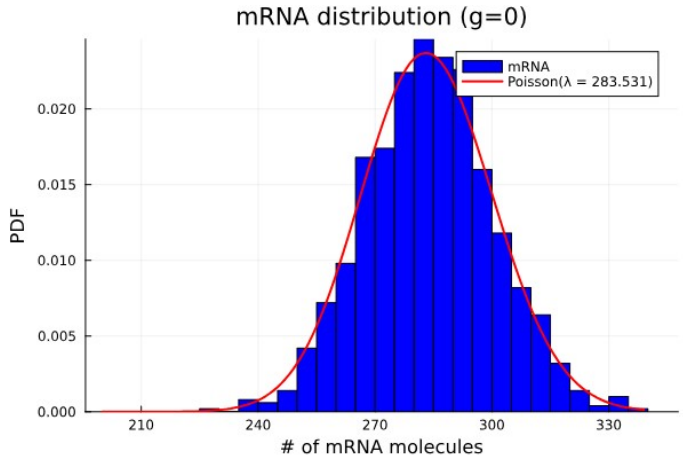
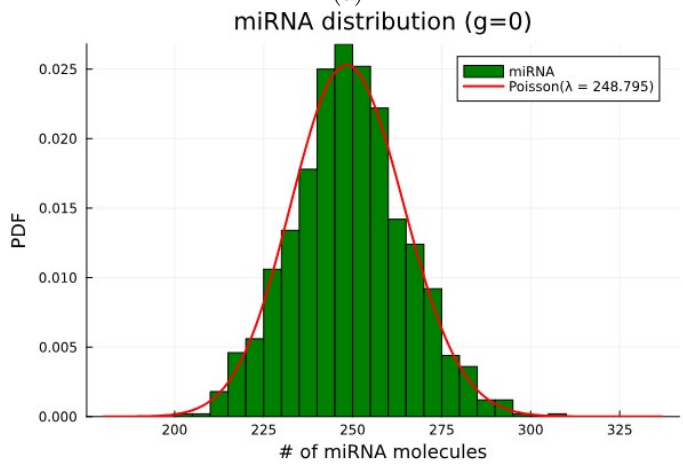


Figure 9: Gillespie simulation for mRNA (blue) and miRNA (green) trajectories for $g = 0\ nM^{-1}min^{-1}$. As the interaction between the two species has been neglected, they evolve independently. Protein production and decay is set to be zero (red line) and all other parameters assume values reported in Appendix, Table 2.



(a)



(b)

Figure 10: Normalized histograms of mRNA (Figure a) and miRNA (Figure b) number of molecules, overlaid with theoretical Poisson model (red line). Parameters set as explained in Figure 9.

5.3 mRNA-microRNA simulations

Upon including the fact that mRNA molecules and microRNA ones may form the mRNA-miRNA complex, the trajectories simulated through the Gillespie algorithm change. This is due to the fact that miRNA molecules, when bound to their targets, reduce drastically the available number of mRNA molecules. Just for completeness, in the first simulation also the number of proteins translated from available mRNA molecules is reported, even though in this study focuses on mRNAs distributions.

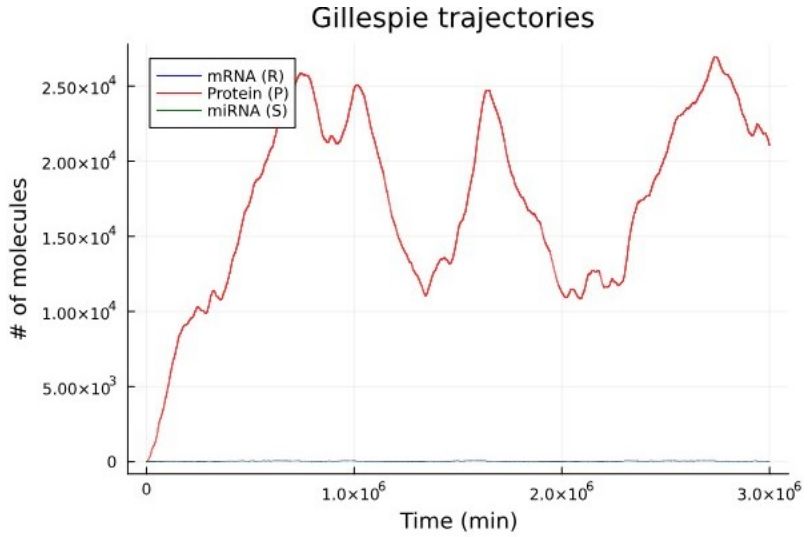


Figure 11: Gillespie simulation of the full miRNA-based genetic circuit. The red line is the protein trajectory. Parameters values reported in Appendix, Table 1.

In order to better understand how parameter g, assuming non zero value, influences the number of mRNA molecules in the following the proteins counts have been neglected.

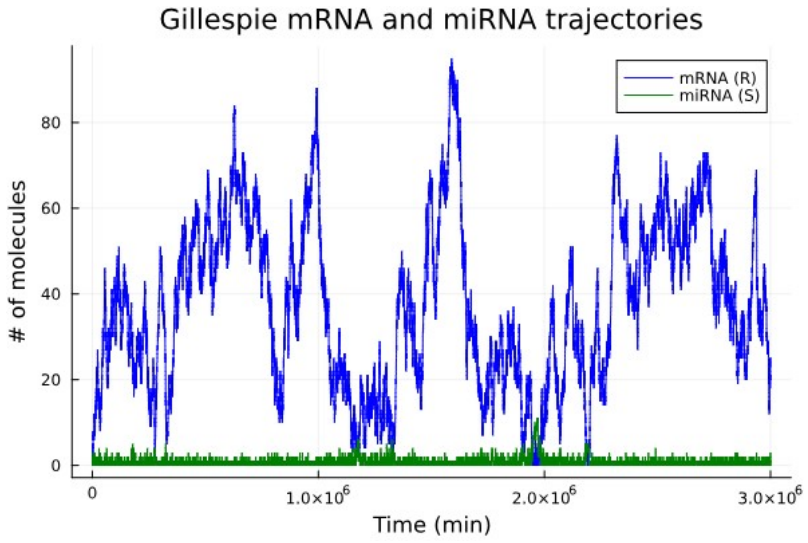


Figure 12: Focus of just mRNA and miRNA trajectories not visible from Figure 11. Parameters values reported in Appendix, Table 1.

As expected, the interaction between the two molecular species is modifying the availability of molecules inside the cell, therefore we are correctly modeling

the miRNA-mediated post-transcriptional control.

5.4 First-Passage Time distribution of mRNAs

As the main aim of this study is to investigate the dynamics of the genetic circuit shown in Figure 4, a First-Passage Time framework is developed within the context of a Gillespie simulation. Particularly, in such a simulation, the output of each run corresponds to the time evolution of the number of molecules for each chemical species in the system. To define the First-Passage Time, a threshold is introduced in terms of the number of mRNA molecules. The first passage time for a given trajectory is then the time at which the number of molecules first reaches this threshold level.

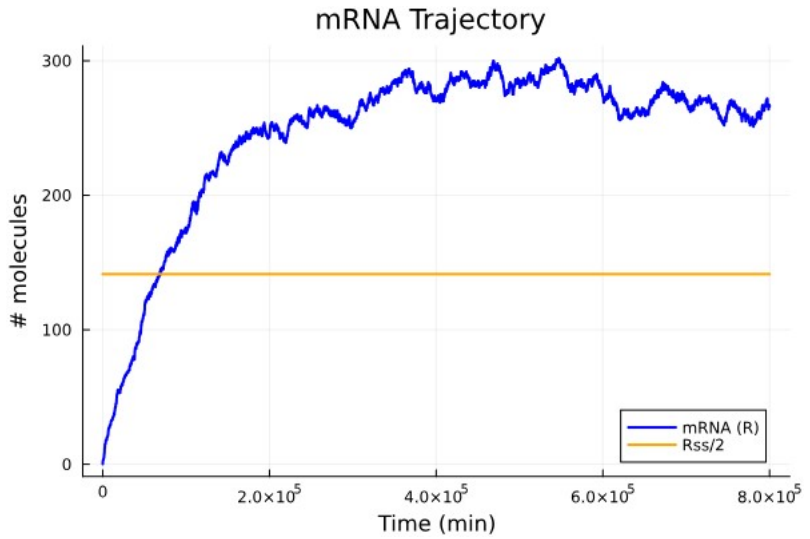


Figure 13: Number of mRNA molecules as a function of time for $g = 0\ nM^{-1}min^{-1}$. The yellow line, representing the chosen threshold level, is the half steady-state value. All other parameters are set to the values reported in the Appendix, Table 2.

Because of the intrinsic stochasticity of gene expression, or better of the underlying reactions, different stochastic realizations will produce trajectories that reach the threshold at different times. Therefore, in order to obtain statistically meaningful results, a large number of independent Gillespie simulations must be performed. The collection of these individual crossing times constitutes the first passage time distribution, which provides quantitative insight into the variability and timing of gene expression.

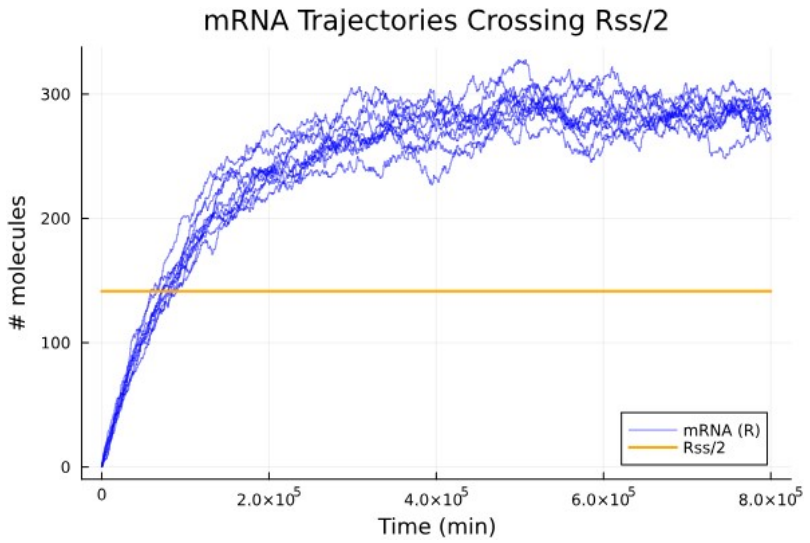


Figure 14: Multiple trajectories for the number of mRNA molecules as a function of time crossing the half steady-state line (in yellow) for $g = 0\ nM^{-1}min^{-1}$. All other parameters are set to the values reported in Appendix, Table 2.

Number of simulations In order to evaluate the reliability of the statistical results and identify a suitable number of stochastic realizations, the mean of the crossing times for a fixed threshold was calculated as a function of the total number of simulations. In this analysis, the chosen threshold corresponds to half of the steady-state value for the considered chemical species:

$$R_{ss} = \frac{k_R V}{g_R} \quad \text{for binding affinity} \quad g = 0.$$

To this end, sets of 100, 200, 300, and up to 1000 simulations were performed, as reported below.

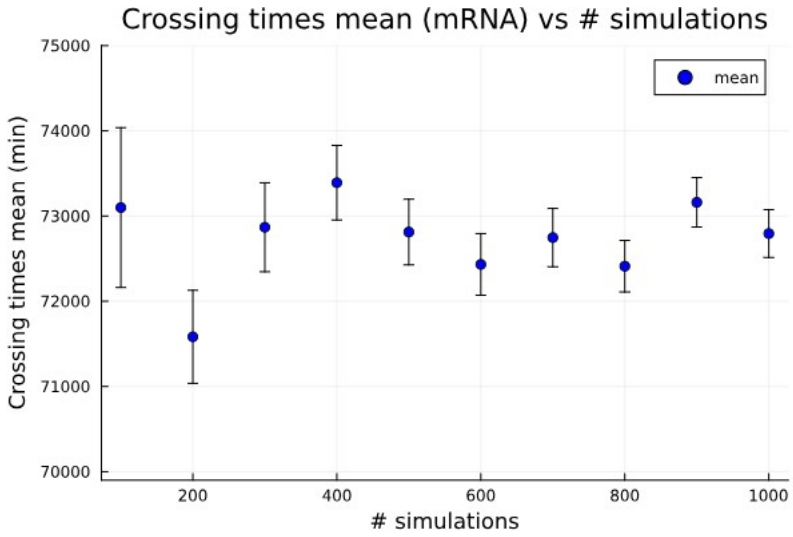


Figure 15: Mean of mRNA crossing times of half of the steady-state as a function of the number of trajectories for $g = 0\ nM^{-1}min^{-1}$. All other parameters are set to the values reported in Appendix, Table 2.

This result shows that the mean value of the crossing times gradually converges after 500 simulations. Based on this observation, 500 realizations were chosen as a reasonable compromise between statistical accuracy and computational cost, and this number was therefore adopted for all subsequent Gillespie simulations.

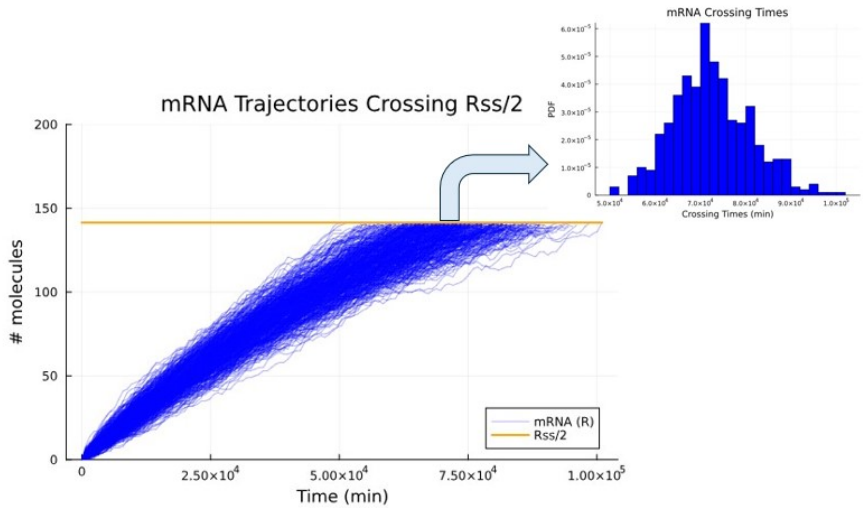


Figure 16: Visualization of all 500 trajectories obtained through Gillespie simulations and FPT distribution of all times crossing the half-steady-state threshold (yellow line) neglecting miRNA-mRNA interaction ($g = 0\ nM^{-1}min^{-1}$). All other parameters are set to the values reported in Appendix, Table 2.

This First-Passage Time framework is applied to an ensemble of trajectories for non-zero g values, enabling the study of the FPT distribution for different interaction strengths. As shown in Figure 10, a Poisson process is recovered for zero binding affinity ($g = 0$). Upon increasing the value of parameter g, the mRNAs distribution transitions to a bimodal one. Intuitively, as more miRNAs bind to their targets, mRNA molecules disappear from the system. The system's stochastic nature—i.e., noise—allows it to settle into two distinct states: one with a high number of free mRNA molecules and one with a lower number depending on miRNA availability.

6 Analytics

6.1 Non interacting case: $g = 0$

Since we are interested in the First-Passage Time distribution of mRNA molecules, from now on the proteins are neglected, as they do not influence the mRNA dynamics.
By setting parameter g to zero, so not considering the mRNA-miRNA interaction, the genetic circuit already numerically analyzed (Figure 4) becomes very simple, as it is just composed by two independent genes that transcribe one for mRNA and one for miRNA. The master equation that describes such a system is the single gene master equation:

$$\frac{dP(n,t)}{dt} = k[P(n-1,t) - P(n,t)] + \frac{g}{V}[(n+1)P(n+1,t) - nP(n,t)], \quad (14)$$

where n can be the number of either mRNA molecules or microRNA ones and k (or g) is the transcription (degradation) rate of the chemical species considered.

With the aim of analytically obtain the First-Passage Time distribution for mRNA molecules, two different approaches were implemented to solve this master equation, which are reported below.

6.1.1 Redner approach

Consider Master Equation (14), which is describing birth and death of mRNA molecules, with following initial condition:

$$P(n,0) = \delta_{n,0}. \quad (15)$$

The idea is to apply the method used by Redner S. [20], [21], so transforming the number of molecules and time following the equation

$$c(k,s) = \int_{-\infty}^{\infty} dx\, e^{ikx} \int_{0}^{\infty} dt\, e^{-st} c(x,t). \quad (16)$$

Notice that n is the number of molecules, so it is discrete and cannot be negative, therefore the transform must be modified accordingly,

$$\sum_{n=0}^{\infty} e^{ikn} \int_0^{\infty} dt\, e^{-st}\, \frac{\partial P(n,t)}{\partial t} = \sum_{n=0}^{\infty} e^{ikn} \int_0^{\infty} dt\, e^{-st} \Big[k\left(P(n-1,t) - P(n,t)\right) + \frac{g}{V}\left((n+1)P(n+1,t) - nP(n,t)\right)\Big]. \tag{17}$$

Focus on the l.h.s. of (17),

$$\begin{aligned}
\int_0^{\infty} dt\, e^{-st} \frac{\partial \left\{\sum_{n=0}^{\infty} e^{ikn} P(n,t)\right\}}{\partial t} &= \int_0^{\infty} dt\, e^{-st} \frac{\partial P(k,t)}{\partial t} \\
&= e^{-st} P(k,t)\Big|_0^{\infty} - \int_0^{\infty} dt\,(-s) e^{-st} P(k,t) \\
&= -P(k,t=0) + s\int_0^{\infty} dt\, e^{-st} P(k,t) \\
&= sP(k,s) - P(k,t=0).
\end{aligned}$$

While the r.h.s. of equation (17) reads

$$\begin{aligned}
&\int_0^{\infty} dt\, e^{-st} \Big\{ k\left[\sum_{n=0}^{\infty} e^{ikn} P(n-1,t) - \sum_{n=0}^{\infty} e^{ikn} P(n,t)\right] + \\
&\qquad + \frac{g}{V}\left[\sum_{n=0}^{\infty} e^{ikn}(n+1)P(n+1,t) - \sum_{n=0}^{\infty} e^{ikn} nP(n,t)\right]\Big\} \\
&= \int_0^{\infty} dt\, e^{-st} \Big\{ k\left[\sum_{n=0}^{\infty} e^{ik} e^{ik(n-1)} P(n-1,t) - P(k,t)\right] + \\
&\qquad + \frac{g}{V}\left[\sum_{n=0}^{\infty} z^n (n+1)P(n+1,t) - \sum_{n=0}^{\infty} z^n nP(n,t)\right]\Big\} \\
&= \int_0^{\infty} dt\, e^{-st} \Big\{ k\left[zP(k,t) - P(k,t)\right] + \frac{g}{V}\Big[\sum_{n=0}^{\infty} \frac{\partial z^{n+1}}{\partial z} P(n+1,t) + \\
&\qquad - \sum_{n=0}^{\infty} z \frac{\partial z^n}{\partial z} P(n,t)\Big]\Big\} \\
&= \int_0^{\infty} dt\, e^{-st} \left\{ k(z-1)P(k,t) + \frac{g}{V}(1-z)\frac{\partial P(k,t)}{\partial z}\right\}
\end{aligned}$$

$$
\begin{aligned}
&= \int_0^\infty dt\, e^{-st} \left[k(z-1) + \frac{g}{V}(1-z)\frac{d}{dz} \right] P(z,t) \\
&= \left[k(z-1) + \frac{g}{V}(1-z)\frac{d}{dz} \right] \int_0^\infty dt\, e^{-st} P(z,t) \\
&= \left[k(z-1) + \frac{g}{V}(1-z)\frac{d}{dz} \right] P(z,s).
\end{aligned}
$$

Notice that in the above calculations $z := e^{ik}$. The final equation reads (substituting k with z inside each probability)

$$
sP(z,s) - P(z,t=0) = \left[k(z-1) + \frac{g}{V}(1-z)\frac{d}{dz} \right] P(z,s), \tag{18}
$$

with

$$
\begin{aligned}
P(z,t=0) &= \sum_{n=0}^{\infty} e^{ikn} P(n,0) \\
&= \sum_{n=0}^{\infty} e^{ikn} \delta_{n,0} \\
&= 1.
\end{aligned} \tag{19}
$$

In the end, the original Master Equation is transformed into

$$
sP(z,s) - 1 = \left[k(z-1) + \frac{g}{V}(1-z)\frac{d}{dz} \right] P(z,s), \tag{20}
$$

with initial condition

$$
P(1,s) = \int_0^\infty dt\, e^{-st} \left[\sum_{n=0}^{\infty} P(n,t) \right] = \frac{1}{s}. \tag{21}
$$

This approach was not actually useful, as transforming back is in general non trivial, therefore obtaining the distribution of interest becomes very hard.

6.1.2 Method of characteristics

In order to find the probability $P(n,t)$, the Master Equation (14) can be transformed just on the number of molecules—neglecting from Redner approach the Laplace transform—becoming

$$
\frac{\partial M(z,t)}{\partial t} = k(z-1)M(z,t) + \frac{g}{V}(1-z)\frac{\partial M(z,t)}{\partial z}. \tag{22}
$$

Following the same calculations reported in (19), one finds the initial condition, $M(z,0) = 1$.

This equation is analytically solvable using the characteristics' method, as reported below:

- $$\frac{dt}{ds} = 1,$$
$$t = s;$$

- $$\frac{dz}{ds} = \frac{dz}{dt} = -\frac{g}{V}(z-1),$$
$$z(t) = z_0 e^{\frac{g}{V}t} + 1,$$
$$z_0 = (z-1)e^{-\frac{g}{V}t};$$

- $$\frac{dM(z(t),t)}{dt} = k(z-1)M(z(t),t),$$
$$M(z,t) = M(z(0),0)exp\left[\frac{k}{g}Vz_0(e^{\frac{g}{V}t}-1)\right],$$
$$M(z,t) = exp\left[\frac{k}{g}V(z-1)(1-e^{-\frac{g}{V}t})\right].$$
One may rewrite the last equation:

$$M(z,t) = exp\left[-\frac{k}{g}V(1-e^{-\frac{g}{V}t})\right]\ exp\left[\frac{k}{g}Vz(1-e^{-\frac{g}{V}t})\right] =$$
$$= exp\left[-\frac{k}{g}V(1-e^{-\frac{g}{V}t})\right]\sum_{n=0}^{\infty} z^n \frac{1}{n!}(\frac{k}{g}V)^n(1-e^{-\frac{g}{V}t})^n.$$

From the solution one may extract the expression of $P(n,t)$,

$$P(n,t) = exp\left[-\frac{k}{g}V(1-e^{-\frac{g}{V}t})\right]\frac{1}{n!}(\frac{k}{g}V)^n(1-e^{-\frac{g}{V}t})^n. \tag{23}$$

Following the procedure outlined previously, it is possible to compute the First-Passage Time distribution, shown in Figure 17 from the probability $P(n,t)$ just gained.

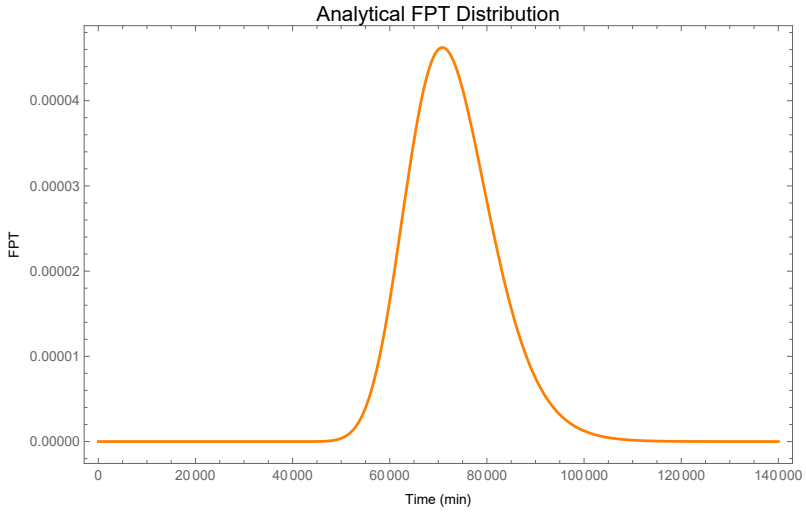


Figure 17: Analytical FPT-distribution for $g = 0$ $nM^{-1}min^{-1}$. All other parameters are set to the values reported in Appendix, Table 2.

6.2 Interacting case: $g \neq 0$

In order to understand how miRNA molecules actually influence translation of mRNA, their interaction can no longer be neglected. In a more quantitative way, this implies that parameter g appearing in the miRNA-based circuit considered must assume a non-zero value.

As in the previous case, where $g = 0$, the main objective is to obtain an analytical solution of the Master Equation, which will then allow the derivation of the First-Passage Time distribution of mRNAs under miRNA regulation.

The dynamics of such a system are still described by the Master Equation (13), which cannot be solved exactly and therefore requires suitable approximations.

6.2.1 The Van Kampen expansion

As previously seen, the fact that the number of reactants is typically low in the cellular environment leads to a stochastic description of the dynamics of the system considered. Upon increasing the number of molecules, the system's evolution becomes smoother and the deterministic description becomes more appropriate. In support of this, consider the single-gene Master Equation (5). This, as already computed, leads to a Fano factor equal to one or to a fractional deviation η, which is a dimensionless measure of the fluctuations, equal to

$$\eta = \sqrt{\frac{\sigma^2}{\langle R \rangle^2}} = \sqrt{\frac{1}{\langle R \rangle}}.$$

This fractional deviation states that relative fluctuations scale roughly as the inverse square-root of the number of reactants, therefore confirming that if reactants were to increase the system's evolution would be smoother.

Since the Master Equation (13) is not exactly solvable, a systematic approximation is needed. In this study, the Van Kampen expansion, also known as Linear Noise Approximation, will be employed [24].
This approximation exploits the behavior outlined above, as it lies on the hypothesis that the deterministic evolution of the reactant concentrations can be separated from the fluctuations. These fluctuations are assumed to scale as the square root of the typical number of molecules consistent with the Poissonian statistics discussed earlier.

In practice, this approximation consists of an expansion of the Master Equation in a power series of an extensive parameter V, the system size, through which the molecules' number of a generic chemical species X may be written as

$$n_X = V\phi_X(t) + \sqrt{V}\xi_X, \tag{24}$$

where:

- n_X: number of molecules of species X;

- $\phi_X(t)$: deterministic concentration of molecules of species X;

- ξ_X: gaussian-distributed noise variable.

Neglecting again translation and degradation of proteins, both component of $\vec{n} = (n_R, n_S)$ must be re-written following equation (24) , so that the probability $P(\vec{n}, t)$ becomes a function of ξ and t only, i.e.:

$$P(\vec{n}, t) = P(V\vec{\phi} + V^{1/2}\vec{\xi}, t) = \Pi(\vec{\xi}, t). \tag{25}$$

With the purpose of re-writing the Master Equation (13) using (24), it is useful to introduce the creation and annihilation operators, namely

$$\begin{aligned} EP(n,t) &= P(n+1,t), \\ E^{-1}P(n,t) &= P(n-1,t). \end{aligned} \tag{26}$$

Therefore, the Master Equation (13) takes the following form

$$
\begin{aligned}
\frac{\partial P(\vec{n},t)}{\partial t} &= k_R(E_R^{-1}-1)P(\vec{n},t) + \frac{g_R}{V}\left[(E_R[n_R P(\vec{n},t)] - n_R P(\vec{n},t)\right] + \\
&+ k_S(E_S^{-1}-1)P(\vec{n},t) + \frac{g_S}{V}\left[(E_S[n_S P(\vec{n},t)] - n_S P(\vec{n},t)\right] + \\
&+ \frac{g\alpha}{V^2}[E_S E_R[n_S n_R P(\vec{n},t)] - n_R n_S P(\vec{n},t)] + \\
&+ \frac{g(1-\alpha)n_S}{V^2}[E_R[n_R P(\vec{n},t)] - n_R P(\vec{n},t)] = \\
&= k_R(E_R^{-1}-1)P(\vec{n},t) + \frac{g_R}{V}(E_R-1)n_R P(\vec{n},t) + \\
&+ k_S(E_S^{-1}-1)P(\vec{n},t) + \frac{g_S}{V}(E_S-1)n_S P(\vec{n},t) + \\
&+ \frac{g\alpha}{V^2}(E_S E_R - 1)n_R n_S P(\vec{n},t) + \\
&+ \frac{g(1-\alpha)n_S}{V^2}(E_R-1)n_R P(\vec{n},t).
\end{aligned}
\tag{27}
$$

Since the noise is assumed to be small, the operator E can be expanded around $\xi = 0$, becoming

$$
E_i = 1 + V^{-1/2}\frac{\partial}{\partial \xi_i} + \frac{1}{2}V^{-1}\frac{\partial^2}{\partial \xi_i^2} + \ldots + \mathcal{O}(V^{-3/2}),
\tag{28}
$$

with i representing either mRNA (index R) or microRNA (index S).
Then, using the expansion of the operator (28) and the function $\Pi(\vec{x_i},t)$ instead of $P(\vec{n},t)$, let's re-express the full Master Equation (27).

First, consider its l.h.s., with only one species present—so that ξ and ϕ in (25) are not vectors for the following calculations.

$$
\begin{aligned}
\frac{\partial \Pi(\xi,t)}{\partial t} &= \frac{\partial P(n(t),t)}{\partial t} = \\
&= \frac{\partial P(V\phi + V^{1/2}\xi, t)}{\partial t} = \\
&= \frac{\partial P}{\partial t} + \frac{dn}{dt}\frac{\partial P}{\partial n} = \\
&= \frac{\partial P}{\partial t} + V\frac{d\phi}{dt}\frac{\partial P}{\partial n} = \\
&= \frac{\partial P}{\partial t} + V^{1/2}\frac{d\phi}{dt}\frac{\partial \Pi}{\partial \xi},
\end{aligned}
\tag{29}
$$

where in the last equality, the following one has been used,

$$
\frac{\partial P}{\partial n} = V^{-1/2}\frac{\partial \Pi}{\partial \xi}.
\tag{30}
$$

Therefore,

$$\frac{\partial P(\vec{n},t)}{\partial t} = \frac{\partial \Pi(\vec{\xi},t)}{\partial t} - V^{1/2}\sum_i \frac{d\phi_i}{dt}\frac{\partial \Pi(\vec{\xi},t)}{\partial \xi_i}, \tag{31}$$

where again, index i covers all species involved, in this case mRNA (index R) and miRNA (index S).

By defining

$$\tau = \frac{t}{V},$$

equation (30) becomes:

$$\frac{\partial P(\vec{n},t)}{\partial t} = \frac{1}{V}\left[\frac{\partial \Pi(\vec{\xi},\tau)}{\partial \tau} - V^{1/2}\sum_i \frac{d\phi_i}{d\tau}\frac{\partial \Pi(\vec{\xi},\tau)}{\partial \xi_i}\right]. \tag{32}$$

The full Master Equation will amount to something of the form

$$\frac{\partial \Pi(\vec{\xi},\tau)}{\partial \tau} - V^{1/2}\sum_i \frac{d\phi_i}{d\tau}\frac{\partial \Pi(\vec{\xi},\tau)}{\partial \xi_i} = V\left[...\right]. \tag{33}$$

Since the calculations on the r.h.s. are more involved, the terms related to mRNA, miRNA and complex miRNA-mRNA are treated separately, particularly:

- mRNA terms:

$$\begin{aligned}
&k_R(E_R^{-1}-1)P(\vec{n},t) + \frac{g_R}{V}(E_R-1)n_R P(\vec{n},t) = \\
&= k_R(-V^{-1/2}\frac{\partial}{\partial \xi_R} + \frac{1}{2}V^{-1}\frac{\partial^2}{\partial \xi_R^2})\Pi(\vec{\xi},t) + \\
&+ \frac{g_R}{V}(V^{-1/2}\frac{\partial}{\partial \xi_R} + \frac{1}{2}V^{-1}\frac{\partial^2}{\partial \xi_R^2})(V\phi_R + V^{1/2}\xi_R)\Pi(\vec{\xi},t) = \\
&= k_R(-V^{-1/2}\frac{\partial}{\partial \xi_R} + \frac{1}{2}V^{-1}\frac{\partial^2}{\partial \xi_R^2})\Pi(\vec{\xi},t) + \\
&+ \frac{g_R}{V}(V^{1/2}\phi_R\frac{\partial}{\partial \xi_R} + \frac{1}{2}\phi_R\frac{\partial^2}{\partial \xi_R^2} + \frac{\partial}{\partial \xi_R}\xi_R + \frac{1}{2}V^{-1/2}\frac{\partial^2}{\partial \xi_R^2}\xi_R)\Pi(\vec{\xi},t) = \\
&\overset{(1)}{=} k_R(-V^{1/2}\frac{\partial}{\partial \xi_R} + \frac{1}{2}\frac{\partial^2}{\partial \xi_R^2})\Pi(\vec{\xi},\tau) + \\
&+ g_R(V^{1/2}\phi_R\frac{\partial}{\partial \xi_R} + \frac{1}{2}\phi_R\frac{\partial^2}{\partial \xi_R^2} + \frac{\partial}{\partial \xi_R}\xi_R + \frac{1}{2}V^{-1/2}\frac{\partial^2}{\partial \xi_R^2}\xi_R)\Pi(\vec{\xi},\tau)
\end{aligned} \tag{34}$$

- miRNA terms:

$$
\begin{aligned}
&k_S(E_S^{-1}-1)P(\vec{n},t)+\frac{g_S}{V}(E_S-1)n_SP(\vec{n},t)= \\
&\overset{(1)}{=} k_S(V^{-1/2}\frac{\partial}{\partial\xi_S}+\frac{1}{2}\frac{\partial^2}{\partial\xi_S^2})\Pi(\vec{\xi},\tau)+ \\
&+g_S(V^{1/2}\phi_S\frac{\partial}{\partial\xi_S}+\frac{1}{2}\phi_S\frac{\partial^2}{\partial\xi_S^2}+\frac{\partial}{\partial\xi_S}\xi_S+\frac{1}{2}V^{-1/2}\frac{\partial^2}{\partial\xi_S^2}\xi_S)\Pi(\vec{\xi},\tau)
\end{aligned}
\tag{35}
$$

- interaction terms:

$$
\begin{aligned}
&\frac{g\alpha}{V^2}(E_SE_R-1)n_Rn_SP(\vec{n},t)= \\
&=\frac{g\alpha}{V^2}\Big[(1+V^{-1/2}\frac{\partial}{\partial\xi_R}+\frac{1}{2}V^{-1}\frac{\partial^2}{\partial\xi_R^2})(1+V^{-1/2}\frac{\partial}{\partial\xi_S}+\frac{1}{2}V^{-1}\frac{\partial^2}{\partial\xi_S^2})+ \\
&-1\Big](V\phi_R+V^{1/2}\xi_R)(V\phi_S+V^{1/2}\xi_S)\Pi(\vec{\xi},t)= \\
&=\frac{g\alpha}{V^2}\Big[V^{-1/2}(\frac{\partial}{\partial\xi_R}+\frac{\partial}{\partial\xi_S})+\frac{1}{2}V^{-1}(\frac{\partial^2}{\partial\xi_R^2}+\frac{\partial^2}{\partial\xi_S^2})+ \\
&+\frac{1}{2}V^{-3/2}(\frac{\partial}{\partial\xi_R}\frac{\partial^2}{\partial\xi_S^2}+\frac{\partial}{\partial\xi_S}\frac{\partial^2}{\partial\xi_R^2})+V^{-1}\frac{\partial^2}{\partial\xi_R\partial\xi_S}+ \\
&+\frac{1}{4}V^{-2}\frac{\partial^2}{\partial\xi_R^2}\frac{\partial^2}{\partial\xi_S^2}\Big](V^2\phi_S\phi_R+V^{3/2}(\phi_S\xi_R+\xi_S\phi_R)+V\xi_R\xi_S)\Pi(\vec{\xi},t)= \\
&=\frac{g\alpha}{V^2}[(\frac{\partial}{\partial\xi_R}+\frac{\partial}{\partial\xi_S})(V^{3/2}\phi_S\phi_R+V(\phi_S\xi_R+\xi_S\phi_R)+V^{1/2}\xi_S\xi_R)+ \\
&+\frac{1}{2}(\frac{\partial^2}{\partial\xi_R^2}+\frac{\partial^2}{\partial\xi_S^2})(V\phi_S\phi_R+V^{1/2}(\phi_S\xi_R+\xi_S\phi_R)+\xi_S\xi_R)+ \\
&+\frac{1}{2}(\frac{\partial}{\partial\xi_R}\frac{\partial^2}{\partial\xi_S^2}+\frac{\partial}{\partial\xi_S}\frac{\partial^2}{\partial\xi_R^2})(V^{1/2}\phi_S\phi_R+\phi_S\xi_R+\xi_S\phi_R+V^{-1/2}\xi_S\xi_R)+ \\
&+(\frac{\partial^2}{\partial\xi_R\partial\xi_S})(V\phi_S\phi_R+V^{1/2}(\phi_S\xi_R+\xi_S\phi_R)+\xi_S\xi_R)+ \\
&+\frac{1}{4}(\frac{\partial^2}{\partial\xi_R^2}\frac{\partial^2}{\partial\xi_S^2})(\phi_S\phi_R+V^{-1/2}(\phi_S\xi_R+\xi_S\phi_R)+V^{-1}\xi_S\xi_R)]\Pi(\vec{\xi},t)=
\end{aligned}
$$

$$
\begin{aligned}
&\overset{(1)}{=} g\alpha[(\frac{\partial}{\partial\xi_R}+\frac{\partial}{\partial\xi_S})(V^{1/2}\phi_S\phi_R+\phi_S\xi_R+\phi_R\xi_S+V^{-1/2}\xi_S\xi_R)+\\
&\quad+\frac{1}{2}(\frac{\partial^2}{\partial\xi_S^2}+\frac{\partial^2}{\partial\xi_R^2})(\phi_S\phi_R+V^{-1/2}(\phi_S\xi_R+\phi_R\xi_S)+V^{-1}\xi_S\xi_R)+\\
&\quad+\frac{1}{2}(\frac{\partial}{\partial\xi_R}\frac{\partial^2}{\partial\xi_S^2}+\frac{\partial}{\partial\xi_S}\frac{\partial^2}{\partial\xi_R^2})(V^{-1/2}\phi_S\phi_R+V^{-1}(\phi_S\xi_R+\phi_R\xi_S)+V^{-3/2}\xi_S\xi_R)+\\
&\quad+(\frac{\partial^2}{\partial\xi_R\partial\xi_S})(\phi_S\phi_R+V^{-1/2}(\phi_S\xi_R+\phi_R\xi_S)+V^{-1}\xi_S\xi_R)+\\
&\quad+\frac{1}{4}(\frac{\partial^2}{\partial\xi_R^2}\frac{\partial^2}{\partial\xi_S^2})(V^{-1}\phi_S\phi_R+V^{-3/2}(\phi_S\xi_R+\phi_R\xi_S)+V^{-2}\xi_S\xi_R)]\Pi(\vec{\xi},\tau)=\\
&=g\alpha[(\frac{\partial}{\partial\xi_R}+\frac{\partial}{\partial\xi_S})(V^{1/2}\phi_S\phi_R+\phi_S\xi_R+\phi_R\xi_S+V^{-1/2}\xi_S\xi_R)+\\
&\quad+\frac{1}{2}(\frac{\partial^2}{\partial\xi_S^2}+\frac{\partial^2}{\partial\xi_R^2})(\phi_S\phi_R+V^{-1/2}(\phi_S\xi_R+\phi_R\xi_S)+V^{-1}\xi_S\xi_R)+\\
&\quad+\frac{1}{2}(\frac{\partial}{\partial\xi_R}\frac{\partial^2}{\partial\xi_S^2}+\frac{\partial}{\partial\xi_S}\frac{\partial^2}{\partial\xi_R^2})(V^{-1/2}\phi_S\phi_R+V^{-1}(\phi_S\xi_R+\phi_R\xi_S))+\\
&\quad+(\frac{\partial^2}{\partial\xi_R\partial\xi_S})(\phi_S\phi_R+V^{-1/2}(\phi_S\xi_R+\phi_R\xi_S)+V^{-1}\xi_S\xi_R)+\\
&\quad+\frac{1}{4}(\frac{\partial^2}{\partial\xi_R^2}\frac{\partial^2}{\partial\xi_S^2})V^{-1}\phi_S\phi_R]\Pi(\vec{\xi},\tau)
\end{aligned}
$$

- re-cycled miRNA terms:

$$
\begin{aligned}
&\frac{g(1-\alpha)n_S}{V^2}(E_R-1)n_RP(\vec{n},t)=\\
&=\frac{g(1-\alpha)}{V^2}(V\phi_S+V^{1/2}\xi_S)(1+V^{-1/2}\frac{\partial}{\partial\xi_R}+\frac{1}{2}V^{-1}\frac{\partial^2}{\partial\xi_R^2}+\cdots-1)(V\phi_R+\\
&\quad+V^{1/2}\xi_R)\Pi(\vec{\xi},t)=\\
&\overset{(1)}{=} g(1-\alpha)(\phi_S+V^{-1/2}\xi_S)(V^{1/2}\phi_R\frac{\partial}{\partial\xi_R}+\frac{\partial}{\partial\xi_R}\xi_R+\frac{1}{2}\phi_R\frac{\partial^2}{\partial\xi_R^2}+\\
&\quad+\frac{1}{2}V^{-1/2}\frac{\partial^2}{\partial\xi_R^2}\xi_R)\Pi(\vec{\xi},\tau)=\\
&=g(1-\alpha)[V^{1/2}\phi_S\phi_R\frac{\partial}{\partial\xi_R}+\phi_S\frac{\partial}{\partial\xi_R}\xi_R+\frac{1}{2}\phi_S\phi_R\frac{\partial^2}{\partial\xi_R^2}\\
&\quad+\frac{1}{2}V^{-1/2}\phi_S\frac{\partial^2}{\partial\xi_R^2}\xi_R+\xi_S\phi_R\frac{\partial}{\partial\xi_R}+V^{-1/2}\xi_S\frac{\partial}{\partial\xi_R}\xi_R\\
&\quad+\frac{1}{2}V^{-1/2}\xi_S\phi_R\frac{\partial^2}{\partial\xi_R^2}+\frac{1}{2}V^{-1}\xi_S\frac{\partial^2}{\partial\xi_R^2}\xi_R]\Pi(\vec{\xi},\tau)
\end{aligned}
\tag{36}
$$

where in each (1) the procedure stated in equation (33) has been imposed.

Taking now all terms of order $\mathcal{O}(V^{1/2})$, both on the right-hand and left-hand sides of the equation, one may verify that the correct deterministic equations are recovered.

$$
\begin{aligned}
&- V^{1/2}\frac{d\phi_S}{d\tau}\frac{\partial \Pi}{\partial \xi_S} - V^{1/2}\frac{d\phi_R}{d\tau}\frac{\partial \Pi}{\partial \xi_R} = \\
&= (-k_R + g_R\phi_R)V^{1/2}\frac{\partial \Pi}{\partial \xi_R} + (-k_S + g_S\phi_S)V^{1/2}\frac{\partial \Pi}{\partial \xi_S} + \\
&+ V^{1/2}g\alpha\phi_S\phi_R(\frac{\partial}{\partial \xi_S} + \frac{\partial}{\partial \xi_R})\Pi + g(1-\alpha)V^{1/2}\phi_S\phi_R\frac{\partial \Pi}{\partial \xi_R} \\
&\qquad\qquad \Leftrightarrow \\
&- \frac{d\phi_S}{d\tau}\frac{\partial \Pi}{\partial \xi_S} - \frac{d\phi_R}{d\tau}\frac{\partial \Pi}{\partial \xi_R} = \\
&= (-k_R + g_R\phi_R + g\alpha\phi_S\phi_R + g(1-\alpha)\phi_S\phi_R)\frac{\partial \Pi}{\partial \xi_R} + \\
&+ (-k_S + g_S\phi_S + g\alpha\phi_S\phi_R)\frac{\partial \Pi}{\partial \xi_S}
\end{aligned}
\tag{37}
$$

$$
\Leftrightarrow
$$

$$
\begin{aligned}
-\dot{\phi_R} &= -k_R + g_R\phi_R + g\phi_S\phi_R, \\
-\dot{\phi_S} &= -k_S + g_S\phi_S + g\alpha\phi_S\phi_R.
\end{aligned}
\tag{38}
$$

Taking instead all terms of order $\mathcal{O}(V^0)$, the resulting equation is a Fokker-Planck Equation.
Particularly, it is possible to show that the solution of such Fokker-Planck equation is a Gaussian [24], therefore only the first and second moments must be determined to gain information about the distribution. For this reason the Linear Noise Approximation is often referred to as the Gaussian approximation. The Fokker-Planck equation obtained in this case of study is

$$
\begin{aligned}
\frac{\partial \Pi(\vec{\xi},\tau)}{\partial \tau} &= (\frac{1}{2}k_R\frac{\partial^2}{\partial \xi_R^2} + \frac{1}{2}g_R\phi_R\frac{\partial^2}{\partial \xi_R^2} + g_R\frac{\partial}{\partial \xi_R}\xi_R)\Pi(\vec{\xi},\tau) + \\
&+ (\frac{1}{2}k_S\frac{\partial^2}{\partial \xi_S^2} + \frac{1}{2}g_S\phi_S\frac{\partial^2}{\partial \xi_S^2} + g_S\frac{\partial}{\partial \xi_S}\xi_S)\Pi(\vec{\xi},\tau) + \\
&+ g\alpha[(\frac{\partial}{\partial \xi_R} + \frac{\partial}{\partial \xi_S})(\phi_S\xi_R + \xi_S\phi_R) + \\
&+ \frac{1}{2}(\frac{\partial^2}{\partial \xi_S^2} + \frac{\partial^2}{\partial \xi_R^2})\phi_S\phi_R + \phi_S\phi_R\frac{\partial^2}{\partial \xi_S\partial \xi_R}]\Pi(\vec{\xi},\tau) + \\
&+ g(1-\alpha)[\phi_S\frac{\partial}{\partial \xi_R}\xi_R + \xi_S\phi_R\frac{\partial}{\partial \xi_R} + \frac{1}{2}\phi_S\phi_R\frac{\partial^2}{\partial \xi_R^2}]\Pi(\vec{\xi},\tau).
\end{aligned}
\tag{39}
$$

First and second moments of ξ gaussian noise Consider the following linear Fokker-Plank Equation, where for simplicity only one chemical species has been involved, therefore ξ is not a vector,

$$
\frac{\partial \Pi(\xi,\tau)}{\partial \tau} = -\alpha_1(\phi)\frac{\partial}{\partial \xi}\xi\Pi(\xi,\tau) + \frac{1}{2}\alpha_2(\phi)\frac{\partial^2\Pi(\xi,\tau)}{\partial \xi^2},
\tag{40}
$$

where $\alpha_j, j = 1, 2$ are two coefficients depending on time through concentration $\phi(t)$. With the aim of determine both first and second moments, multiply by ξ and ξ^2 to respectively obtain

$$
\begin{aligned}
&\frac{\partial\langle\xi\rangle}{\partial\tau} = \alpha_1(\phi)\langle\xi\rangle, \\
&\frac{\partial\langle\xi^2\rangle}{\partial\tau} = 2\alpha_1(\phi)\langle\xi^2\rangle + \alpha_2(\phi).
\end{aligned}
\tag{41}
$$

Both equations can be solved, provided that the initial condition is known. Particularly, since the initial problem was to solve the Master Equation for $P(n,t)$ condition to the fact that at initial time $P(n,0) = \delta_{n,0}$, then at the beginning of the process fluctuations vanish, i.e.:

$$
\langle\xi\rangle_0 = \langle\xi^2\rangle_0 = 0.
\tag{42}
$$

As a consequence,

$$
\begin{aligned}
&\langle\xi\rangle_t = A\, e^{-\int_0^t \alpha_1(\phi(s))ds}, \\
&\text{ therefore, using initial condition (42),} \\
&\langle\xi\rangle_0 = 0 = A \;\;\Rightarrow\;\; \langle\xi\rangle_t = 0 \;\; \forall t.
\end{aligned}
\tag{43}
$$

The fact that the noise has zero mean implies that the variance is equal to

the second moment, i.e.:

$$Var[\xi] = \langle\langle \xi^2 \rangle\rangle = \langle \xi^2 \rangle - \langle \xi \rangle^2 = \langle \xi^2 \rangle. \tag{44}$$

Then, passing from distribution $\Pi(\xi, t)$ to distribution $P(n, t)$, according to relation (25), the connection between number of molecules n and noise ξ is needed. Particularly, the following relations will be used:

- $$\langle n \rangle_t = V\phi(t) + V^{\frac{1}{2}} \cancel{\langle \xi \rangle_t} = V\phi(t) \tag{45}$$

- $$\langle\langle n^2 \rangle\rangle_t = (V\phi(t))^2 + V\langle\langle \xi^2 \rangle\rangle_t - (V\phi(t))^2 \overset{(1)}{=} V\langle \xi^2 \rangle_t \tag{46}$$

  where in step (1) relation (44) was used.

For the Fokker-Plank Equation (39), the noise is a vector with two components $\vec{\xi} = (\xi_R, \xi_S)$, for each of which relations (45) and (46) must hold.

Going back to equation (39) and applying the procedure outlined in the above paragraph, therefore multiplying by ξ_R^2, ξ_S^2 and $\xi_R \xi_S$ and integrating over all their possible values, one obtains equations for autocorrelation and cross-correlation, i.e.:

$$\begin{aligned}
&\dot{\langle \xi_R^2 \rangle} = k_R + g_R\phi_R - 2g_R\langle \xi_R^2 \rangle - 2g\phi_S\langle \xi_R^2 \rangle - 2g\phi_R\langle \xi_R\xi_S \rangle + g\phi_S\phi_R, \\
&\dot{\langle \xi_S^2 \rangle} = k_S + g_S\phi_S - 2g_S\langle \xi_S^2 \rangle - 2g\alpha\phi_S\langle \xi_R\xi_S \rangle - 2g\alpha\phi_R\langle \xi_S^2 \rangle + g\alpha\phi_S\phi_R, \\
&\dot{\langle \xi_R\xi_S \rangle} = -(g_R + g_S + g\phi_S + g\alpha\phi_R)\langle \xi_S\xi_R \rangle - g\phi_R\langle \xi_S^2 \rangle - g\alpha\phi_S\langle \xi_R^2 \rangle + \\
&+ g\alpha\phi_R\phi_S.
\end{aligned} \tag{47}$$

These three equations can be solved when paired with the deterministic ones (38) describing the evolution of the concentration of mRNA and miRNA.

As a result, the probability $P(\vec{n}, t)$ of having $\vec{n}$ at time t is obtained as a Gaussian distribution [24] with

$$\langle \vec{n} \rangle = (\langle n_R \rangle, \langle n_S \rangle) = V(\phi_R, \phi_S),$$

$$\langle\langle \vec{n}^2 \rangle\rangle = (\langle\langle n_R^2 \rangle\rangle_t, \langle\langle n_S^2 \rangle\rangle_t) = V(\langle \xi_R^2 \rangle_t, \langle \xi_S^2 \rangle_t).$$

Following again the procedure outlined in Part II, the First-Passage Time distribution is obtained.

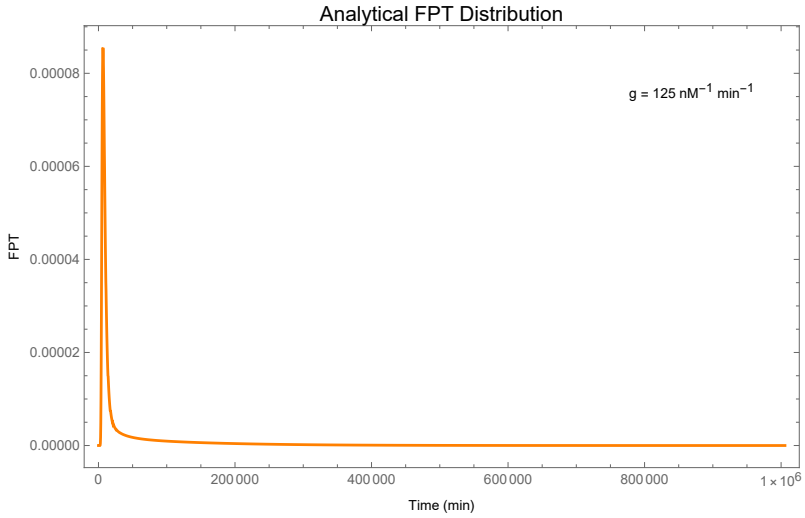


Figure 18: Analytical First-Passage Time distribution with binding affinity $g = 125\ nM^{-1}min^{-1}$. All other parameters are set to the values reported in Appendix, Table 2.

7 Results

In this Chapter all numerical and analytical results will be presented and compared.
The analytical First-Passage Time distribution will be compared with the one gained from Gillespie simulations both for the single-gene and for non-zero binding affinity. Particularly, for the interacting circuit it will be established if and when the approximation implemented, i.e. the Van Kampen expansion, is a good and valid approximation for the case of study.

7.1 The non-interacting circuit

The system represented in Figure 4 has been studied while keeping a null binding affinity ($g = 0$) between mRNA and miRNA molecules, therefore the number of messenger-RNA molecules is not regulated. The analytical First-Passage Time distribution is below reported and compared with the one obtained through Gillespie simulations.

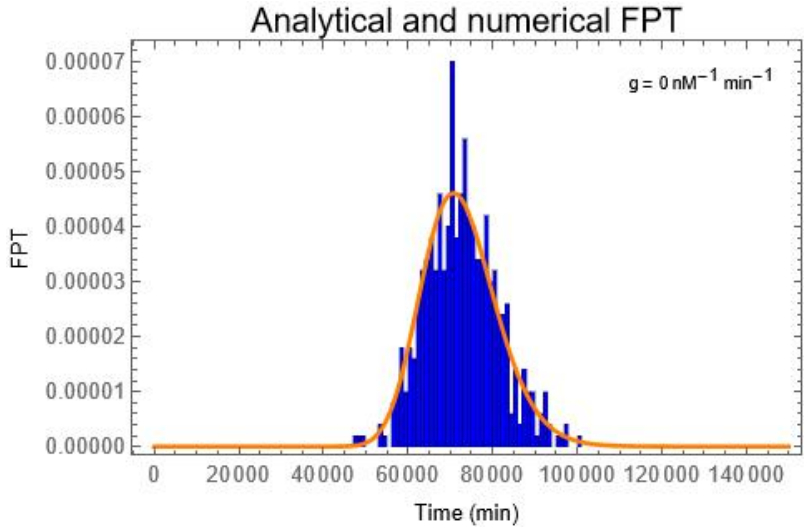


Figure 19: Analytical (orange) and numerical (blue histogram) First-Passage Time distribution for $g = 0\ nM^{-1}min^{-1}$. All other parameters are set to the values reported in Appendix, Table 2.

Also the properties of such distribution may be computed and compared with simulations results. Particularly mean, variance, coefficient of variation and fano factor of the First-Passage Time distribution have been computed upon varying the value of the transcription rate k_R of mRNA molecules, specifically form 0.001 to 0.015 $nM\ min^{-1}$. All parameters, with exception of k_R

and g, are set to the values reported in Appendix, Table 2. The results are reported in the following.

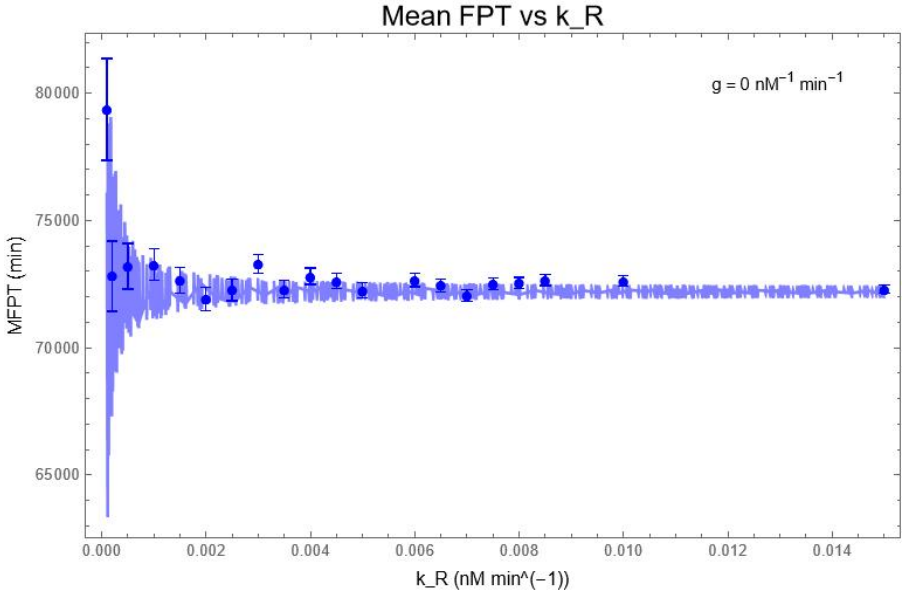


Figure 20: Analytical mean First-Passage Time as a function of transcription rate k_R for $g = 0\ nM^{-1}min^{-1}$. The blue dots are the numerical simulations results.

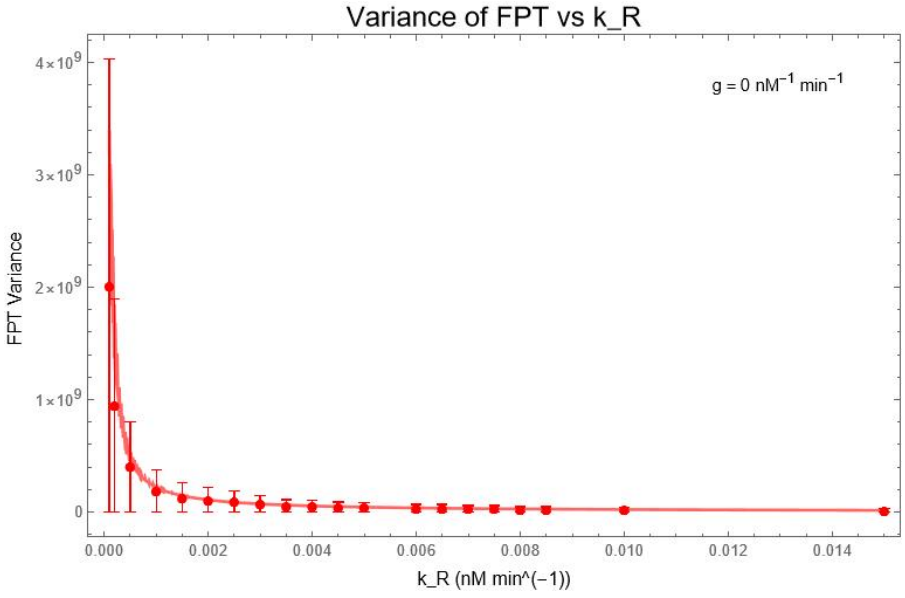


Figure 21: Analytical variance of First-Passage Time distribution as a function of transcription rate k_R for $g = 0\ nM^{-1}min^{-1}$. The red dots are the numerical simulations results.

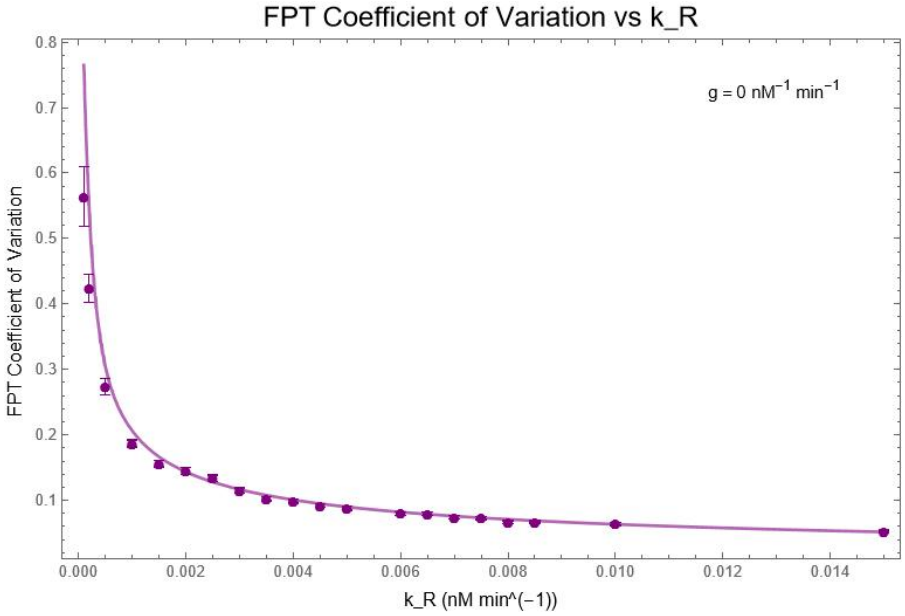


Figure 22: Analytical Coefficient of Variation (CV) of First-Passage Time distribution as a function of transcription rate k_R for $g = 0\ nM^{-1}min^{-1}$. The purple dots are the numerical simulations results.

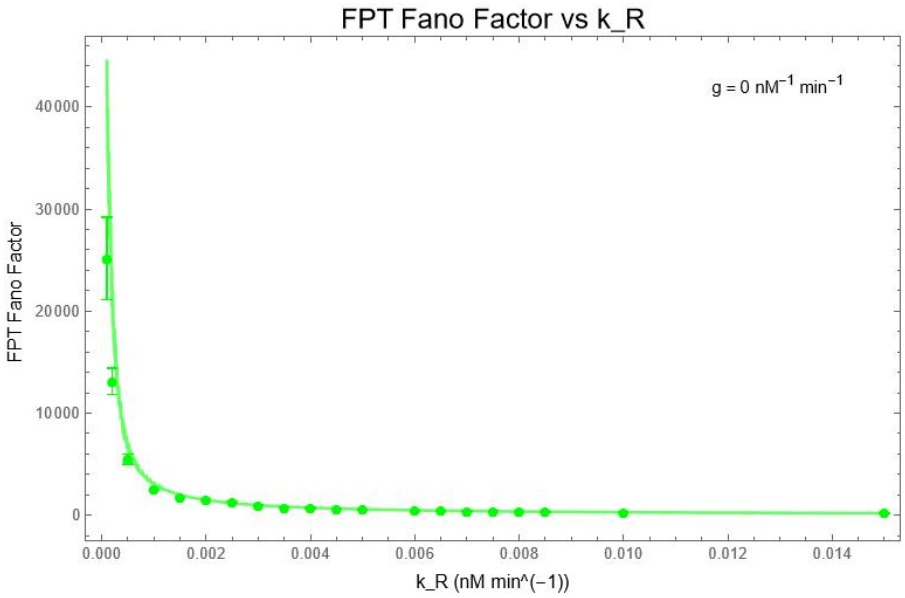


Figure 23: Analytical Fano Factor of First-Passage Time distribution as a function of transcription rate k_R for $g = 0\ nM^{-1}min^{-1}$. The green dots are the numerical simulations results.

7.2 The interacting circuit

Upon including the post-transcriptional regulation in the Gillespie simulations, namely setting parameter g to a non-zero value, the trajectories of mRNA and miRNA change accordingly, as is shown in Figure 24.

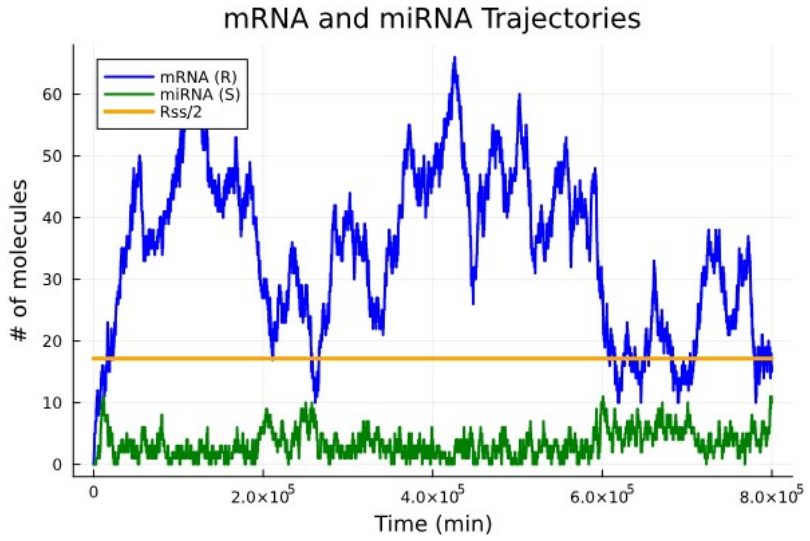


Figure 24: Trajectories of miRNA and mRNA molecules, respectively green and blue line, for a value of binding affinity $g = 150\ nM^{-1}min^{-1}$. The yellow line is the half-steady state line for mRNA. All other parameters are set to the values reported in Appendix, Table 2.

With the aim of comparing the analytical approach carried out and the numerical results obtained through the Gillespie algorithm, the First-Passage Time distribution has been computed for several values of binding affinity g. For each chosen value the analytical and numerical FPT distributions can be compared, as shown in Figure 25 where the two are overlaid.

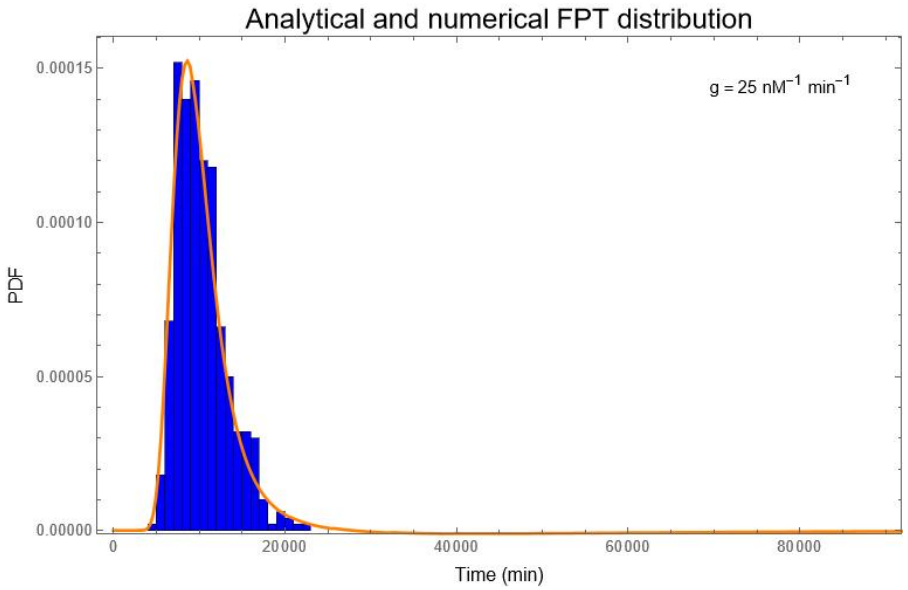


Figure 25: Comparison of analytical (orange line) and numerical (blue histogram) FPT distributions for $g = 25\ nM^{-1}min^{-1}$. All other parameters are set to the values reported in Appendix, Table 2.

To verify the consistency of the two approaches, the mean of the First-Passage Time distribution (MFPT) has been computed both analytically and numerically for different values of binding affinity g. The analytical approximation is reproducing the behavior of the numerical MFPT (i.e., obtained with Gillespie algorithm) only for small values of interaction strength. Upon increasing this parameter, the mean FPT computed via the Van Kampen approximation overestimates the correct Gillespie mean First-Passage Time, as shown in Figure 26.

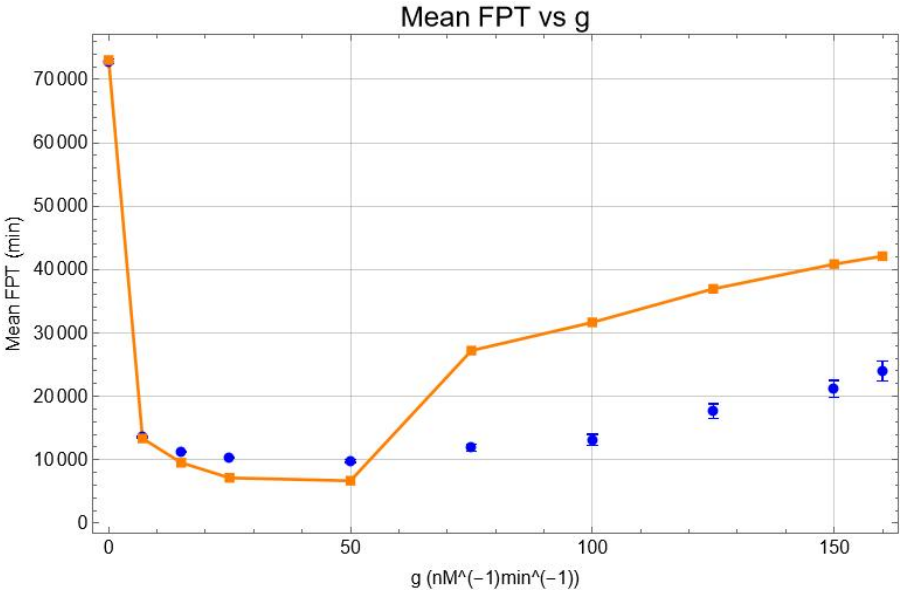


Figure 26: Mean of the First-Passage Time distribution (MFPT), which has been computed for different values of binding affinity $g \in \{0, 7, 15, 25, 50, 75, 100, 125, 150, 160\}$ nM^{-1} min^{-1}. The orange dots are the analytical MFPT gained through the Linear Noise Approximation, while the blue dots are the ones gained through Gillespie simulations. All other parameters are set to the values reported in Appendix, Table 2.

Looking at the comparison of analytical and numerical MFPTs, we may see that the approximation is not valid above a certain critical value of g.

Particularly, all analytical data beyond $g = 50$ nM^{-1} min^{-1} are roughly double the corresponding Gillespie MFPT. This observation could suggest a systematic error in the computation of the analytical FPT distribution. To be more precise, the variances of all FPT-distributions for the several values of binding affnity have been computed. As expected, this quantity increases as parameter g increases; in particular, for all distributions characterized by a higher value of binding affinity than the critical one (roughly beyond $g = 50$ nM^{-1} min^{-1}), variances become very large, definitely confirming that the Linear Noise Approximation is failing at describing the mRNA-miRNA interaction correctly in this regime.

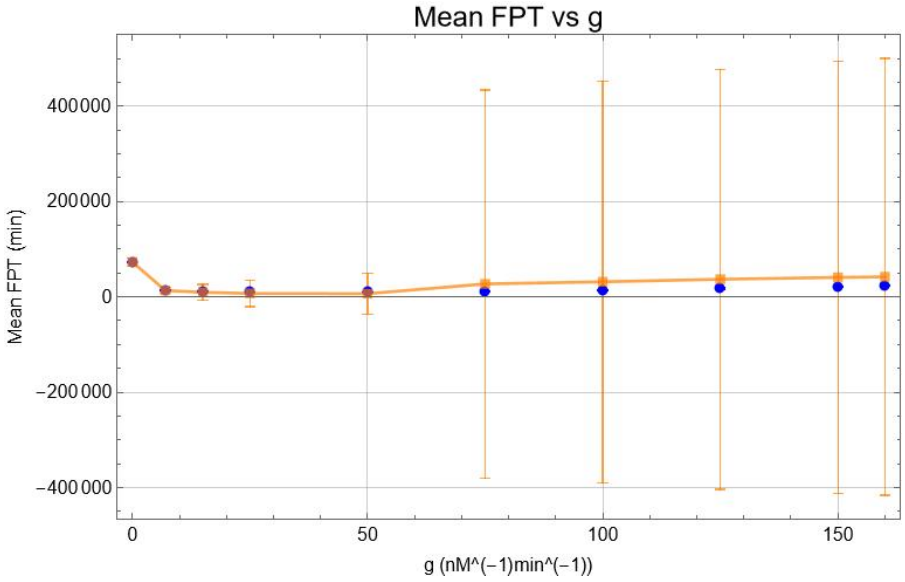


Figure 27: Mean of the First-Passage Time distribution (MFPT) as a function of g. The analytical data (orange) and the numerical data (blue) are computed in the same parameters set up of Figure 26.

Furthermore, the dependence of the First-Passage Time distribution on mRNA transcription rate k_R was studied. Upon increasing this parameter, for a predetermined value binding affinity, the MFPT increases and converges to the MFPT value characterizing the single-gene distribution (blue line of Figure 28).

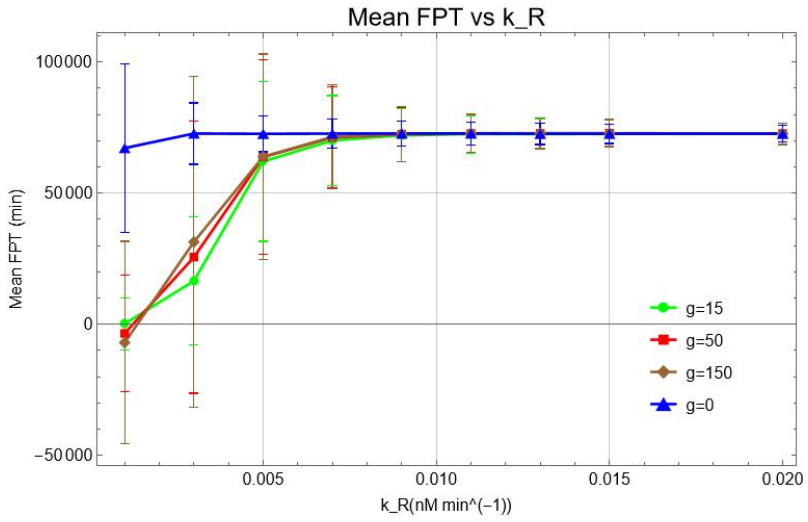
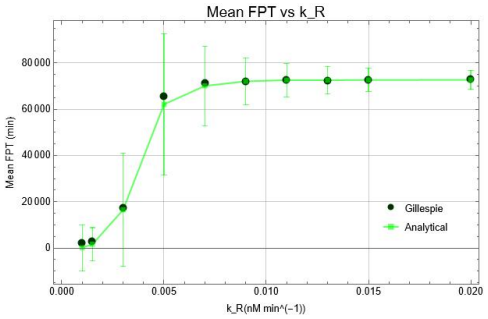
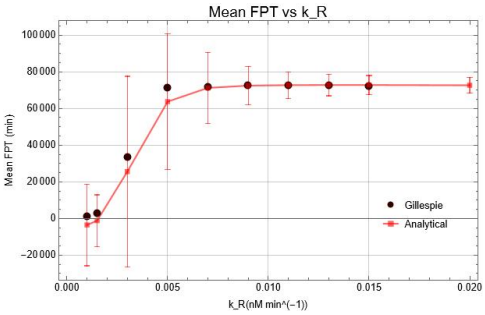


Figure 28: Mean of the analytical First-Passage Time distribution as a function of mRNA transcription rate k_R, for different values of binding affinity, i.e.: $g \in \{0, 15, 50, 150\}\ nM^{-1}\ min^{-1}$. All other parameters values in Appendix, Table 2.

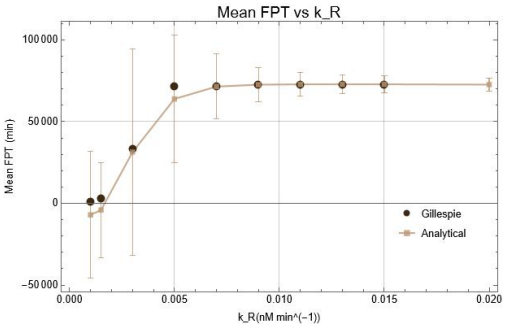
For completeness trajectories of Figure 28 were compared with the mean of the FPT distribution computed with the Gillespie algorithm. The comparison of the analytical and numerical unregulated MFPT behavior has already been shown in Figure 20.



(a) $g = 15\ nM^{-1}min^{-1}$



(b) $g = 50\ nM^{-1}min^{-1}$



(c) $g = 150\ nM^{-1}min^{-1}$

Figure 29: Mean of the analytical and numerical FPT distribution as a function of mRNA transcription rate k_R, for three different values of$g$. All other parameters are set to the values reported in Appendix, Table 2.

8 Conclusions

The main aim of this thesis was to shift the focus from the steady-state properties to the dynamical properties of microRNA-mediated gene regulation networks. To this end, a First-Passage Time framework for the mRNA molecules was developed, as this metric is fundamental for understanding the timing at which cells execute crucial biological processes. According to this framework, in fact, the timing of a process is established as soon as the level of expression of the corresponding molecular species reaches a certain threshold.

Specifically, to address this objective a dual approach was implemented: numerical and analytical.

The numerical approach is based on the implementation of the Gillespie algorithm, which allows for the generation of exact stochastic trajectories of all molecular species present in the system of interest. Therefore, this enabled the reconstruction of the First-Passage Time distribution for both the unregulated single-gene circuit and the more complex miRNA-regulated circuit.

The challenge was to compute the First-Passage Time distributions not only via numerical simulation, but also analytically. For the unregulated case, the procedure carried out leads to the exact First-Passage Time distribution, as the comparison with the Gillespie FPT distribution confirms (Figure 19). Also, the properties of both the analytical and numerical distribution were calculated, specifically by varying the transcription rate k_R of mRNA molecules. All the analytical properties studied—mean, variance, coefficient of variation and fano factor—are in excellent agreement with the numerical predictions, therefore confirming once again that the analytical distribution obtained is consistent with the one provided by the Gillespie algorithm.

Upon introducing the post-transcriptional regulation mediated by microRNA molecules, the Master Equation which describes the process is hard-to-solve, therefore an approximation is needed to gain an analytical expression for the First-Passage Time distribution of interest. The analytical approximation implemented is the Van Kampen expansion, also known as Linear Noise or Gaussian Approximation.

To systematically compare the numerical and analytical results of the regulated case, the mean of the FPT distribution (MFPT) has been computed for different values of mRNA-miRNA interaction strength. As shown in Figure 27, the approximation proves valid for low values of binding affinity. As further confirmation, the unregulated FPT distribution is recovered within this

approximation by setting a null binding affinity.

However, as the interaction strength increases, the validity of the approximation is compromised by the rapid growth of the variance. The increase in variance upon increasing the binding affinity is physically expected, as this implies stronger miRNA-mediated regulation. In this regime, the strong activity of miRNAs induces large stochastic fluctuations in the mRNA molecule number, resulting in a broad distribution of crossing times. However, the fact that the variance grows in time much faster than the mean leads to the loss of temporal precision of the process, making the mean First-Passage Time highly unpredictable. This behavior indicates that the mean value ceases to be a representative descriptor of the system's dynamics.
The breakdown of the approximation could also be attributed to the nature of the Van Kampen expansion, which models the probability of having a certain number of mRNA molecules at a given time as a Gaussian distribution. In contrast, increasing the mRNA-miRNA interaction strength drives the system away from Gaussian behavior, so much so that the mRNA molecule number distributions eventually exhibits bimodality. Although full bimodality may not yet emerge within the parameter range investigated, the distributions begin to develop long tails and marked asymmetry, features that the Gaussian Approximation fails to capture.

Therefore, while for small values of interaction the approximation is still valid, beyond a certain critical value—roughly $g = 50\ nM^{-1}min^{-1}$—the Van Kampen expansion fails to correctly describe the behavior of the system.

Nonetheless, an interesting finding is that a cell, by employing a miRNA-mediated post-transcriptional control, can actually reduce the time at which the number of mRNAs reaches a given threshold, while maintaining low noise levels. Therefore, this shows that microRNA regulation plays a crucial and fundamental role in event timing, particularly accelerating the initiation of a given biological process and also acting as powerful noise buffer.

Furthermore, the MFPT curve (Figure 26) shows a minimum, suggesting that there is an optimal value of binding affinity such that the MFPT is minimized.
For completeness, the mRNA MFPT has been studied upon varying the transcription rate k_R for different binding affinity values, as shown in Figure 28. Reasonably, as the transcription rate increases enough, the regulatory effect is washed out. The limiting factor becomes the intrinsic degradation rather than the miRNA interaction, causing the statistical properties of the regulated

system—including the MFPT—to asymptotically approach those of the unregulated circuit.

It is worth noting that the analytical approximation implemented could be refined to gain better insights into this behavior. A first option could be to extend this same procedure, but including higher orders in the expansion of the Master Equation. While this would increase the analytical complexity, it would provide corrections to the Gaussian behavior.

Alternatively, the limitations of the approximation could be overcome by developing a field-theoretic framework, for example following the methodology of B. Walter et al (2023) [25]. By mapping the stochastic dynamics onto a Doi-Peliti formalism equipped with a specific tracing mechanism, this method allows for the analytical derivation of the visit probability, from which the exact First-Passage Time distribution can be obtained.

Appendix

Parameters definition

Definition	Symbol	Value	Unit
mRNA transcription rate	k_R	2.7×10^{-3}	$nM\ min^{-1}$
miRNA transcription rate	k_S	1.2×10^{-3}	$nM\ min^{-1}$
mRNA degradation rate	g_R	2.4×10^{-2}	min^{-1}
miRNA degradation rate	g_S	2.4×10^{-2}	min^{-1}
mRNA-miRNA binding affinity	g	1.5×10^{3}	$nM^{-1}\ min^{-1}$
protein translation rate	k_P	6.0	min^{-1}
protein degradation rate	g_P	1.2×10^{-2}	min^{-1}
Fraction of degraded miRNAs from mRNA-miRNA complex	α	0.5	–

Table 1: Parameters values, adapted from [6].

Definition	Symbol	Value	Unit
mRNA transcription rate	k_R	2.7×10^{-3}	$nM\ min^{-1}$
miRNA transcription rate	k_S	1.2×10^{-3}	$nM\ min^{-1}$
mRNA degradation rate	g_R	2.4×10^{-2}	min^{-1}
miRNA degradation rate	g_S	2.4×10^{-2}	min^{-1}
mRNA-miRNA binding affinity	g	1.5×10^{3}	$nM^{-1}\ min^{-1}$
Fraction of degraded miRNAs from mRNA-miRNA complex	α	0.5	–

Table 2: Parameters values for numerical simulations and analytical calculation, adapted from [6]. As proteins are not discussed, their translation and degradation rates are not needed for numerical neither analytical procedure.

Transform of full master equation

An attempt to solve the full master equation was made following Redner approach, but once gained an equation for the transformed probability, going back to P(n,t) is not straightforward as thought. Here are reported the calculations made for the case $g \neq 0$:

As did for the $g = 0$ case, the Master Equation can be transformed according to the following rule:

$$M(z_R, z_S, z_P, t) = \sum_{\vec{n}=0}^{\infty} P(\vec{n}, t) z_R^{n_R} z_S^{n_S} z_P^{n_P} \tag{48}$$

where to simplify the notation $z_j := e^{ik_j}$ with $j = (R, S, P)$ was defined.

So, Master Equation 13 becomes:

$$\begin{aligned}
\frac{\partial M(\vec{z}, t)}{\partial t} = \sum_{\vec{n}=0}^{\infty} \Big[& k_R[P(n_R{-}1, t) - P(n_R, t)] + \frac{g_R}{V}[(n_R{+}1)P(n_R{+}1, t) - n_R P(n_R, t)] \\
& + k_S[P(n_S{-}1, t) - P(n_S, t)] + \frac{g_S}{V}[(n_S{+}1)P(n_S{+}1, t) - n_S P(n_S, t)] \\
& + \frac{k_P n_R}{V}[P(n_P{-}1, t) - P(n_P, t)] + \frac{g_P}{V}[(n_P{+}1)P(n_P{+}1, t) - n_P P(n_P, t)] \\
& + \frac{g\alpha}{V}[(n_S{+}1)(n_R{+}1)P(n_R{+}1, n_S{+}1, t) - n_S n_R P(n_R, n_S, t)] \\
& + \frac{g(1{-}\alpha)n_S}{V^2}[(n_R{+}1)P(n_R{+}1, t) - n_R P(n_R, t)] \Big] z_R^{n_R} z_S^{n_S} z_P^{n_P}
\end{aligned} \tag{49}$$

Here below are reported all calculations needed to transform the r.h.s. of equation 49, line by line:

•

$$
\begin{aligned}
&k_R \sum_{\vec{n}=0}^{\infty} P(n_R-1,t) z_R^{n_R} z_S^{n_S} z_P^{n_P} - k_R \sum_{\vec{n}=0}^{\infty} P(n_R,t) z_R^{n_R} z_S^{n_S} z_P^{n_P} + \\
&+ \frac{g_R}{V} \sum_{\vec{n}=0}^{\infty} (n_R+1) P(n_R+1,t) z_R^{n_R} z_S^{n_S} z_P^{n_P} - \frac{g_R}{V} \sum_{\vec{n}=0}^{\infty} n_R P(n_R,t) z_R^{n_R} z_S^{n_S} z_P^{n_P} = \\
&= k_R \sum_{\vec{n}=0}^{\infty} z_R P(n_R-1,t) z_R^{n_R-1} z_S^{n_S} z_P^{n_P} - k_R \sum_{\vec{n}=0}^{\infty} P(n_R,t) z_R^{n_R} z_S^{n_S} z_P^{n_P} + \\
&\frac{g_R}{V} \sum_{\vec{n}=0}^{\infty} \frac{d}{dz_R} P(n_R+1,t) z_R^{n_R+1} z_S^{n_S} z_P^{n_P} - \frac{g_R}{V} \sum_{\vec{n}=0}^{\infty} z_R \frac{d}{dz_R} P(n_R,t) z_R^{n_R} z_S^{n_S} z_P^{n_P} = \\
&= \left[k_R z_R - k_R + \frac{g_R}{V} \frac{d}{dz_R} - \frac{g_R}{V} z_R \frac{d}{dz_R} \right] M(\vec{z},t) = \\
&= \left[k_R (z_R - 1) + \frac{g_R}{V} (1 - z_R) \frac{d}{dz_R} \right] M(\vec{z},t)
\end{aligned}
\tag{50}
$$

• Analogous procedure leads to:

$$
\left[k_s (z_S - 1) + \frac{g_S}{V} (1 - z_S) \frac{d}{dz_S} \right] M(\vec{z},t) \tag{51}
$$

• from now on the protein are to be neglected, but just for completeness:

$$
\left[\frac{k_P}{V} z_R (z_P - 1) \frac{d}{dz_R} + \frac{g_P}{V} (1 - z_P) \frac{d}{dz_P} \right] M(\vec{z},t)
$$

•

$$
\begin{aligned}
&\sum_{\vec{n}=0}^{\infty} \frac{g\alpha}{V^2} [(n_S+1)(n_R+1) P(n_R+1, n_S+, t) - n_S n_R P(n_R, n_S, t)] z_R^{n_R} z_S^{n_S} = \\
&= \frac{g\alpha}{V^2} \sum_{\vec{n}=0}^{\infty} \frac{d}{dz_R} \frac{d}{dz_S} P(n_R+1, n_S+1, t) z_R^{n_R+1} z_S^{n_S+1} + \\
&\quad - \frac{g\alpha}{V^2} \sum_{\vec{n}=0}^{\infty} z_R z_S \frac{d}{dz_R} \frac{d}{dz_S} P(n_R, n_S, t) z_R^{n_R} z_S^{n_S} = \\
&= \frac{g\alpha}{V^2} \left[\frac{d^2}{dz_R dz_S} M(\vec{z},t) - z_R z_s \frac{d^2}{dz_R dz_S} M(\vec{z},t) \right] = \\
&= \frac{g\alpha}{V^2} (1 - z_R z_S) \frac{d^2}{dz_R dz_S} M(\vec{z},t)
\end{aligned}
\tag{52}
$$

•

$$\begin{aligned}
&\sum_{\vec{n}=0}^{\infty} \frac{g(1-\alpha)}{V^2} \left[n_S(n_R+1)P(n_R+1,t) - n_R P(n_R,t) \right] z_R^{n_R} z_S^{n_S} = \\
&= \frac{g(1-\alpha)}{V^2} \left[\sum_{\vec{n}=0}^{\infty} z_S \frac{d}{dz_S} \frac{d}{dz_R} P(n_R+1,t) z_R^{n_R+1} z_S^{n_S} + \right. \\
&\quad \left. - z_R z_S \frac{d}{dz_R} \frac{d}{dz_S} P(n_R,t) z_R^{n_R} z_S^{n_S} \right] = \\
&= \frac{g(1-\alpha)}{V^2} z_S(1-z_R) \frac{d^2}{dz_S dz_R} M(\vec{z},t)
\end{aligned} \tag{53}$$

The full transformed Master Equation (neglecting protein part) reads:

$$\begin{aligned}
\frac{\partial M(\vec{z},t)}{\partial t} = &\left\{ \left[k_R(z_R-1) + \frac{g_R}{V}(1-z_R)\frac{d}{dz_R} \right] + \right. \\
&+ \left[k_s(z_S-1) + \frac{g_S}{V}(1-z_S)\frac{d}{dz_S} \right] + \left[\frac{g\alpha}{V^2}(1-z_R z_S)\frac{d^2}{dz_R dz_S} \right] + \\
&\left. + \left[\frac{g(1-\alpha)}{V^2} z_S(1-z_R)\frac{d^2}{dz_S dz_R} \right] \right\} M(\vec{z},t)
\end{aligned} \tag{54}$$

References

[1] B. Alberts, A. Johnson, J. Lewis, M. Raff, K. Roberts, and P. Walter. *Molecular Biology of the Cell*. Garland Science, 5 edition, 2008.

[2] Oswald T. Avery, Colin M. MacLeod, and Maclyn McCarty. Studies on the chemical nature of the substance inducing transformation of pneumococcal types: Induction of transformation by a desoxyribonucleic acid fraction isolated from pneumococcus type iii. *The Journal of Experimental Medicine*, 79(2):137–158, 1944.

[3] Carla Bosia, Francesco Sgrò, Laura Conti, Carlo Baldassi, Davide Brusa, Federica Cavallo, Ferdinando Di Cunto, Emilia Turco, Andrea Pagnani, and Riccardo Zecchina. Rnas competing for microrna mutually influence their fluctuations in a highly non-linear microrna-dependent manner in single cells. *Genome Biology*, 18(1):37, 2017.

[4] F. H. C. Crick, L. Barnett, S. Brenner, and R. J. Watts-Tobin. General nature of the genetic code for proteins. *Nature*, 192(4809):1227–1232, 1961.

[5] A. Dal Co, M. Cosentino Lagomarsino, M. Caselle, and M. Osella. Stochastic timing in gene expression for simple regulatory strategies. *Nucleic Acids Research*, 45(3):1069–1078, 2017.

[6] M. Del Giudice, S. Bo, S. Grigolon, and C. Bosia. On the role of extrinsic noise in microrna-mediated bimodal gene expression. *PLOS Computational Biology*, 14(4):e1006063, 2018.

[7] M. B. Elowitz, A. J. Levine, E. D. Siggia, and P. S. Swain. Stochastic gene expression in a single cell. *Science*, 297(5584):1183–1186, 2002.

[8] E. Ferro, C. Enrico Bena, S. Grigolon, and C. Bosia. microrna-mediated noise processing in cells: A fight or a game? *Computational and Structural Biotechnology Journal*, 18:642–649, 2020.

[9] D. T. Gillespie. A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. *Journal of Computational Physics*, 22:403–434, 1976.

[10] Daniel T. Gillespie. Exact stochastic simulation of coupled chemical reactions. *The Journal of Physical Chemistry*, 81(25):2340–2361, 1977.

[11] Daniel T. Gillespie. Stochastic simulation of chemical kinetics. *Annual Review of Physical Chemistry*, 58:35–55, 2007.

[12] M. Bilal Hanif. *Biochemistry*. Toronto Academic Press, 1 edition, 2024.

[13] W. A. Haseltine and R. Patarca. The rna revolution in the central molecular biology dogma evolution. *International Journal of Molecular Sciences*, 25(23):12695, 2024.

[14] R. C. Lee, R. L. Feinbaum, and V. Ambros. The c. elegans heterochronic gene lin-4 encodes small rnas with antisense complementarity to lin-14. *Cell*, 75:843–854, 1993.

[15] L. Mesin. *Mathematical models for biomedicine*. Ilmiolibro self publishing, 2017.

[16] S. Mukherji, M. S. Ebert, G. X. Y. Zheng, J. S. Tsang, P. A. Sharp, and A. van Oudenaarden. Micrornas can generate thresholds in target gene expression. *Nature Genetics*, 43(9):854–859, 2011.

[17] J. O'Brien, H. Hayder, Y. Zayed, and C. Peng. Overview of microrna biogenesis, mechanisms of actions, and circulation. *Frontiers in Endocrinology*, 9:402, 2018.

[18] M. Osella, C. Bosia, D. Corá, and M. Caselle. The role of microrna in creating gene expression noise buffers. *PLOS Computational Biology*, 7:e1001101, 2011.

[19] A. E. Pasquinelli, B. J. Reinhart, F. Slack, M. Q. Martindale, M. I. Kuroda, B. Maller, D. C. Hayward, E. E. Ball, B. Degnan, P. Müller, J. Spring, A. Balla, J. Grosshanks, D. D'Evelyn, and G. Ruvkun. Conservation of the sequence and temporal expression of let-7 heterochronic regulatory rna. *Nature*, 408(6809):86–89, 2000.

[20] S. Redner. *A Guide to First-Passage Processes*. Cambridge University Press, 2001.

[21] S. Redner. A first look at first-passage processes, 2023. Testo non pubblicato o preprint.

[22] M. Scott. Tutorial: Genetic circuits and noise. Appunti di lezione, Quantitative Approaches to Gene Regulatory Systems Summer School, University of California, San Diego, 2006.

[23] Velia Siciliano, Immacolata Garzilli, Chiara Fracassi, Stefania Criscuolo, Simona Ventre, and Diego di Bernardo. mirnas confer phenotypic robustness to gene networks by suppressing biological noise. *Nature Communications*, 4:2364, 2013.

[24] N. G. Van Kampen. *Stochastic Processes in Physics and Chemistry*. Elsevier, 3 edition, 2007.

[25] B. Walter, G. Pruessner, and G. Salbreux. Field theory of survival probabilities, extreme values, first passage times, and mean span of non-markovian stochastic processes, 2023.

[26] J. D. Watson and F. H. C. Crick. Molecular structure of nucleic acids: A structure for deoxyribose nucleic acid. *Nature*, 171(4356):737–738, 1953.

[27] Hongyan Zhu and Guo-Chang Fan. Extracellular/circulating micrornas and their potential role in cardiovascular disease. *American Journal of Cardiovascular Disease*, 1(2):138–149, 2011.