Stochastic models of genetic circuits

by

Luis Alberto Gutiérrez López

Submitted to the Department of Physics
in partial fulfillment of the requirements for the degree of

Bachelor of Science in Physics

at the

UNIVERSIDAD DE LOS ANDES

May 12, 2016

Certified by .....
Juan Manuel Pedraza Leal
Associate Professor
Monograph Supervisor

Stochastic models of genetic circuits

by

Luis Alberto Gutiérrez López

Submitted to the Department of Physics
on May 12, 2016 in partial fulfillment of the
requirements for the degree of
Bachelor of Science in Physics

Abstract

All living beings store their genetic information in the DNA and use similar basic mechanisms to read it and, according to its sequence, build their structure and develop their functions. Nevertheless, the information codified on the DNA is not the only aspect that makes an organism what it is. A big and complex network of gene regulation determines which genes are read at a particular moment and the intensity of their activity making it possible, for instance, to differentiate between our muscle cells and our neurons, very different cells but with the same genetic material. Since gene expression is mediated by reactions between molecules, which at microscopic levels happen due to random collisions of the reactants, gene expression and regulation is subjected to noise. A cell also regulates the expression of many genes according to the environment, which may change randomly. In response to these important sources of noise, living beings have developed their regulatory networks to work properly under its presence. This work explores the models that have been done on the last years related to stochasticity in gene expression, the insights they have given to us into the principles of biology and the design of synthetic biological circuits.

Monograph Supervisor: Juan Manuel Pedraza Leal
Title: Associate Professor

Contents

List of Figures

List of Tables

Introduction

Stochasticity, or noise, in biological circuits occurs due to of fluctuations during transcription, translation [1] and other processes that affect gene expression. As a consequence of noise, genetically identical cells and on the same environment may have notorious phenotypical variations [1] [2] [3]. This noise has been classified in two groups: intrinsic and extrinsic [2] [4]. The former is the variability inherent to systems with discrete components and low numbers (e.g. RNA and proteins). The latter is related to external factors as environmental fluctuations, cell growing and cell division.

Recent works have shown the importance of noise for living beings. They have adapted their genetic circuits to develop their respective functions correctly regardless of its presence (robustness) [5], or to take advantage of it to produce variability [6]. Also, when designing synthetic genetic circuits it is important to consider the stochasticity that the circuit may have.

For those reasons, in the last years, several stochastic models of gene expression have been developed. In a pioneer work, Thattai and van Oudenaarden [7] a linearized model for intrinsic noise in the amounts of RNA and proteins that can be applied to some basic motifs. Also, Pedraza and van Oudenaarden [3] developed a model that includes extrinsic noise and showed how fluctuations are propagated through a cascade of regulation.

Most recent models have focused in other aspects that could induce noise. For instance, the bursting in the production of the molecules involved in gene expression, their senescence [8], and the partition of molecules during cell division [9] [10]. One of the most important conclusions of these works is that when considering different factors, the

behavior of noise is similar. Therefore, by studying only the fluctuations it is difficult to know the mechanisms that produce them.

Altought many important results have been made, most of the models used are linearized around the fixed points due to the non-linearity of the equations used to model molecular kinetics. With this, information about the full dynamics of fluctuations. it would be useful then to develop stochastic models that consider the non-linearities, that include the time evolution of noise and that consider more factors like the cell growing and division together with gene expression.

Chapter 1

Preliminary concepts

TODO: Find better images

1.1 The central dogma of molecular biology

The central dogma explains how genetic information flows within a living being. It states that DNA, the molecule that stores the genetic information, is replicated by the enzyme DNA Polymerase. Also, RNA Polymerase produces messenger RNA (mRNA) from DNA in a process called transcription. Finally, the ribosome build the proteins following the sequence of the mRNA and according to a genetic code that translates from the language of nucleotides (the structural blocks of DNA and RNA) to the language of aminoacids, the structural blocks of proteins [11].

Proteins are the structural and functional elementary units of living beings. Therefore, according to the proteins that are being produced in a certain cell it will develop certain functions. The central dogma is summarized in fig. 1.1.1.

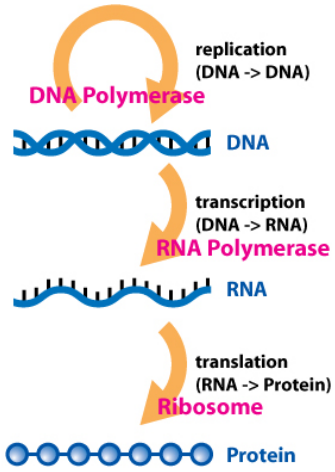

Figure 1.1.1: Scheme of the central dogma of molecular biology. By Dhorspool at en.wikipedia, CC BY-SA 3.0, $3.

An important fact is that the central dogma is valid for all the living beings. The encoding of information in DNA and the mechanisms by which proteins are made according to that information, including the genetic code, are very similar between different organisms.

1.2 Gene regulation and biological circuits

DNA contains all the information necessary to build a living being and let him develop his functions. But genetically identical cells may differ a lot. For example, our neurons are very different in form and function than our skin cells, even though they have the same DNA and thus the same genetic information. This differentiation happens because they are expressing different sets of genes. The genetic information encoded in the DNA is called genotype, while the observable characteristics of an organism are called phenotype. In this terminology, both kinds of cells have the same genotype but differ in their genotypes [11] [12].

Those differences lie on the genes that each cell is expressing at a certain time and how much they are being expressed (measured in the rate of production of proteins corresponding to a gene). There are proteins (and even RNA molecules) that inhibit the

production of other protein by stoping transcription of the corresponding gene. On the contrary, there are proteins that enhance the production of other proteins by increasing the rate of transcription. Both activators and inhibitors are called *transcription factors*, both cases can be seen on fig. 1.2.1.

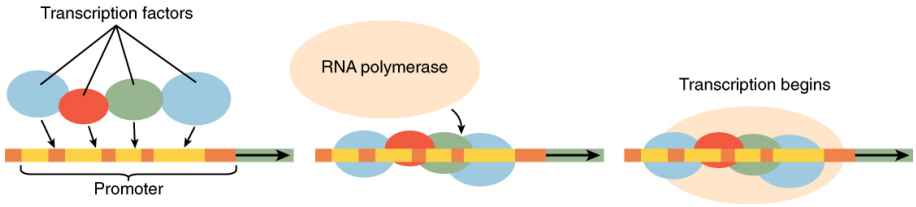

Figure 1.2.1: Scheme of the mechanisms of transcription factors. Retrieved from: `http://biowiki.ucdavis.edu`

The mechanisms of gene regulation can be very complex. For example, a molecule can change the conformation of another protein that when affected by the first, inhibits the transcription of certain gene. Those molecules may be signals from the environment and with mechanisms of this type, the cell process environmental signals to express the optimal genes according to the environment. It is also important to point out that the inhibition and activation is not necessarily done individually. A certain gene may need more than one different protein to enhance its activity, or there may be genes that are activated by a protein and inhibited by others, whose production is in turn mediated by other molecules and transcription factors, a well studied case is the *lac* operon in *E. coli* whose mechanism is explained on fig. 1.2.2.

From the biochemical point of view, transcription factors bind specific sites on the *promoter*, a region of the DNA which is upstream the gene (or set of genes for prokaryotes), and it is where the RNA Polymerase binds to initiate transcription (see fig. 1.2.3). The binding of the transcription factor may enhance or obstruct the binding of RNA Polymerase to the promoter.

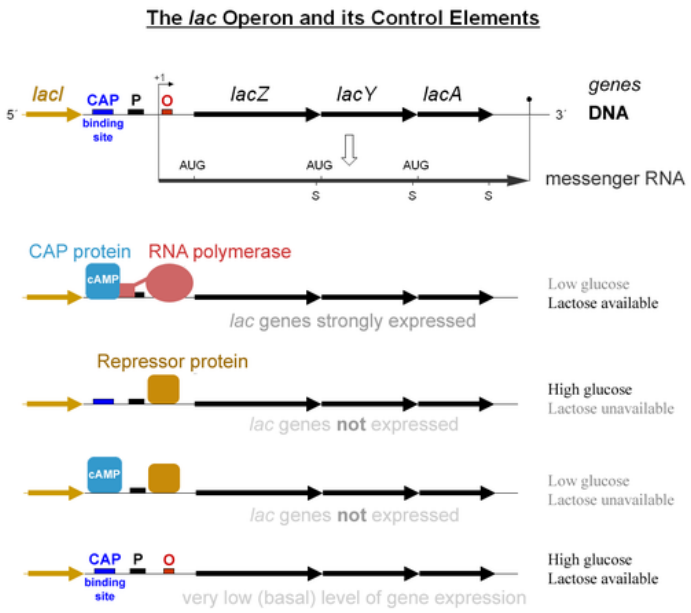

Figure 1.2.2: Example of gene regulation (Lac operon). Retrieved from `upload.wikimedia.org/wikipedia/commons/thumb/d/d2/Lac_operon-2010-21-01.png/550px-Lac_operon-2010-21-01.png`

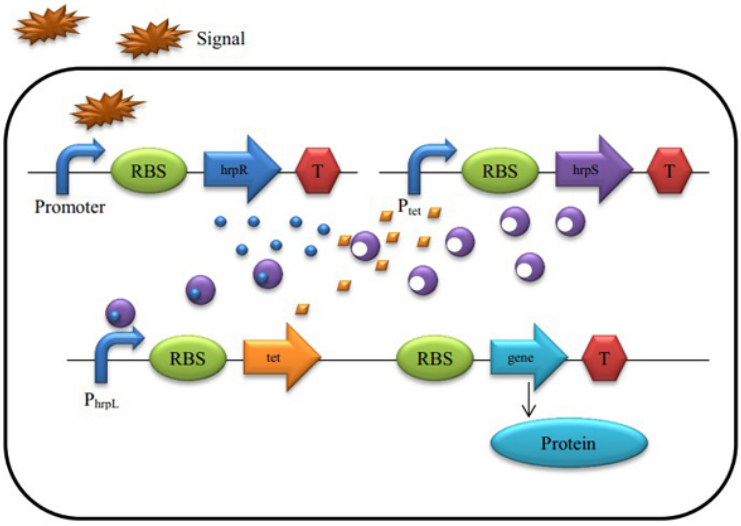

Figure 1.2.3: The promoter, RBS, stop codon are shown. Retrieved from `http://2013.igem.org/wiki/images/c/c6/HIT-Harbin_Project_Schematic.png`

Therefore, in addition to the genotype, gene expression is very important for the cells to develop properly. And, together with the genetic information, defines its structure

and behavior. With this in mind, and the fact that those networks may be very large and complex, the approach that Systems Biology is proposing consists on focusing on the interactions between the different genes and components of a cell rather than on the details of the structure of the molecules involved. The set of interactions may be visualized as biological circuits, that are groups of genes that regulate each other's expression. Figure 1.2.4 shows some of the conventions used in the schemes of biological circuits.

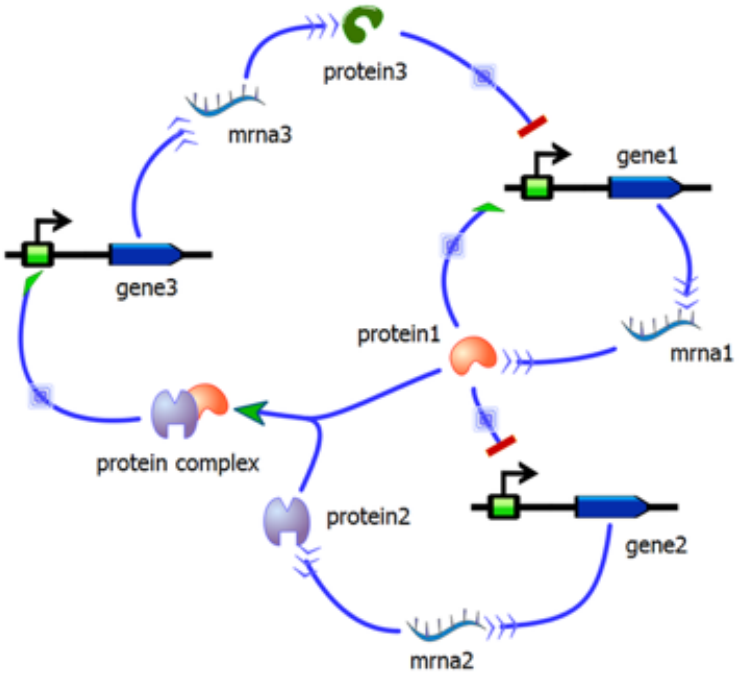

Figure 1.2.4: Typical conventions for biological circuits used in Systems Biology. Retrieved from `http://beacon-center.org/wp-content/uploads/2012/10/SyntheticGeneCircuit.png`

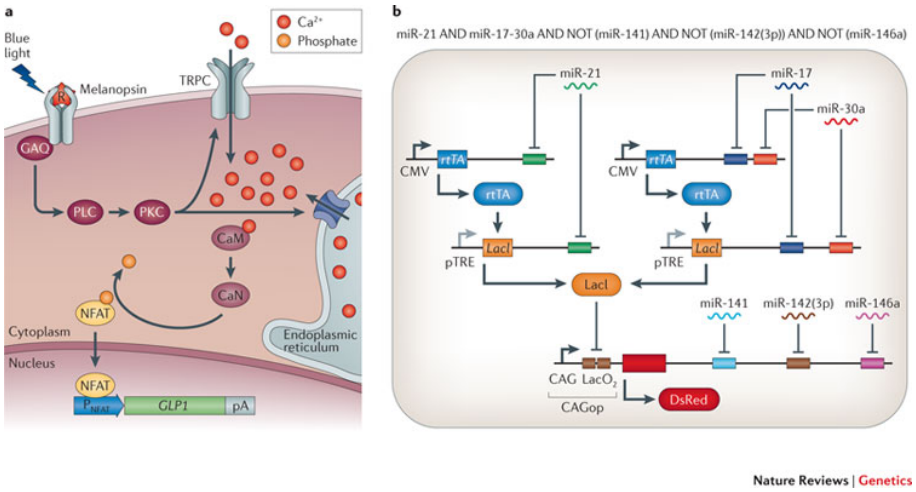

Figure 1.2.5: Example of a biological circuit. Retrieved from `http://www.nature.com/nrg/journal/v13/n6/images/nrg3227-i2.jpg`

1.3 Hill functions

To model the regulation on a gene by a transcription factor, a widely used model is the Hill equation. We will derive it for a particular case that allows a phenomenological understanding of the principles [12].

Consider a transcription factor X that binds to the promoter of some gene, we will label the promoter (gene) as D. Also, suppose that X has n binding sites on the promoter and ignore the intermediate states, where less than n molecules of X are bound. The chemical equation is

$$\mathrm{n[X] + [D] \underset{k_-}{\overset{k_+}{\rightleftharpoons}} [nXD].}$$

Hence, $[nXD]$ changes over time as

$$\frac{\mathrm{d}[nXD]}{\mathrm{d}t} = k_+[X]^n[D] - k_-[nXD],$$

which in steady state yields

$$[X]^n[D] = \frac{k_-}{k_+}[nXD]. \tag{1.3.1}$$

Taking the total number D_T of copies of the gene (promoter and DNA molecules) as a constant we obtain

$$[D_T] = [D] + [nXD].$$

Solving for the free DNA concentration $[D]$ and replacing in eq. (1.3.1)

$$[X]^n\left([D_T] - [nXD]\right) = \frac{k_-}{k_+}[nXD].$$

$[nXD]/[D_T]$ and $[D]/[D_T]$ are the fractions of DNA bound and unbound to the transcription factor, respectively, solving for those quantities we obtain

$$\frac{[nXD]}{[D_T]} = \frac{[X]^n}{K_d^n + [X]^n}, \qquad \frac{[D]}{[D_T]} = \frac{K_d^n}{K_d^n + [X]^n} = \frac{1}{1 + \left(\frac{[X]}{K_d}\right)^n},$$

where $K_d^n \coloneqq k_-/k_+$. In a timescale such that many bindings and unbindings of the transcription factor to the promoter have occurred, those fractions can be interpreted as the probability of having n bound molecules of X, and the probability for being unbound, respectively. If the increasing in transcription rate with respect to the basal rate a is proportional to the probabilities of being bound for an activator, and of being unbound for a repressor, the net rates are

$$f([X]) = a + b\frac{[X]^n}{K_d^n + [X]^n}, \tag{1.3.2}$$

for an activator, and

$$f([X]) = a + b\frac{1}{1 + \left(\frac{[X]}{K_d}\right)^n}. \tag{1.3.3}$$

for a repressor. $b+a$ is the maximum transcription rate, wich happens when $[X] \to \infty$

for the case of an activator, and when $[X] = 0$ for the repressor. K_d is called the *dissociation constant*, which is the concentration of $[X]$ needed for half activation or repression. Biologically it represents the chemical affinity between x and the promoter. n is called the *Hill coefficient* and from the derivation can be said that it is related to the cooperativity of the transcription factor, being larger if the binding of a molecule of $[X]$ enhances more the binding of another one. A larger value of n give a more step-like Hill function. Figures 1.3.1 and 1.3.2 show the shape of typical Hill functions given by eqs. (1.3.2) and (1.3.3).

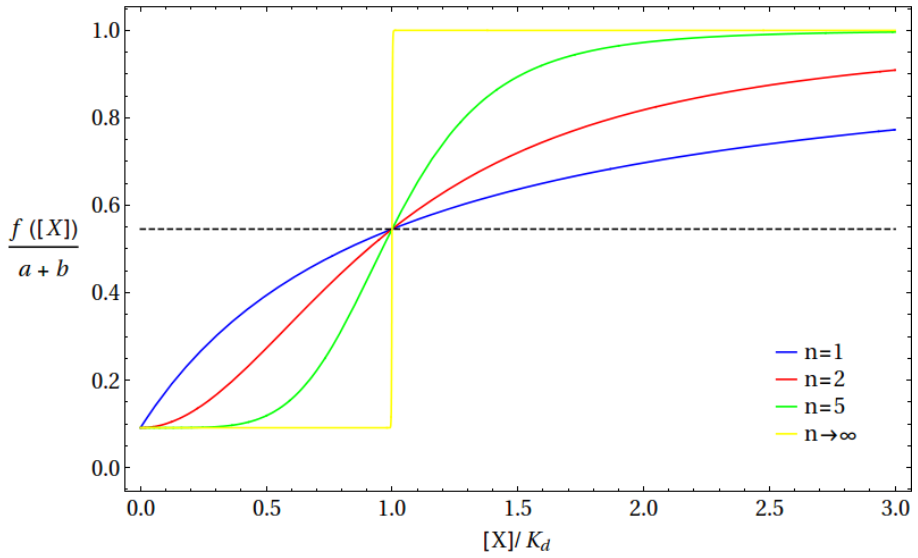

Figure 1.3.1: Hill functions for an activator. Various values of n are shown. The dashed line shows the point of half activation corresponding to $[X] = K_d$. All have the same value of K_d, a and b with $b/a = 10$.

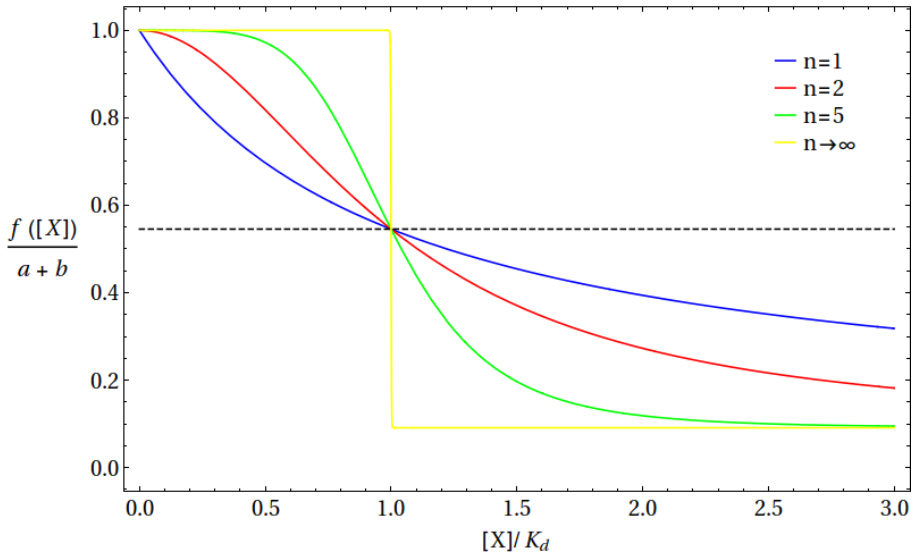

Figure 1.3.2: Hill functions for a repressor. With the same parameters as fig. 1.3.1.

Notice in both graphs that as $n \to \infty$, the function appears more like a Heaviside function with the step in $[X] = K_d$. This approximation can be very useful on a first qualitative analysis of biological circuits but such high values of n are biologically unrealistic. The case $n = 1$ corresponds to the Michaelis-Menten equation.

1.4 Probability

Consider an experiment in which the set of possible outcomes is clearly known. A random variable X is a quantity that can take values from that set of possible outcomes $\{x\}$ of the experiment. How likely is that any value x happens in a given trial of the experiment is determined by the probability mass function (PMF) $P(x)$ if the variable is discrete. For a discrete random variable, $P(x)$ represents the fraction of trials of the experiment in which X has the value x when the number of trials is large. The PMFs follow the axioms of nonnegativity, additivity and normalization. [13] [1]

TODO: Explain CDFs?

[1] We will focus here on discrete random variables. The continous case is very similar, it reduces almost entirely to change $\sum$ by $\int \mathrm{d}x$ and $P(x)$ by $\rho(x)$, where $\rho(x)$ is the probability density function (PDF), the analogous of the PMF.

For several random variables $X_1, \ldots, X_n$, the joint PMF $P(x_1, \ldots, x_n)$ is defined as the probability that $X_1 = x_1, \ldots, X_n = x_n$. The set of random variables are **independent** if

$$P(x_1, \ldots, x_n) = P(x_1) \cdots P(x_n)$$

The conditional probability of the r. v.s $X_1, \ldots, X_k$ given the variables $X_{k+1}, \ldots, X_n$ is denoted by $P(x_1, \ldots, x_k | x_{k+1}, \ldots, x_n)$ and it is defined as

$$P(x_1, \ldots, x_k | x_{k+1}, \ldots, x_n) \coloneqq \frac{P(x_1, \ldots, x_n)}{P(x_{k+1}, \ldots, x_n)},$$

provided that the denominator is different from 0. It can be thought as the probability of a certain outcome for $x_1, \ldots, x_k$ when certain given values of $x_{k+1}, \ldots, x_n$ have been obtained. Notice that if all the random variables are independent it reduces to

$$P(x_1, \ldots, x_k | x_{k+1}, \ldots, x_n) = P(x_1, \ldots, x_k).$$

The conditional and unconditional probabilities are equal, meaning that the outcome of $(x_{k+1}, \ldots, x_n)$ does not affect the outcome of $(x_1, \ldots, x_k)$.

To find the probability of a certain outcome of X, sometimes it is easier to use the **total probability theorem**

$$P(x) = \sum_y P(x, y) = \sum_y P(x|y)P(y).$$

The **expected value** (also called average) of a function of a random variable $f(X)$ is defined as

$$\langle f(X) \rangle \coloneqq \sum_x f(x)P(x). \tag{1.4.1}$$

From the definition can be noticed that the expected value is linear, i.e., for a pair of random variables X and Y, and a constant c

$$\langle X + cY \rangle = \langle X \rangle + c\langle Y \rangle.$$

The variance $\sigma^2(X)$ of X measures the dispersion of the possible outcomes of the random variable, it is defined as

$$\sigma^2(X) \coloneqq \left\langle (X - \langle X \rangle)^2 \right\rangle.$$

It can be easily shown that

$$\sigma^2(X) = \langle X^2 \rangle - \langle X \rangle^2. \tag{1.4.2}$$

From the previous equation it can be seen that for a constant c

$$\sigma^2(cX) = c^2\sigma^2(X) \tag{1.4.3}$$

If X and Y are independent, it can be shown that

$$\langle XY \rangle = \langle X \rangle \langle Y \rangle \quad \text{and} \quad \sigma^2(X + Y) = \sigma^2(X) + \sigma^2(Y). \tag{1.4.4}$$

The conditional expectation of X given a random variable Y, $\langle X|Y \rangle$ is itself a random variable which depends on Y, which when Y is fixed in some y is given by

$$\langle X|y \rangle \coloneqq \sum_x xP(x|y),$$

and it follows that

$$\langle \langle X|Y \rangle \rangle = \langle X \rangle \tag{1.4.5}$$

which is the **law of total expectation**, for the variance there is an analogous theorem, called the **law of total variance**

$$\sigma^2(x) = \left\langle \sigma^2(x|y) \right\rangle + \sigma^2\left(\langle x|y \rangle\right) \tag{1.4.6}$$

The **covariance** of X and Y is defined as

$$\mathrm{cov}(X,Y) \coloneqq \langle (X - \langle X \rangle)(Y - \langle Y \rangle) \rangle = \langle XY \rangle - \langle X \rangle \langle Y \rangle.$$

It can be thought as a measure of how correlated is their behaviour. For example, if the value of Y is known, the value of X will be more likely to be known if the covariance is high in absolute value. The covariance will be 0 if it does not give us any information. Consider the extreme case, if $Y = X$, $\mathrm{cov}(X,X) = \sigma^2(X)$, while if X and Y are independent by eq. (1.4.4) we get $\mathrm{cov}(X,Y) = 0$.

1.5 Noise

Intuitively, we may expect that a random variable is more 'random' or noisy, when the deviations relative to the expected value are bigger. With this in mind, the noise in a random variable X must increase as the variance increases and decrease as the mean increases (the same deviation from a smaller expected value contributes more to the noise than from a bigger one). The quantities that have benn used to measure noise in biology are the Fano factor ν and the coefficient of variation η, which are defined by

$$\nu_X \coloneqq \frac{\sigma^2(X)}{\langle X \rangle}. \tag{1.5.1}$$

$$\eta_X \coloneqq \frac{\sigma(X)}{\langle X \rangle}. \tag{1.5.2}$$

The Fano factor has been used in the first studies of noise in biology since it had the particular property that for a random variable with a Poisson distribution $\nu_X = 1$ and hence it measures deviations from a Poissonian behavior. In more recent studies the coefficient of variation is being used because it is dimensionless and for that reason it

does not depend on the units used for the random variable. For this reason, the generic term 'noise' is now used to refer to η.

1.6 Moment generating functions

Let $n_1, \ldots, n_N$ be discrete random variables over $\mathbb{N}$ and let $f(n_1, \ldots, n_N)$ be the joint probability mass function. The moment generating function $F(z_1, \ldots, z_N)$ is defined as [2]

$$F(z_1, \ldots, z_N) \coloneqq \sum_{n_1=0}^{\infty} \cdots \sum_{n_N=0}^{\infty} z_1^{n_1} \cdots z_N^{n_N} f(n_1, \ldots, n_N). \tag{1.6.1}$$

Evaluating the function on $z_1 = \cdots = z_N = 1$ (denoted by $|_1$) we obtain

$$F|_1 = \sum_{n_1,\ldots,n_N} f(n_1, \ldots, n_N) = 1. \tag{1.6.2}$$

by the axiom of normalization. Taking the derivative of eq. 1.6.1 with respect to z_i, $i = 1, \ldots, N$ we get

$$\left.\frac{\partial F}{\partial z_i}\right|_1 = \sum_{n_1,\ldots,n_N} n_i z_1^{n_1} \cdots z_i^{n_i - 1} \cdots z_N^{n^N} f(n_1, \ldots, n_N) \Bigg|_1 = \sum_{n_1,\ldots,n_N} n_i f(n_1, \ldots, n_N) = \langle n_i \rangle. \tag{1.6.3}$$

Differentiating again with respect to z_j, $j = 1, \ldots, N$ with $j \neq i$ we obtain

$$\left.\frac{\partial F}{\partial z_i z_j}\right|_1 = \langle n_i n_j \rangle. \tag{1.6.4}$$

Differentiating eq. 1.6.1 twice with respect to z_i we obtain similarly

$$\left.\frac{\partial^2 F}{\partial z_i^2}\right|_1 = \langle n_i (n_i - 1) \rangle. \tag{1.6.5}$$

These properties will be very useful in the next sections to find the noise of a genetic

[2] Not to be confused with the cumulative distribution function

system.

1.7 Characteristic function

Another transformation of the PMF that has properties similar to the mentioned for the moment generating function is the characteristic function. It is defined for a N-tuple of random variables $(X_1, \ldots, X_N)$ as[3].

$$\phi(s_1, \ldots, s_N) \coloneqq \left\langle e^{\sum_{i=1}^{N} s_i x_i} \right\rangle = \sum_{x_1} \cdots \sum_{x_N} \exp\left(\sum_{i=1}^{N} s_i x_i \right).$$

We will denote the evaluation at $s_1 = \cdots = s_N = 0$ by $|_0$ It is easy to see that by normalization

$$\phi(s)|_0 = 1.$$

Differentiating once with respect to s_i for $i = 1, \ldots, N$.

$$\left. \frac{\partial \phi(s_1, \ldots, s_N)}{\partial s_i} \right|_0 = \left. \left\langle x_i e^{\sum_{i=1}^{N} s_i x_i} \right\rangle \right|_0 = \langle x_i \rangle. \tag{1.7.1}$$

Each differentiation with respect to s_i produces a factor x_i in the average, hence

$$\left. \frac{\partial^2 \phi(s_1, \ldots, s_N)}{\partial s_i \partial s_j} \right|_0 = \langle x_i x_j \rangle. \tag{1.7.2}$$

This equation is valid for any $i, j = 1, \ldots, N$, even for the case $i = j$. In that case the right hand side becomes $\langle x_i^2 \rangle$. By calculating higher order derivatives we can find higher order moments in the same way.

[3]We will consider here the case with the real exponent

1.8 Stochastic processes

A stochastic process $X(t)$ [4] is a set of random variables indexed by another variable, which usually is taken to be the time. An outcome of the stochastic process is a function of time which varies randomly between different repetitions of the experiment [14] [15].

The **autocorrelation** C_X of a stochastic process $X(t)$ is given by

$$C_X(t, t') \coloneqq \langle X(t)X(t')\rangle.$$

It measures the degree of correlation between outcomes of the random variables at different times. If the process X is **stationary**, the autocorrelation only depends on the time difference, i.e.

$$C(\tau) \coloneqq \langle X(t)X(t+\tau)\rangle,$$

where $\tau \coloneqq t' - t$.

The **power spectrum** S_X of a stochastic process is defined as average of the square norm of its the Fourier transform

$$S_X(\omega) \coloneqq \left|\langle \hat{X}\rangle\right|^2,$$

where the hat denotes Fourier transform. A mathematical tool that will be very useful in the calculations is the **Wiener-Khinchin theorem** It states that the power spectrum and the autocorrelation are Fourier-Transform pairs, that is

$$\mathscr{F}(C_X(\tau)) = S_X(\omega), \quad \text{and} \quad \mathscr{F}^{-1}(S_X(\omega)) = C_X(\tau). \tag{1.8.1}$$

[4] or $\{X\}_n$ if the time steps are discrete

1.9 The Poisson process

Many of the processes of creation and destruction that occur in this context are modeled as Poisson processes. For example, the creation and destruction of mRNA and proteins.The Poisson process is a continous-time stochastic process that is used to model arrivals when there is some known arrival rate and when the arrivals at different time intervals are independent.

Mathematically, we define $P(k, \tau)$ as the probability that there are $k \in \mathbb{Z}$ arrivals during a time interval τ and assume that it is the same for any interval of the same length. We denote by λ the arrival rate for the process. The Poisson process satisfy the following properties

- $P(k, \tau)$ is the same for all intervals of length τ.
- The value of $P(k, \tau)$ during some particular interval is independent of other intervals.
- $P(k, \tau)$ satisfies the following

$$\begin{aligned} P(0, \tau) &= 1 - \lambda\tau + o_0(\tau), \\ P(1, \tau) &= \lambda\tau + o_1(\tau), \\ P(k, \tau) &= o_k(\tau). \quad \text{for } k > 1. \end{aligned}$$

  Where $o_k(\tau)$, $k = 0, 1, \ldots$ are functions that become negligible compared to τ as it becomes small, that is

$$\lim_{\tau \to 0} \frac{o_k(\tau)}{\tau} = 0.$$

It can be proven that according to the previous properties, $P(k, \tau)$ is given by

Prove it? Is that really true?

$$P(k,\tau) = e^{-\lambda\tau}\frac{(\lambda\tau)^k}{k!}, \quad k = 0, 1, \ldots$$

which is a Poisson distribution, using the definition for the expected value and variance (eqs. (1.4.2) and (1.4.1)), letting N_τ be the number of arrivals during a time interval τ we get after a little algebra

$$\langle N_\tau \rangle = \sigma^2(N_\tau) = \lambda\tau.$$

The average number of arrivals in a time τ is, as intuitively expected, the arrival rate times the length of the interval. From eqs. (1.5.1) and (1.5.2), the noise and Fano factor for N_τ are

$$\nu(N_\tau) = 1, \qquad \eta(N_\tau) = \frac{1}{\sqrt{N_\tau}}. \tag{1.9.1}$$

This proves what was said on section 1.5 about the previous use of the Fano factor as the standard measure of noise. From the form of η we conclude that if the number of events is large, the Poisson noise is negligible. But for a biological system, for instance, the average number of copies of mRNA of a given gene is of the order of 10. This makes the effect of noise a significant factor. Also, altought the mean number of proteins is of the order of 1000, the Poisson noise is negligible, but there is an important contribution of noise coming from the RNA and other sources, as we will see on chapter 2.

Now we will find the time between events, by now let $t = 0$ the time of the last event and let $t = T$ the time of the next event, then

$$P(T \leq t) = \int_0^t f_T(t')dt' = 1 - P(T > t) = 1 - P(0,t) = 1 - e^{-\lambda t}.$$

Differentiating and applying the fundamental theorem of calculus, we get the **exponential PDF**, which is given by

$$f_T(t) = \lambda e^{-\lambda t}.$$

Therefore, the time T until the first arrival follows an exponential distribution.

The exponential distribution is **memoryless** in the following sense: suppose an arrival happened at time t', therefore, the probability distribution for the remaining time until the next arrival is an exponential with the same rate.

Explain better, prove it?

. With this in mind, not only the time until the first arrival, but all the interarrival times (the times between arrivals), follow the distribution given by eq. (1.9).

Suppose we k independent Poisson processes with rates $\lambda_1, \ldots, \lambda_k$ and we record an arrival each time an arrival occur in either process. This merged process is also a Poisson process with rate $\Lambda := \sum_{i=1}^{k} \lambda_i$. Also, any arrival of the merged process has a probability λ_i/Λ, $i = 1, \ldots, k$ of being an arrival of the ith process.

In the models we will consider there might be several creation and destruction events which are Poissonian and independent, for example, the synthesis and degradation of different kinds of RNA and protein, the binding of transcription factors, etc. For these models, we can take advantage of the merging properties of Poisson processes to make efficient and precise simulations.

1.10 The Gillespie algorithm

To simulate the models we will use the Gillespie algorithm [16] which improves speed a lot with respect to brute-force stochastic algorithms. It is used to simulate simultaneous Poissonian events that occur with a certain rate (probability per unit time), e.g. synthesis and degradation of RNA or proteins, binding of an enzyme to a substrate, etc.

In the brute-force approach we consider a fixed time interval that must be sufficiently small, and for every possible event we sample a random number that depending on the probabilities of the events, will tell us which of the events happened or if nothing

happened at that interval. This procedure is repeated for all the intervals. Since time intervals must be small and for each interval we must sample as many random numbers as events, this approach is computationally inefficient.

In the Gillespie algorithm, we take advantage of the mentioned properties of the Poisson process [13]. There are not fixed time intervals in this case, a random number is sampled with an exponential distribution whose rate Λ is the sum of the rates of the individual processes to find the time of occurence of the next event and another uniform one is sampled to evaluate which of the events occured. Using the properties of the exponential CDF, to sample an exponential random number X with parameter Λ from an uniform U between 0 and 1 we apply the following equation

Prove it?

$$X = -\frac{1}{\Lambda}\ln(U). \tag{1.10.1}$$

The Gillespie algorithm is by far more efficient. First, it does not need to have a fixed small time interval, it finds the time intervals between events. Second, the number of random numbers sampled per time interval in the brute-force approach is equal to the number of events (which can be large), and many of the intervals there will not be an event, while in the Gillespie algorithm only two random numbers are sampled by event and the way in which the time interval is sampled allows to go forward in time in many fewer steps. Finally, a very important aspect is that the Gillespie algorithm is that it is exact, while the precision of the brute-force algorithm depends on how small is the time interval in consideration.

Chapter 2

Basic genetic circuits in steady state - The master equation

In this chapter, we develop a model of a genetic system considering only the intrinsic noise using the master equation (ME). This is an approach where there are certain states in which the system can be found and it evolves probabilistically between those states according to some determined transition rates. In the context of genetic circuits the possible states are characterized by number of mRNAs and proteins and the transitions by their rates of creation and degradation.

We introduce the ME approach applying it to the simplest possible genetic circuit: a single gene with regulated by a constitutive promoter. Next we explain a generalization of the analytic method that can be applied to an arbitrary genetic circuit that satisfies certain conditions. Finally, we ilustrate the general method using it on an negatively autorregulated gene.

Define constitutive promoter

This chapter is based on the work done by M. Thattai and A. van Oudenaarden in [7].

2.1 Single gene

Consider the processes of transcription and translation of a certain gene. Let the number of mRNA molecules and proteins be n_1 and n_2, respectively. The assumptions of the model are the following: the number d of copies of certain gene is a constant, the rate k_r of synthesis of mRNA per copy of the gene is a constant. In the same way, the rate k_p of production of proteins per mRNA molecule is a constant. There is also a degradation rate for each molecule proportional to its concentration labeled as γ_r and γ_p for mRNAs and proteins,respectively. The model is schematized in fig. 2.1.1. The mentioned assumptions hold for many genetic systems.

Cite something here

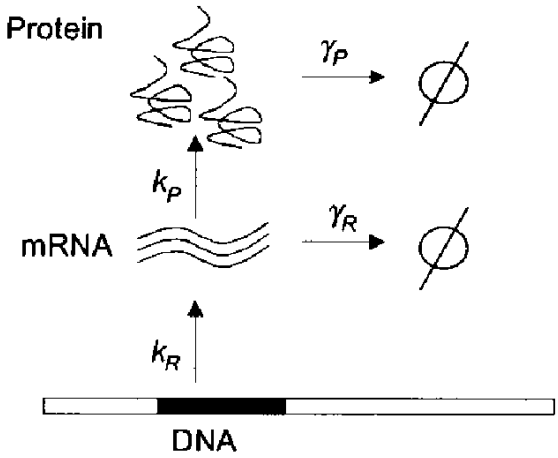

Figure 2.1.1: Steps of gene expression considered in the model. Taken from [7].

According to the assumptions, the deterministic equations for n_1 and n_2 are

$$\dot{n_1}(t) = k_r d - \gamma_r n_1(t), \tag{2.1.1}$$

$$\dot{n_2}(t) = k_p n_1(t) - \gamma_p n_2(t). \tag{2.1.2}$$

Hence, on steady state

Explain steady state on concepts, what does it mean?

$$n_{1,s} = \frac{k_r}{\gamma_r}, \tag{2.1.3}$$

$$n_{2,s} = \frac{k_p}{\gamma_p} n_{1,s} = \frac{k_p k_r}{\gamma_p \gamma_r}. \tag{2.1.4}$$

Let $n_1(0) = n_2(0) = 0$. Then, the solutions of the differential equation for n_1 is

$$n_1(t) = n_{1,s} \left(1 - e^{-\gamma_r t}\right).$$

Also, assuming n_1 is fixed at some value n_1^*, the solution for n_2 is similar

$$n_2(t) = n_{2,s} \left(1 - e^{-\gamma_r t}\right),$$

with $n_{2,s} = \frac{k_p}{\gamma_p} n_1^*$. Figure 2.1.2 shows a plot of the solution.

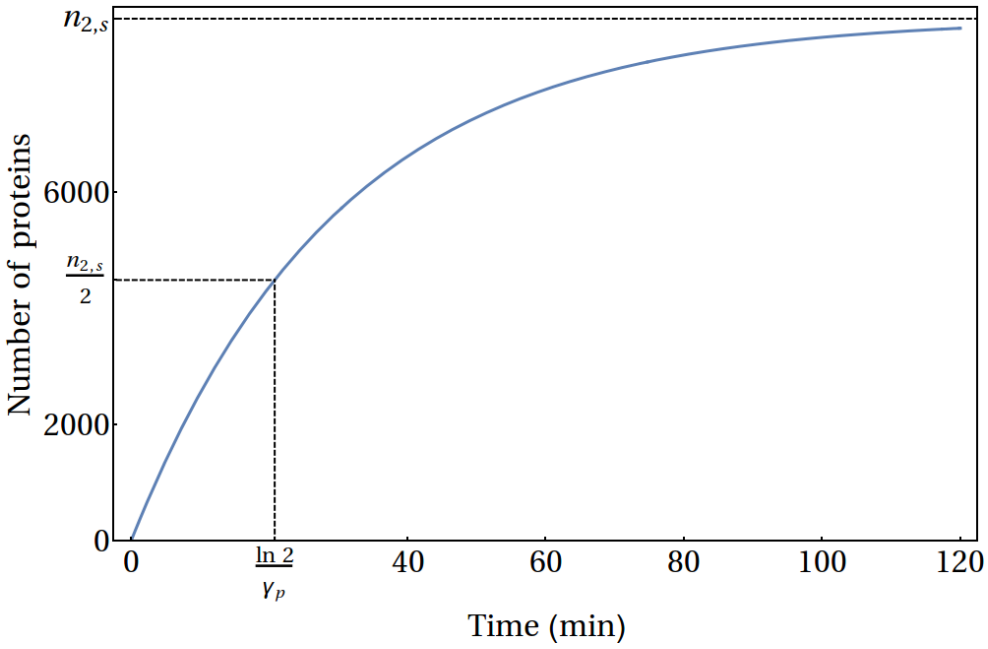

Figure 2.1.2: Deterministic solution for the number of proteins as a function of time for a fixed number of mRNAs. The time for reaching half the steady state value, $\ln 2/\gamma_p$, and the steady state value are marked with dashed lines.

The number of protein reaches half its steady state in a time of $\ln 2/\gamma_p$. Then, the speed at which n_2 reaches its steady state is proportional to $1/\gamma_p$ and it is the same independently of whether the levels start above or below $n_{2,s}$. If the number of mRNA molecules changes, the steady state level for n_2 also changes and the proteins reach the new steady state within a timescale of $1/\gamma_p$. This fact is very important in noise propagation as we will see in the next chapter.

A parameters that living beings might want to tune in their genetic circuits is their response time. In an unregulated gene γ_p must be increased to have a faster response time. To do that, proteins must be actively destroyed and this raises the energetic cost. However, there are more sophisticated mechanisms to control the response time such as feed forward loops and autorregulation [12].

The solutions of the differential equations (2.1.1) and (2.1.2) are continous functions of time that are uniquely determined for some given initial conditions. However, in the reality n_1 and n_2 are discrete numbers. This fact determines an important part of their

stochastic behaviour. Also, the synthesis and degradation of the molecules are, at a fundamental level, a consequence of random collisions of molecules that are difussing through the cell (e.g. mRNA polymerases with the promoter and enzymes with their substrates).

With this in mind, to account for the noise the quantities $n_1(t)$ and $n_2(t)$ are defined as stochastic processes instead of deterministic functions. Also, they only take values on $\mathbb{N}$ instead of $\mathbb{R}^+$. In the ME aproach, the system can be thought as a set of states labeled with all the possible pairs (n_1, n_2). The rates of synthesis and degradation are thought as the probabilities per unit time of a transition between states occuring. Figure 2.1.3 shows the possible transitions from and into the state (n_1, n_2).

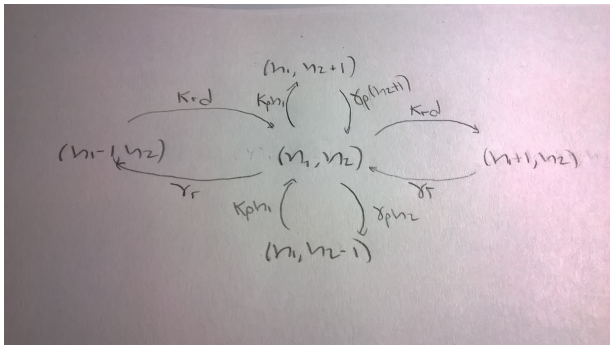

Figure 2.1.3: Scheme of the possible transitions involving n_1 RNA molecules and n_2 protein molecules.

The transitions shown in figure 2.1.3 can be interpreted as follows: there is a probability $p(n_1, n_2, t)$ of being at the state (n_1, n_2) at time t which changes according to the probabilities of being in the adjacent states and the transition probabilities given by the reaction rates. Also, it is assumed that the probability of being at a state is independent of the transition probabilities. Formally speaking, applying the master equation approach implies that the processes are Markovian.

Therefore, to completely characterize the system under this scheme, the PMF $p(n_1, n_2, t)$ must be found for all possible n_1, n_2 and t. We can write a difference-differential equation for p called the Master Equation, it is given by

$$\begin{aligned}\frac{\mathrm{d}p(n_1,n_2,t)}{\mathrm{d}t} &= k_r dp(n_1-1,n_2,t) - k_r dp(n_1,n_2,t)\\ &+ k_p n_1 p(n_1,n_2-1,t) - k_p n_1 p(n_1,n_2,t)\\ &+ \gamma_r(n_1+1)p(n_1+1,n_2,t) - \gamma_r n_1 p(n_1,n_2,t)\\ &+ \gamma_p(n_2+1)p(n_1,n_2+1,t) - \gamma_p n_2 p(n_1,n_2+1,t).\end{aligned} \tag{2.1.5}$$

The first term refers to a transition from state (n_1-1,n_2,t) to (n_1,n_2,t) via a creation of a mRNA molecule, whereas the second term involves a transition $(n_1,n_2,t)\rightarrow(n_1+1,n_2,t)$ via mRNA synthesis. The third and fourth terms have the meaning but related to protein synthesis. The other terms are related to transitions due to degradation.

Define moments

The PMF can be found from the ME for certain systems. Nevertheless, to find the noise we only need the first and second moments. To do that we write the master equation in terms of the moment generating function $F(z_1,z_2)$, defined in eq. (1.6.1). Multiplying by $z_1^{n_1}z_2^{n_2}$ and summing over n_1 and n_2, both from 0 to ∞ we obtain for the left hand side simply $\dot{F}(z_1,z_2)$. For the first term on the right hand side we obtain [1]

$$\sum_{n_1=0,n_2=0}^{\infty} z_1^{n_1}z_2^{n_2} f(n_1-1,n_2) = \sum_{n_1=-1,n_2=0}^{\infty} z_1^{n_1+1}z_2^{n_2} f(n_1,n_2),$$

but since n_1 represents number of molecules, it must be a positive quantity. Hence $f(-1,n_2)=0$ and the last sum can be taken from $n_1=0$ yielding

$$z_1\sum_{n_1=0,n_2=0}^{\infty} z_1^{n_1}z_2^{n_2} f(n_1,n_2) = z_1F(z_1,z_2).$$

For the second term of eq. (2.1.5) the result is trivial, for the third term we get

$$\sum_{n_1=0,n_2=0}^{\infty} n_1 f(n_1,n_2-1) = \sum_{n_1=0,n_2=-1}^{\infty} n_1 z_1^{n_1}z_2^{n_2+1} f(n_1,n_2).$$

[1] The time dependence is not shown for simplicity.

Using the same argument as above, $f(n_1, -1) = 0$. Rearranging it becomes

$$z_1 z_2 \sum_{n_1=0,n_2=0}^{\infty} z_1^{n_1-1} z_2^{n_2} f(n_1, n_2) = z_1 z_2 \frac{\partial F(z_1, z_2)}{\partial z_1}.$$

For the fifth term

$$\begin{aligned}\sum_{n_1=0,n_2=0}^{\infty} (n_1+1) z_1^{n_1} z_2^{n_2} f(n_1+1, n_2) &= \sum_{n_1=1,n_2=0}^{\infty} n_1 z_1^{n_1-1} z_2^{n_2} f(n_1, n_2) \\ &= z_1 \sum_{n_1=0,n_2=0}^{\infty} n_1 z_1^{n_1-1} z_2^{n_2} f(n_1, n_2) = \frac{\partial F(z_1, z_2)}{\partial z_1}.\end{aligned}$$

The other terms are treated in a similar fashion. Putting all of this together in we obtain the master equation in terms of the moment generating function F

$$\begin{aligned}\dot{F}(z_1, z_2, t) = k_r d(z_1 - 1) F(z_1, z_2, t) + k_p z_1 (z_2 - 1) \frac{\partial F(z_1, z_2, t)}{\partial z_1} \\ + \gamma_r (1 - z_1) \frac{\partial F(z_1, z_2, t)}{\partial z_1} + \gamma_p (1 - z_2) \frac{\partial F(z_1, z_2, t)}{\partial z_2}.\end{aligned} \tag{2.1.6}$$

We transformed a difference equation in (n_1, n_2) into a partial differential equation in (z_1, z_2). The PMF p can be found by solving for F and transforming back. However, in this case we will use the properties of F (eqs. (1.6.2) - (1.6.5)) to find only the first two moments. Differentiating with respect to z_1

$$\begin{aligned}\frac{\partial \dot{F}}{\partial z_1} = k_r d \left(F + (z - 1) \frac{\partial F}{\partial z_1} \right) + k_p (z_2 - 1) \left(\frac{\partial F}{\partial z_1} + z_1 \frac{\partial^2 F}{\partial z_1^2} \right) \\ + \gamma_r \left(-\frac{\partial F}{\partial z_1} + (1 - z_1) \frac{\partial^2 F}{\partial z_1^2} \right) + \gamma_p (1 - z_2) \frac{\partial^2 F}{\partial z_1 \partial z_2},\end{aligned} \tag{2.1.7}$$

and with respect to z_2

$$\begin{aligned}\frac{\partial \dot{F}}{\partial z_2} &= k_r d(z_1-1)\frac{\partial F}{\partial z_2} + k_p z_1\left(\frac{\partial F}{\partial z_1} + (z_2-1)\frac{\partial^2 F}{\partial z_1 \partial z_2}\right) \\ &+ \gamma_r(1-z_1)\frac{\partial^2 F}{\partial z_1 \partial z_2} + \gamma_p\left(-\frac{\partial F}{\partial z_2} + (1-z_2)\frac{\partial^2 F}{\partial z_2^2}\right).\end{aligned} \tag{2.1.8}$$

Evaluating eqs. (2.1.7) and (2.1.8) at $z_1 = z_2 = 1$ and using properties (1.6.2) and (1.6.3) we obtain

$$\begin{aligned}\dot{\langle n_1 \rangle} &= k_r d - \gamma_r \langle n_1 \rangle, \\ \dot{\langle n_2 \rangle} &= k_p \langle n_1 \rangle - \gamma_p \langle n_2 \rangle.\end{aligned}$$

Which are the same as eqs. (2.1.1) and (2.1.2). Therefore, the averages follow the deterministic behavior and the steady state values are thus given by eqs. (2.1.3) and (2.1.4).

Differentiating eq. (2.1.7) with respect to z_2, eq. (2.1.7) with respect to z_1 and eq. (2.1.8) with respect to z_2 and evaluating at $z_1 = z_2 = 1$ we obtain, respectively

$$\dot{\langle n_1 n_2 \rangle} = k_r d\langle n_2 \rangle + k_p\left(\langle n_1 \rangle + \langle n_1(n_1-1)\rangle\right) - (\gamma_r + \gamma_p)\langle n_1 n_2 \rangle, \tag{2.1.9}$$

$$\dot{\langle n_1(n_1-1)\rangle} = 2k_r\langle n_1 \rangle - 2\gamma_r\langle n_1(n_1-1)\rangle, \tag{2.1.10}$$

$$\dot{\langle n_2(n_2-1)\rangle} = 2k_p\langle n_1 n_2 \rangle - 2\gamma_p\langle n_2(n_2-1)\rangle. \tag{2.1.11}$$

The previous equations will be treated in steady state. From eq. (2.1.10)

$$0 = k_r d\langle n_1 \rangle_s - \gamma_r\left(\langle n_1^2 \rangle_s - \langle n_1 \rangle_s\right) \Rightarrow \langle n_1^2 \rangle_s = \frac{k_r d}{\gamma_r}\langle n_1 \rangle_s + \langle n_1 \rangle_s = \langle n_1 \rangle_s^2 + \langle n_1 \rangle_s. \tag{2.1.12}$$

Therefore, in steady state $\sigma_1^2 = \langle n_1 \rangle$. Hence, the Fano factor and the squared CV for

the mRNA are given by

$$\nu_1 = 1, \qquad \eta_1^2 = \frac{1}{\langle n_1 \rangle}. \tag{2.1.13}$$

Which is the noise for a Poisson process as we saw on eq. (1.9.1). This makes sense since the assumptions made for the mRNA dynamics correspond to the ones made for the Poisson process in section 1.9. In *E. coli*, $\langle n_1 \rangle_s$ 5 mRNA molecules. Since the number is low, the mRNA number present large fluctuations.

From eq. (2.1.9) we have

$$0 = k_r d \langle n_2 \rangle_s + k_p \langle n_1^2 \rangle_s - (\gamma_p + \gamma_r) \langle n_1 n_2 \rangle_s \Rightarrow \langle n_1 n_2 \rangle_s = \frac{k_r d \langle n_2 \rangle_s + k_p \langle n_1^2 \rangle_s}{\gamma_r + \gamma_p}.$$

But from eq. (2.1.12) and (2.1.4),

$$\langle n_1^2 \rangle_s = \langle n_1 \rangle_s \left(\langle n_1 \rangle_s + 1 \right) = \frac{\gamma_p}{k_p} \langle n_2 \rangle_s \left(\langle n_1 \rangle_s + 1 \right). \tag{2.1.14}$$

Hence, the covariance is given by

$$\begin{aligned} \langle n_1 n_2 \rangle_s - \langle n_1 \rangle_s \langle n_2 \rangle_s &= \langle n_2 \rangle_s \left(\frac{k_r d + \gamma_p \left(\langle n_1 \rangle_s + 1 \right)}{\gamma_r + \gamma_p} - \langle n_1 \rangle_s \right) \\ &= \langle n_2 \rangle_s \frac{k_r d + \gamma_p - \gamma_r \langle n_1 \rangle_s}{\gamma_r + \gamma_p}. \end{aligned}$$

From eq. (2.1.3) the first and third term of the numerator cancel out, therefore

$$\boxed{\mathrm{cov}(n_1, n_2)_s = \langle n_2 \rangle_s \frac{\frac{\gamma_p}{\gamma_r}}{1 + \frac{\gamma_p}{\gamma_r}}}. \tag{2.1.15}$$

From eq. 2.1.11 we have

$$kp \langle n_1 n_2 \rangle_s = \gamma_p \langle n_2^2 \rangle_s - \gamma_p \langle n_2 \rangle_s$$

Replacing eq. 2.1.15 in the previous equation we get after rearranging

$$\begin{aligned}\langle n_2^2\rangle_s &= \frac{k_p}{\gamma_p}\left(\langle n_1\rangle_s\langle n_2\rangle_s + \frac{\langle n_2\rangle_s\gamma_p}{\gamma_r+\gamma_p}\right) + \langle n_2\rangle_s \\ &= \langle n_2^2\rangle_s + \frac{k_p\langle n_2\rangle_s}{\gamma_r+\gamma_p} + \langle n_2\rangle_s.\end{aligned}$$

Hence substracting $\langle n_2\rangle_s^2$ from the previous equation we obtain

$$\sigma_2^2 = \langle n_2\rangle\left(\frac{k_p/\gamma_r}{1+\gamma_p/\gamma_r}+1\right).$$

Therefore, the noise for the proteins in steady state is given by

$$\nu_2 = \frac{b}{1+\gamma_p/\gamma_r}+1, \qquad \eta_2^2 = \frac{1}{\langle n_2\rangle}\left(\frac{b}{1+\gamma_p/\gamma_r}+1\right). \tag{2.1.16}$$

Explain something about orders of magnitude

where $b \coloneqq k_p/\gamma_r$ is the average number of proteins that are produced during the lifetime of a transcript, often called the *burst size*. The second part (1 in ν_p) is the Poisson noise, which measured as the CV is negligible because the average number of proteins for a given gene is usually large (of the order of 1000 to 10000 in *E. coli*).

The first part is the contribution to the noise that arises in mRNA fluctuations and is propagated to the proteins since they are related by the rate of production of proteins $k_p n_1$. The noise in the number of proteins is thus much larger than Poissonian since b is about 15 proteins/mRNA. The factor $1/(1+\gamma_p/\gamma_r)$ represent the so called time averaging of fluctuations, it takes values between 0 and 1. Usually $\gamma_p < \gamma_r$, meaning that fluctuations in mRNA number happen on a characteristic timescale $\tau_r \propto 1/\gamma_r$ that is smaller than the response time of the proteins $\tau_p \propto 1/\gamma_p$. The proteins are thus unable to respond immediately to the mRNA fluctuations. As a consequence, the protein make a time average over many ups and downs of mRNA. This process reduces fluctuations in the number of proteins. This very important phenomenon will be considered with more

detain in section 3.3.

The analytical results are compared with the results of Gillespie simulations in fig. 2.1.4. It can be seen that the Fano factor is strongly dependent on the burst size b, independent of k_r and dependent on the protein half-life $\tau_p = \ln 2/\gamma_r$ only for low values of $\langle p \rangle$. Also, ν_p can become much larger than 1 for biologically relevant parameters, verifying that the noise of proteins is much larger than Poisson.

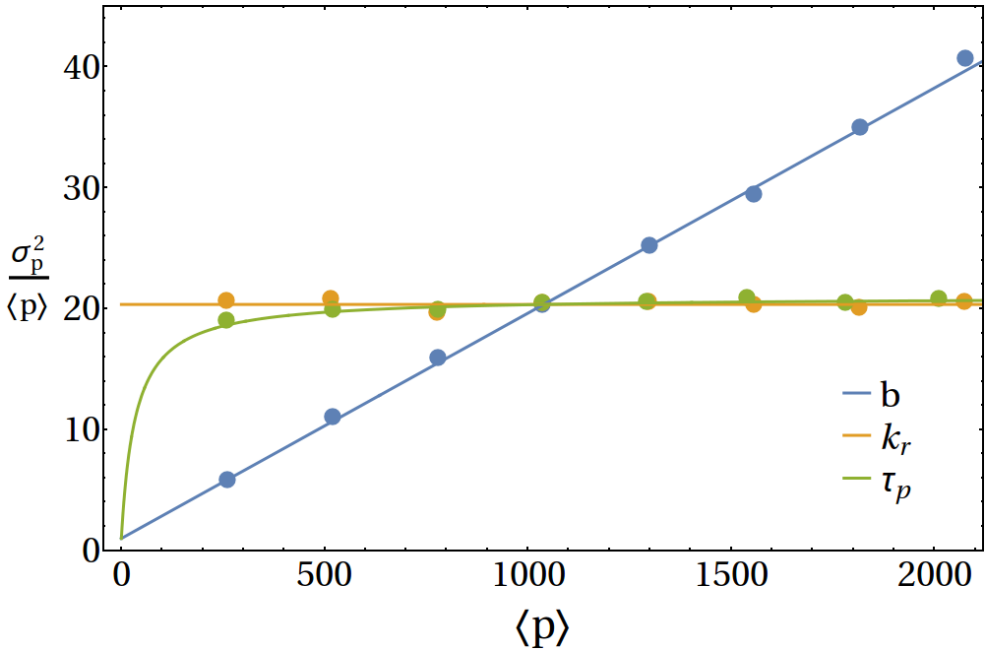

Figure 2.1.4: Comparison between the results of Gillespie simulations (dots) and the analytical results (lines) given by eq. (2.1.16). The Fano factor is plotted vs. the mean number of proteins in steady state. The base values of the parameters are $\tau_r = \ln 2/\gamma_r = 2$ min, $b = 20$ proteins/mRNA $k_r = 0.01$ mRNA/s and $\tau_p = \ln 2/\gamma_p = 1$ h. For each curve, the parameter indicated in the legend is varied while the others are fixed. b is varied between 5 and 40 proteins/mRNA; k_r varies between 0.0025 and 0.02 mRNA/s; and τ_p varies from 15 min to 2 h. Each point correspond to 10000 trials where each one was evolved until a time of $10\tau_p$ in order to be sure that the system has reached its steady state [3].

Define the thing with kr and krd

The noise and the steady state average in protein numbers can be controlled inde-

[3]The simulations were programmed in C and the graphics were done with Wolfram Mathematica. The scripts can be found here.

pendently by controlling the burst size b. If a cell produces many mRNAs and a few proteins per transcript (small b), the noise is reduced. On the contrary, the same average number of proteins can be reached by producing a few mRNAs and many proteins per mRNA (large b). In this case the noise is larger. Nevertheless, reducing noise in this case requires a constant synthesis and degradation of many mRNAs. This is inefficient for the cells since they are spending energy in the production of mRNAs from which there will be only a few proteins translated. This suggests that there is an interchange between fitness and noise reduction in the cells that has been tuned by evolution according to the necessity of reliability of the particular genetic component. However, we will see that there are other mechanisms, like negative autorregulation (sec. 2.3), that allow cells to reduce noise in a more efficient way.

Define fitness, word for pay-off

2.2 Several species with linear interactions

In this section we generalize the previous results to arbitrary genetic network in which the interactions between its components are linear. Consider eqs. (2.1.1) and (2.1.2). In matrix notation, they can be written as

$$\dot{\mathbf{n}} = (\mathbf{A} - \mathbf{\Gamma})\,\mathbf{n}, \tag{2.2.1}$$

where $\mathbf{n}^T = (d, n_1, n_2)$ is the vector of chemical species and the matrices $\mathbf{A}$ and $\mathbf{\Gamma}$ are defined as

Try to write eqs. on the same line

$$\mathbf{A} = \begin{pmatrix} 0 & 0 & 0 \\ k_r & 0 & 0 \\ 0 & k_p & 0 \end{pmatrix}, \tag{2.2.2}$$

$$\mathbf{\Gamma} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & \gamma_r & 0 \\ 0 & 0 & \gamma_p \end{pmatrix}. \tag{2.2.3}$$

A contains the creation rates and describes how each rate depends on the different species and **Γ** has the degradation rates, which is diagonal whenever the degradation is not mediated by interactions among different kinds of molecules.

For an arbitrary circuit with an arbitrary number N of species and linear interaction between them, we can write its deterministic equations in the form of eq. (2.2.1). In the ME approach for genetic circuits, the state space has now N dimensions and the ME has now $4N$ terms. This is because for each kind of molecule there are two transitions given by synthesis and two given by degradation as can be seen in fig. 2.1.3. With this in mind, the ME is given by [4]

$$\dot{p}(n_i) = \sum_{i=1}^{N} \left(\sum_{j=1}^{N} A_{ij} n_j \left(p(n_i - 1) - p(n_i) \right) + \sum_{j=1}^{N} \Gamma_{ij} ((n_j + 1) p(n_i + 1) - n_j p(n_i)) \right) \tag{2.2.4}$$

When for a fixed specie i, the terms in parentheses represent having $n_i - 1$ molecules of the ith type and creating one; having n_i and creating one; having $n_i + 1$ and destroying one; and having n_i and destroying one, respectively. Since this is possible for each of the N types of molecules, we must sum over i as well.

Assuming that degradation does not involve interactions between various molecules,

[4] To reduce the complexity of the expressions, we define $p(n_i) \coloneqq p(\mathbf{n}) = p(n_1, \ldots, n_i, \ldots, n_N)$, and $p(n_i \pm 1) \coloneqq p(n_1, \ldots, n_i \pm 1, \ldots, n_N)$. The same convention is used for $F(z_i)$.

the matrix Γ can be taken to be diagonal, i.e. $\Gamma_{ij} = \delta_{ij}\Gamma_j$. With this eq. (2.2.4) becomes

$$\dot{p}(n_i) = \sum_{i=1}^{N} \left(\sum_{j=1}^{N} A_{ij} n_j \left(p(n_i - 1) - p(n_i) \right) + \Gamma_i((n_i + 1)p(n_i + 1) - n_i p(n_i)) \right). \quad (2.2.5)$$

To find the mean and variances, we write the master equation in terms of the moment generating function and use its properties to find equations for the moments. In this case we multiply by $z_1^{n_1} \cdots z_N^{n_N}$ and sum over $n_1, \ldots n_N$, all from 0 to ∞. First we will consider the term in parentheses of eq. (2.2.5) and later we will perform the outer sum over i. For the first term we obtain the following expression [5]

$$\sum n_j z_1^{n_1} \cdots z_i^{n_i} \cdots z_j^{n_j} \cdots z_N^{n_N} f(n_i) = z_i z_j \sum n_j z_j^{n_j - 1} z_i^{n_i} f(n_i) = z_i z_j \frac{\partial F}{\partial z_j}. \quad (2.2.6)$$

Where the index switching trick used previously was used. Similarly, for the second term

$$\sum n_j z_j^{n_j} z_i^{n_i} f(n_i) = z_j \frac{\partial F}{\partial z_j}. \quad (2.2.7)$$

For the third and fourth terms

$$\sum (n_i + 1) z_i^{n_i} f(n_i + 1) = \sum n_i z_i^{n_i - 1} f(n_i) = \frac{\partial F}{\partial z_i}, \quad (2.2.8)$$

$$\sum n_i z_i^{n_i} f(n_i) = z_i \frac{\partial F}{\partial z_i}. \quad (2.2.9)$$

Replacing eqs. (2.2.6) - (2.2.9) in eq. (2.2.5) yields the equation for the moment generating function

[5] To avoid unnecesarily long expressions, we write $\sum$ refering to $\sum_{n_1=0}^{\infty} \cdots \sum_{n_N=0}^{\infty}$.

$$\dot{F}(z_i) = \sum_i \left(z_i \sum_j A_{ij} \frac{\partial F}{\partial z_j} - \sum_j A_{ij} z_j \frac{\partial F}{\partial z_j} + \Gamma_i \frac{\partial F}{\partial z_i} - \Gamma_i z_i \frac{\partial F}{\partial z_i} \right), \tag{2.2.10}$$

which after factoring becomes

$$\dot{F}(z_i) = \sum_i (z_i - 1) \left(\sum_j A_{ij} z_j \frac{\partial F}{\partial z_j} - \Gamma_i \frac{\partial F}{\partial z_i} \right). \tag{2.2.11}$$

We have to differentiate it and use the properties (1.6.3) - (1.6.5) to obtain equations for the moments. Differentiating with respect to z_l

$$\begin{aligned} \frac{\partial \dot{F}}{\partial z_l} = \sum_i \Bigg[(z_i - 1) \left[\sum_j A_{ij} \left(\delta_{jl} \frac{\partial F}{\partial z_j} + z_j \frac{\partial^2 F}{\partial z_j \partial z_l} \right) - \Gamma_i \frac{\partial^2 F}{\partial z_i \partial z_l} \right] \\ + \delta_{il} \left(\sum_j A_{ij} z_j \frac{\partial F}{\partial z_j} - \Gamma_i \frac{\partial F}{\partial z_i} \right) \Bigg]. \end{aligned}$$

$$\begin{aligned} \frac{\partial \dot{F}}{\partial z_l} = \sum_i (z_i - 1) \left[A_{il} \frac{\partial F}{\partial z_l} + \sum_j A_{ij} z_j \frac{\partial^2 F}{\partial z_j \partial z_l} - \Gamma_i \frac{\partial^2 F}{\partial z_i \partial z_l} \right] \\ + \sum_j A_{lj} z_j \frac{\partial F}{\partial z_j} - \Gamma_l \frac{\partial F}{\partial z_l}. \end{aligned}$$

Evaluating in $z_i = 1$, $i = 1, \ldots, N$ we otain after applying the properties of F

$$\dot{\langle n_l \rangle} = \sum_j A_{lj} \langle n_j \rangle - \Gamma_l \langle n_l \rangle,$$

and writing in matrix form

$$\dot{\langle \mathbf{n} \rangle} = (\mathbf{A} - \mathbf{\Gamma}) \langle \mathbf{n} \rangle, \tag{2.2.12}$$

as expected. It has the same form of the deterministic equations (2.2.1). Differentiating again with respect to z_m and doing some algebra

$$\frac{\partial^2 \dot{F}}{\partial z_l \partial z_m} = \sum_i (z_i - 1) \left(A_{im} \frac{\partial^2 F}{\partial z_i \partial z_m} + \sum_j A_{ij} z_j \frac{\partial^3 F}{\partial z_j \partial z_l \partial z_m} + A_{il} \frac{\partial^2 F}{\partial z_l \partial z_m} - \Gamma_i \frac{\partial^3 F}{\partial z_i \partial z_l \partial z_m} \right)$$
$$+ \sum_j A_{mj} z_j \frac{\partial^2 F}{\partial z_j \partial z_l} + A_{ml} \frac{\partial F}{\partial z_l} - \Gamma_m \frac{\partial^2 F}{\partial z_l \partial z_m}$$
$$+ A_{lm} \frac{\partial F}{\partial z_m} + \sum_j A_{lj} z_j \frac{\partial^2 F}{\partial z_j \partial z_m} - \Gamma_l \frac{\partial^2 F}{\partial z_l \partial z_m}.$$

Evaluating at $z_i = 1$ for all i

$$\frac{\partial^2 \dot{F}}{\partial z_l \partial z_m} = \sum_j A_{mj} z_j \frac{\partial^2 F}{\partial z_j \partial z_l} + A_{ml} \frac{\partial F}{\partial z_l} - \Gamma_m \frac{\partial^2 F}{\partial z_l \partial z_m} + A_{lm} \frac{\partial F}{\partial z_m} + \sum_j A_{lj} z_j \frac{\partial^2 F}{\partial z_j \partial z_m} - \Gamma_l \frac{\partial^2 F}{\partial z_l \partial z_m}.$$

Rearranging the previous eq. and using again the fact that $\mathbf{\Gamma}$ is diagonal

$$\frac{\partial^2 \dot{F}}{\partial z_l \partial z_m} = \sum_j \left(A_{mj} z_j - \Gamma_{mj} \right) \frac{\partial^2 F}{\partial z_j \partial z_l} + \sum_j A_{mj} \delta_{jl} \frac{\partial F}{\partial z_j}$$
$$+ \sum_j \left(A_{lj} z_j - \Gamma_{lj} \right) \frac{\partial^2 F}{\partial z_j \partial z_m} + \sum_j A_{lj} \delta_{jm} \frac{\partial F}{\partial z_j},$$

This is valid for all l and m. Evaluating at $z_i = 1$ for all i we get in matrix form

$$\nabla \nabla^T \dot{F}|_1 = \left(\left(\mathbf{\Gamma} - \mathbf{A} \right) \nabla \nabla^T F|_1 - \mathbf{A} \mathbf{\Theta} F|_1 \right) + \left(\left(\mathbf{\Gamma} - \mathbf{A} \right) \nabla \nabla^T F|_1 - \mathbf{A} \mathbf{\Theta} F|_1 \right)^T, \quad (2.2.13)$$

where $\Theta_{ij} \coloneqq \delta_{ij} \frac{\partial}{\partial z_i}$. The set of linear equations can be solved for the moments and correlation using a computer program. Given the specific form of the matrices $\mathbf{A}$ and $\mathbf{\Gamma}$, we only have to replace on eq. (2.2.12) and (2.2.13) to find the moments and the noise.

2.3 Application to non-linear interactions: negative autorregulation

As we have seen on sec. 1.3, the interactions between the component of a genetic circuits are non-linear. A consequence of this is that there might be several steady states. To apply the developed linear approach to such kind of systems we need to identify the stable fixed points, linearize about each of them, each linearization yields to a pair of matrices $\mathbf{A}$ and $\mathbf{\Gamma}$ that are replaced on eqs. (2.2.12) and (2.2.13) to find the noise.

We will consider the case of negative autorregulation, in this case k_r is now a Hill function for a repressor (eq. (1.3.2)) that depends on the number of proteins of the same gene. Assuming that the basal transcription rate is 0, the deterministic equations are

$$\begin{aligned} \dot{n_1}(t) &= k_r(n_2)d - \gamma_r n_1(t), \\ \dot{n_2}(t) &= k_p n_1(t) - \gamma_p n_2(t), \end{aligned} \tag{2.3.1}$$

where

$$k_r(n_2) \coloneqq \frac{k_r^{\max}}{1 + \left(\frac{n_2(t)}{K_d}\right)^n}.$$

Since the protein represses the transcription of his own gene, it is expected that the steady state number of proteins $\langle n_2 \rangle_s$ is reduced with respect to the unrepressed case. To evidence this, fig. 2.3.1 shows the distributions at steady state for the number of proteins of an unregulated and a negatively autorregulated gene resulting from Gillespie simulations. Both the average and the noise (width of the distribution) is reduced. The reduction of the noise occurs because a fluctuation that, for instance, increases the number of protein above the steady state level also increases the repression. This lowers the production rate, thus driving the levels back to the steady state with more strenght than in the unrepressed case. The effect is analogous for a random decrease of the level

of proteins below steady state.

See if noise is actually reduced (reducing the width)

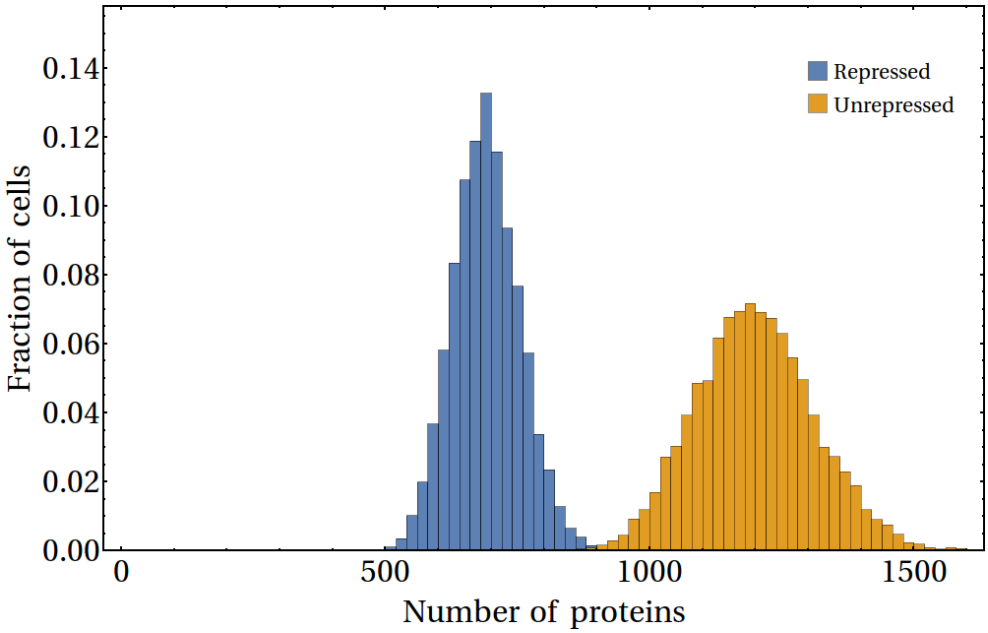

Figure 2.3.1: Distribution for the number of proteins for an autorrepressed gene (eq. 2.3.1, blue), and an unrepressed gene (eqs. 2.1.1 and 2.1.2 with $k_r = k_r^{\max}$, orange). The values of the parameters are $\tau_r = 2$ min, $\tau_p = 1$ h, $b = 10$ proteins/mRNA, $n = 2$, $K_d = 800$ proteins, and the average number of proteins without repression $\langle p \rangle_{\text{unrep}} := k_r^{\max} k_p/(\gamma_r \gamma_p) = 1200$ proteins. A sample of 10000 cells (trials) evolved until a time of $10\tau_p$ was run on each histogram.

Next we find the average and Fano factor of n_2 according to the linearized model. Making a first order Taylor expansion of $k_r(n_2)$ about its value at the steady state $k_r(\langle n_2 \rangle_s)$

$$\begin{aligned} k_r(n_2) &\approx k_r(\langle n_2 \rangle_s) + \left.\frac{\mathrm{d}k_r(n_2)}{\mathrm{d}n_2}\right|_{\langle n_2 \rangle_s} (n_2 - \langle n_2 \rangle_s) \\ &= \frac{k_r^{\max}}{1 + \left(\frac{\langle n_2 \rangle_s}{K_d}\right)^n} - \frac{k_r^{\max} n \left(\frac{\langle n_2 \rangle_s}{K_d}\right)^{n-1}}{K_d \left(1 + \left(\frac{\langle n_2 \rangle_s}{K_d}\right)^n\right)^2} (n_2 - \langle n_2 \rangle_s) \\ &= k_0 - \frac{k_1}{d} n_2, \end{aligned} \tag{2.3.2}$$

where k_0 and k_1 are given by

$$k_0 \coloneqq k_r(\langle n_2 \rangle_s) - \left.\frac{\mathrm{d}k_r(n_2)}{\mathrm{d}n_2}\right|_{\langle n_2 \rangle_s} \langle n_2 \rangle_s, \qquad k_1 \coloneqq -\left.\frac{\mathrm{d}k_r(n_2)}{\mathrm{d}n_2}\right|_{\langle n_2 \rangle_s} n_2.$$

Therefore, the matrix $\mathbf{A}$ becomes

$$\mathbf{A} = \begin{pmatrix} 0 & 0 & 0 \\ k_0 & 0 & -k_1 \\ 0 & k_p & 0 \end{pmatrix}$$

and $\mathbf{\Gamma}$ is the same as in (2.2.3). Solving eqs. (2.2.12) and (2.2.13) we get for the proteins in steady state

$$\boxed{\langle n_2 \rangle = \frac{1}{1+b\phi}\frac{k_0 b}{\gamma_p}}, \qquad \boxed{\nu_2 = \frac{1-\phi}{1+b\phi}\frac{b}{1+\frac{\gamma_p}{\gamma_r}}+1}, \tag{2.3.3}$$

with $\phi \coloneqq k_1/\gamma_p$ represents the strength of the feedback. In figure 2.3.2, we compare eq. (2.3.3) with a Gillespie simulation using the exact rates. Altought the analytic expressions are approximate, there is an excellent match with the simulations.

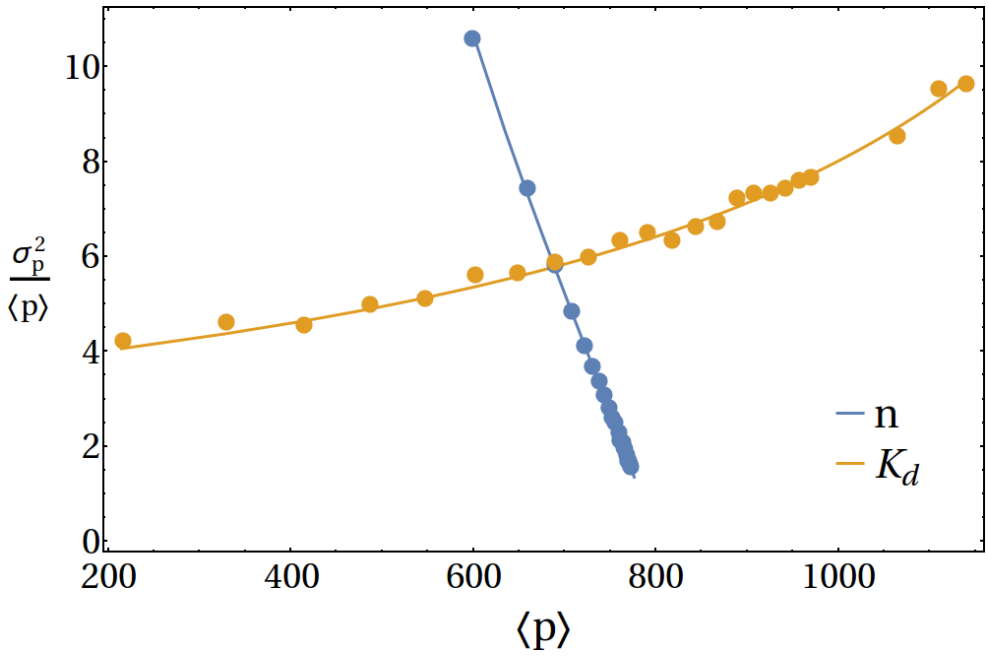

Figure 2.3.2: Comparison between the results of Gillespie simulations (dots) using the exact eqs. (2.3.1) and the analytical results given by eq. (2.3.3). The Fano factor is plotted vs. the mean number of proteins in steady state. The base values of the parameters are the same as in fig. 2.3.1. For each curve, the parameter indicated in the legend is varied while the others are fixed. K_d is varied from 100 to 2000 in increments of 100 and from 2000 to 5000 in increments of 1000, while n is varied from 0 to 20 in unit increments. Each point correspond to 10000 trials. Each one evolved until a time of $10\tau_p$.

For a given average number of proteins, this circuit has a more efficient control of fluctuations than the one for the unrepressed gene. In this case, the random fluctuations are controlled by the same number of proteins and there is no need to spend energy in unused mRNA molecules. Evolution might have developed circuits of this kind, and more sophisticated, to tune the noise according to the function of the circuits trying to reduce the cost in fitness.

2.4 Limitations of the model

The linearized model has reliably reproduced the noise in steady state of two simple genetic circuits. We have also seen that it allows to consider an arbitrary number of species, allowing then to treat networks of arbitrary complexity. However, the model have some limitations. If the system is nonlinear, The nonlinearities should be such that the linearization correctly reproduces the behavior near the fixed points i.e. the nonlinearities should not be too large. Besides, this approach does not work for oscillating systems or unstable steady state points because in both mechanisms there may be large sudden changes in the average values that are not reproducible with a linear model.

However, phenomena such that big nonlinearities and oscillations is common in biological circuits and has allowed living beings to have many complex and interesting behaviors. Those features can not be studied with this model. Also, for many systems it would be important to model the complete time dynamics of noise, not just the steady state values. For example, the stochastic switching between different steady state values.

Define fixed points, write a better justification in the above paragraph

Regardless of its limitations, the linearized ME approach can be used to model a variety of biological circuits and has enligthened some important facts about the mechanisms by which living beings control noise. In the following chapters more sources of noise will be included in the models and additional mathematical tools will be applied.

Chapter 3

The Fluctuation-Dissipation theorem (FDT)

This chapter is based on the work of J. Paulsson in [17] and [4].

3.1 Statement of the FDT

Consider a (genetic) system with N species whose concentrations are $n_1, \ldots, n_N$. The FDT states that if σ is the matrix of covariances (i.e. $\sigma_{ij} \coloneqq \langle n_i - \langle n_i \rangle \rangle \langle n_j - \langle n_j \rangle \rangle$), then it follows that

$$\frac{\mathrm{d}\sigma}{\mathrm{d}t} = \mathbf{A}\sigma + \sigma\mathbf{A}^{\mathbf{T}} + \mathbf{B}. \tag{3.1.1}$$

The elements of $\mathbf{A}$ are,

$$A_{ij} \coloneqq \frac{\partial}{\partial\langle n_j\rangle}\frac{\partial\langle n_i\rangle}{\partial t} = \frac{\partial}{\partial\langle n_j\rangle}\left(\langle J_i^+\rangle - \langle J_i^-\rangle\right). \tag{3.1.2}$$

where $J_i^{\pm}$ are the total fluxes of synthesis and degradation of species i. For $\mathbf{B}$

$$B_{ij} = \sum_k v_{jk} v_{ik} R_k,$$

where k runs over all the possible reactions for the system. R_k is the rate of reaction k, which produces v_{ik} molecules of species i.

In steady state, the FDT becomes

$$\mathbf{A}\sigma + \sigma\mathbf{A}^{\mathbf{T}} + \mathbf{B} = 0 \tag{3.1.3}$$

where $\mathbf{A}$ and $\mathbf{B}$ are now evaluated at steady state. We illustrate the concept with an example in the next section.

3.2 Example: simple gene

Consider a single gene such as the explained on section 2.1. We will treat the same case without the suposition of a fixed number of active copies of the gene. In this case, there is a fixed number $n_1^{\max}$ and n_1 copies that are active, that is, that are available for transcription. There are several ways in which genes can be turned on and off including binding of transcription factors or chromatin remodeling.

Explain chromatin remodeling

We assume that the activation and activation for each gene follows a telegraph process, i.e. each gene switches from *off* to *on* with rate k_1^+ and in the opposite sense with rate k_1^-. This can be illustrated as

$$\text{off} \underset{k_1^-}{\overset{k_1^+}{\rightleftharpoons}} \text{on}.$$

Let $P_{\text{on}}(t)$ and $P_{\text{off}}(t)$ be the probabilities for a gene to be active or inactive, respectively. Clearly $P_{\text{on}}(t) + P_{\text{off}}(t) = 1$ and the ME for $P_{\text{on}}(t)$ is given by

$$\dot{P}_{\text{on}}(t) = k_1^+ P_{\text{off}}(t) - k_1^- P_{\text{on}}(t).$$

In steady state,

$$P_{\text{on}} = \frac{k_1^+}{k_1^-} P_{\text{off}} = \frac{k_1^+}{k_1^-}(1 - P_{\text{on}}),$$

hence,

$$P_{\text{on}} = \frac{k_1^+}{k_1^+ + k_1^-}.$$

Besides, suppose that the $n_1^{\max}$ copies of the gene are independent. Then, in steady state n_1 follows a binomial distribution

$$P(n_1) = \binom{n_1^{\max}}{n_1} P_{\text{on}}^{n_1} (1 - P_{\text{on}})^{n_1^{\max} - n_1}.$$

Recalling that the averages satisfy the deterministic equations, $\langle n_1 \rangle$ follows

$$\dot{\langle n_1 \rangle} = k_1^+ (n_1^{\max} - \langle n_1 \rangle) - k_1^- \langle n_1 \rangle = k_1^+ n_1^{\max} - (k_1^+ + k_1^-)\langle n_1 \rangle.$$

For the number of mRNA n_2 and proteins n_3, we use the same equations as on section 2.1. To make the analysis clearer, we define $\tau_i \coloneqq 1/\gamma_i$, for $i = 2, 3$ and $\tau_1 \coloneqq (k_1^+ + k_1^-)^{-1}$, then

$$\begin{aligned}
\dot{\langle n_1 \rangle} &= k_1^+ n_1^{\max} - \frac{1}{\tau_1}\langle n_1 \rangle, \\
\dot{\langle n_2 \rangle} &= k_2^+ \langle n_1 \rangle - \frac{1}{\tau_2}\langle n_2 \rangle, \\
\dot{\langle n_3 \rangle} &= k_3^+ \langle n_2 \rangle - \frac{1}{\tau_3}\langle n_2 \rangle.
\end{aligned} \tag{3.2.1}$$

k_2 and k_3 are the rates of transcription per active gene and translation per mRNA, respectively. A master equation can be easily written and treated using the methods presented on chapter 2, but we will use the FDT instead.

First, we need to write the matrices $\mathbf{A}$ and $\mathbf{B}$, according to their definition, we have for instance

$$A_{11} = \partial_{n_1}\langle \dot{n_1}\rangle = -\frac{1}{\tau_1},$$
$$A_{12} = \partial_{n_1}\langle \dot{n_2}\rangle = k_2^+$$

and so on. In this case, the fluxes of synthesis of degradation are $\langle J_1^+\rangle = (n_1^{\max} - \langle n_1\rangle)$ and $\langle J_i^-\rangle = \langle n_1\rangle/\tau_1$ (the minus sign is not included in $\langle J_i^-\rangle$). The definition is analogous for the other species. Notice that there are several terms that are zero, including A_{12}, A_{13}, A_{21}, etc. Evaluating all the elements results in

$$\mathbf{A} = \begin{pmatrix} -1/\tau_1 & 0 & 0 \\ k_2 & -1/\tau_2 & 0 \\ 0 & k_3 & -1/\tau_3 \end{pmatrix}. \tag{3.2.2}$$

For the matrix $\mathbf{B}$, there are six reactions in this case: activation and deactivation of a gene, synthesis and destruction of mRNA, and the same for proteins. There are not reactions involving different species neither creating or destroying more than one molecule, thus

$$v_{ik}v_{jk} = \begin{cases} 1 \text{ for } i = j, \\ 0 \text{ for } i \neq j. \end{cases}$$

Where the first case stands because $v_{ik} = \pm 1$ depending on whether the reaction is of synthesis or destruction, but in either case $v_{ik}^2 = 1$.

The rates R_k are given by the deterministic equations. In steady state the rates of creation and destruction are equal. For example, for the reactions involving modifications in n_2:

$$B_{22} \coloneqq \sum_k v_{2k}^2 R_k = \sum_k R_k = k_2\langle n_1\rangle + \frac{1}{\tau_2}\langle n_2\rangle = \frac{2\langle n_2\rangle}{\tau_2},$$

where the last equality holds from the steady state assumption. For B_{33} it is analo-

gous. For B_{11},

$$B_{11} = k_1^+(n_1^{\max} - \langle n_1 \rangle) + k_1^- \langle n_1 \rangle = 2k_1^- \langle n_1 \rangle = 2\frac{1 - P_{\text{on}}}{\tau_1}\langle n_1 \rangle,$$

since

$$k_1^- = \left(1 - \frac{k_1^+}{k_1^+ + k_1^-}\right)(k_1^+ + k_1^-) = \frac{1 - P_{\text{on}}}{\tau_1}.$$

Putting all the expressions together, $\mathbf{B}$ becomes

$$\mathbf{B} = \begin{pmatrix} 2(1 - P_{\text{on}})\langle n_1 \rangle/\tau_1 & 0 & 0 \\ 0 & 2\langle n_2 \rangle/\tau_2 & 0 \\ 0 & 0 & 2\langle n_3 \rangle/\tau_3 \end{pmatrix}. \tag{3.2.3}$$

Replacing on (3.1.3) we obtain a linear system that can be solved for σ using a computer program. The diagonal elements are the variances, from which the noises can be found. The next expressions for the coefficients of variation can be trivially found following the procedure, although the calculation are quite extensive. The results are

$$\eta_1^2 = \frac{1 - P_{\text{on}}}{\langle n_1 \rangle}, \tag{3.2.4}$$

$$\eta_2^2 = \frac{1}{\langle n_2 \rangle} + \frac{1 - P_{\text{on}}}{\langle n_1 \rangle}\frac{\tau_1}{\tau_1 + \tau_2}, \tag{3.2.5}$$

$$\eta_3^2 = \frac{1}{\langle n_3 \rangle} + \frac{1}{\langle n_2 \rangle}\frac{\tau_2}{\tau_2 + \tau_3} + \frac{1 - P_{\text{on}}}{\langle n_1 \rangle}\frac{\tau_2}{\tau_2 + \tau_3}\frac{\tau_1}{\tau_1 + \tau_3}\frac{\tau_1 + \tau_3 + \tau_1\tau_3/\tau_2}{\tau_1 + \tau_2}.. \tag{3.2.6}$$

For eq. (3.2.4) the relative width of the distribution is smaller that Poissonian for a given value of $\langle n_1 \rangle$. If P_{on} is close to 1 the noise could be low even if n_1 is small.

The first term in eq. (3.2.5) is the Poissonian noise (see eq. (1.9.1)). The second term is the noise arising from gene activation multiplied by a factor of time averaging. That depends on the decay rates of each molecule. Due to its importance, we will devote a section to this phenomenon.

3.3 Time averaging

We have seen in sec. 2.1 that the response time depends on the degradation rate and that its inverse is proportional to the speed for the levels of some molecule to reach its steady state levels. In this example, the fluctuations in the number of active genes n_1 change the steady state levels of the number of mRNA molecules. The mRNA therefore continuously has to reach its new stationary level with a timescale given by τ_2. Hence, if τ_2 is small relative to τ_1, the mRNAs respond quickly to fluctuations in the number of active genes, increasing noise in mRNA. In the opposite case, what occurs is that in a time of τ_2 the number of active genes fluctuated many times. Since the mRNA levels is unable to follow all of them, it makes a time average that reduces its fluctuations.

This can be seen in the term of time averaging of eq. (3.2.5):

$$\frac{\tau_1}{\tau_1 + \tau_2} = \frac{1}{1 + \tau_2/\tau_1}$$

In the first case $\tau_2 \ll \tau_1$, thus the time averaging term goes to 1 and the noise in n_1 is fully propagated. In the second case $\tau_2 \gg \tau_1$, then the term goes to zero, and there is no propagated noise from n_1 to n_2.

Add figures

For the number of proteins n_3 we can see in eq. (3.2.6) that the squared noise is the sum of its intrinsic Poisson noise ($(1/n_3)$), the propagated noise from mRNA that is modulated by a time average factor and the noise from gene activation that is propagated first to the mRNAs and then to proteins. This two-step propagation explains the more complex time averaging factor.

Altought the noise in mRNA is usually large due to its low numbers. The effect of time averaging could reduce considerably the noise in the number of proteins with respect to the noise in mRNA. Without this, the reliability of biological circuits would be very low. Besides, since a time average factor takes values between 0 and 1, the noise propagated across multiple steps is reduced with the number of steps. This could be

another strategy for noise reduction.

3.4 The logarithmic gain

An important quantity for the analysis of genetic circuits is the logarithmic gain or elasticity H_{ij}. It measures the amount by which changes in the quantities of the jth component produces changes in the ith component. It is defined as

$$H_{ij} = \frac{\partial \ln(\langle J_i^- \rangle / \langle J_i^+ \rangle)}{\partial \ln \langle n_j \rangle}. \tag{3.4.1}$$

Recall that $\partial \ln x = \partial x / x$. Thus, H_{ij} measures by how fractional changes in $\langle n_j \rangle$ cause the levels of species i to decrease due to fractional changing its degradation to synthesis ratio $\langle J_i^+ \rangle / \langle J_i^- \rangle$. If $H_{ij} = h$, then a 1% increase in n_j will increase the degradation to synthesis ratio by approximately h%.

Include figure of the stepness

Eq. (3.1.3) can be rewritten in terms of the logarithmic gain to a more compact form using the following definitions

$$\eta_{ij} \coloneqq \frac{\sigma_{ij}}{\langle n_i \rangle \langle n_j \rangle}, \quad M_{ij} \coloneqq \frac{\langle n_j \rangle}{\langle n_i \rangle} A_{ij}, \quad D_{ij} \coloneqq \frac{B_{ij}}{\langle n_i \rangle \langle n_j \rangle}.$$

It can be easily shown that in steady state the equation

$$\mathbf{M}\eta + \eta \mathbf{M}^T + \mathbf{D} = 0 \tag{3.4.2}$$

corresponds to eq. (3.1.3) after dividing by $\langle n_i \rangle \langle n_j \rangle$. The diagonal terms of η are the squared CVs. Therefore, using these definitions it is more direct to calculate them. The elements of $\mathbf{M}$ can be written in terms of the logarithmic gain. From eq. (3.1.2)

$$
\begin{aligned}
A_{ij} &= \frac{\partial}{\partial\langle n_j\rangle}\left(\langle J_i^+\rangle - \langle J_i^-\rangle\right) = \frac{1}{\langle n_j\rangle}\frac{\partial}{\frac{\partial\langle n_j\rangle}{\langle n_j\rangle}}\left(\langle J_i^+\rangle - \langle J_i^-\rangle\right) \\
&= \frac{1}{\langle n_j\rangle}\left(\frac{\langle J_i^+\rangle\frac{\partial\langle J_i^+\rangle}{\langle J_i^+\rangle}}{\partial\ln\langle n_j\rangle} - \frac{\langle J_i^-\rangle\frac{\partial\langle J_i^-\rangle}{\langle J_i^-\rangle}}{\partial\ln\langle n_j\rangle}\right) = \frac{1}{\langle n_j\rangle}\left(\frac{\langle J_i^+\rangle\partial\ln\langle J_i^+\rangle}{\partial\ln\langle n_j\rangle} - \frac{\langle J_i^-\rangle\partial\ln\langle J_i^-\rangle}{\partial\ln\langle n_j\rangle}\right) \\
&= \frac{\langle J_i\rangle}{\langle n_j\rangle}\left(\frac{\partial\ln\langle J_i^+\rangle}{\partial\ln\langle n_j\rangle} - \frac{\partial\ln\langle J_i^-\rangle}{\partial\ln\langle n_j\rangle}\right) = -\frac{\langle J_i\rangle}{\langle n_j\rangle}\frac{\partial\ln\left(\langle J_i^-\rangle/\langle J_i^+\rangle\right)}{\partial\ln\langle n_j\rangle} = -\frac{\langle J_i\rangle}{\langle n_j\rangle}H_{ij}.
\end{aligned}
\tag{3.4.3}
$$

In steady state, $\langle J_i^+\rangle = \langle J_i^-\rangle := \langle J_i\rangle$ and for the kind of systems considered the flux of degradation is of the form $\langle J_i^-\rangle = \frac{\langle n_i\rangle}{\tau_i}$. Hence replacing on the previous eq. and on the definition of $\mathbf{M}$ we get

$$
M_{ij} = -\frac{H_{ij}}{\tau_i}.
$$

Also, for genetic systems of this kind, where there are not reactions producing several components and the relations are linear. The only nonzero elements of the matrix $\mathbf{D}$ are the diagonal, which according to eq. (3.1) are

$$
D_{ii} = \frac{\sum_k R_k}{\langle n_i\rangle^2} = \frac{\langle J_i^+\rangle + \langle J_i^-\rangle}{\langle n_i\rangle^2} = 2\frac{\langle J_i\rangle}{\langle n_i\rangle^2} = \frac{2}{\langle n_i\rangle\tau_i}.
$$

In the example used in this chapter. The matrices can be easily found using the previous results

$$
\mathbf{H} = \begin{pmatrix} -1 & 0 & 0 \\ 1 & -1 & 0 \\ 0 & 1 & -1 \end{pmatrix}, \qquad \mathbf{M} = \begin{pmatrix} -1/\tau_1 & 0 & 0 \\ 1/\tau_2 & -1/\tau_2 & 0 \\ 0 & 1/\tau_3 & -1/\tau_3 \end{pmatrix},
$$

$$\mathbf{D} = \begin{pmatrix} 2/\langle n_1 \rangle \tau_1 & 0 & 0 \\ 0 & 2/\langle n_2 \rangle \tau_2 & 0 \\ 0 & 0 & -1/\langle n_3 \rangle \tau_3 \end{pmatrix}.$$

Replacing in eq. (3.4.2) we obtain equivalently eqs. (3.2.4) - (3.2.6).

In this example the logarithmic gain only takes values ± 1 so its effect in noise propagation is not so clear. In the next chapter we will consider a model that involves nonlinear interaction of Hill type between different species where it becomes explicit that the noise propagation between different species depens on the intrinsic noise of each species, the effect of time averaging, and the logarithmic gain, If the logarithmic gain is high in absolute value, the noise will be amplified as it is propagated and viceversa.

Chapter 4

Cascade of regulation - The Langevin equation

4.1 Intrinsic noise in a single gene using the Langevin approach

In the Langevin approach, a term representing the noise of the system is added to the deterministic equations instead of considering the transition rates between states explicitely. Knowing some of the statistical properties of the noise term will allow us to find the noise in the number of species. In this section we will use the Langevin approach to find the intrinsic noise for a single gene.

Consider the model used on section 2.1. For each of the deterministic equations (eqs. (2.1.1) and (2.1.2)), a stochastic process $\mu(t)$ is added that accounts for the intrinsic noise. Including the terms $\mu_1(t)$ and $\mu_2(t)$ results in

$$\dot{n_1}(t) = k_r d - \gamma_r n_1(t) + \mu_1(t), \tag{4.1.1}$$

$$\dot{n_2}(t) = k_p n_1(t) - \gamma_p n_2(t) + \mu_2(t). \tag{4.1.2}$$

Now $n_1(t)$ and $n_2(t)$ are stochastic processes. We need some information about the noise terms in order to proceed. First, as we saw in chapter 2, the averages must follow the deterministic behavior. By taking averages on both sides of eqs. (4.1.1) and (4.1.2) and imposing this condition we get

$$\langle \mu_1 \rangle(t) = \langle \mu_2 \rangle(t) = 0.$$

Also, assuming white noise statistics, the autocorrelations are given by

$$\langle \mu_1(t)\mu_1(t+\tau) \rangle = q_1^2 \delta(\tau), \tag{4.1.3}$$

$$\langle \mu_2(t)\mu_2(t+\tau) \rangle = q_2^2 \delta(\tau). \tag{4.1.4}$$

This means that we will assume that there is no correlation between the values of the noise term at different times. The coefficients q_1 and q_2 determine the strenght of the noise and will be treated later. Also, the intrinsic noise must be fully uncorrelated among mRNA and proteins, hence

$$\langle \mu_1(t)\mu_2(t+\tau) \rangle = 0. \tag{4.1.5}$$

We will use eqs. (4.1.3) - (4.1.5) to find the noise in mRNA and protein numbers. Define the difference with respect to steady state average as δn_1 and δn_2, i.e. $\delta n_i \coloneqq n_i - \langle n_i \rangle_s$, for $i = 1, 2$. In terms of these quantities eqs. (4.1.1) and (4.1.2) become

$$\dot{\delta n_1}(t) = -\gamma_r \delta n_1(t) + \mu_1(t), \tag{4.1.6}$$

$$\dot{\delta n_2}(t) = k_p \delta n_1(t) - \gamma_p \delta n_2(t) + \mu_2(t). \tag{4.1.7}$$

Fourier transforming and using the fact that $[\mathscr{F}(\dot{x}(t))](\omega) = i\omega\hat{x}$, where $\hat{x} \coloneqq \mathscr{F}(x)$ we get

$$\delta\hat{n}_1(\omega) = \frac{\hat{\mu}_1(\omega)}{\gamma_r + i\omega}, \qquad \delta\hat{n}_2(\omega) = \frac{\delta\hat{n}_1(\omega) + \hat{\mu}_2(\omega)}{\gamma_p + i\omega}. \tag{4.1.8}$$

Taking the average of the square norm of the first expression we obtain the power spectrum for δn_1

$$\langle|\delta\hat{n}_1|^2\rangle = \frac{\langle|\hat{\mu}_1|^2\rangle}{\omega^2 + \gamma_r^2} = \frac{q_1^2}{\omega^2 + \gamma_r^2}, \tag{4.1.9}$$

because by using the Wiener-Khinchin theorem (eq. (1.8.1)) and eq. (4.1.3) we obtain $\langle|\hat{\mu}_1|^2\rangle = q_1^2$. Making the inverse Fourier transform and evaluating at $t = 0$ yields the variance of n_1

$$\sigma^2(n_1) = \langle\delta n_1^2\rangle = q_1^2\left[\mathscr{F}^{-1}\left(\frac{1}{\omega^2 + \gamma_r^2}\right)\right](t=0) = q_1^2\int_{-\infty}^{\infty}\frac{1}{\omega^2 + \gamma_r^2}\frac{\mathrm{d}\omega}{2\pi} = \frac{q_1^2}{2\gamma_r}. \tag{4.1.10}$$

The integral can be easily solved in the complex plane or by trigonometric substitution. Keeping in mind that mRNA creation and destruction are single step Poisson processes and as we saw on chapter 2, the variance must be equal to the average. Imposing this condition, we find that q_1^2 is given in steady state by

$$q_1^2 = 2\gamma_r\sigma_s^2(n_1) = 2\gamma_r\langle n_1\rangle_s = 2k_r d$$

There is a more general procedure for finding the noise strenght q in steady state. It can be shown that for single step Poisson processes it is given by the square root of the sum of the rates for all the events evaluated at the steady state average. If amount x of some molecule follows the deterministic equation

$$\dot{x} = f(x) - g(x),$$

then

$$q_x = \sqrt{f(\langle x \rangle_s) + g(\langle x \rangle_s)} \tag{4.1.11}$$

In this case, we have

$$q_1 = \sqrt{k_r d + \gamma_r \langle n_1 \rangle_s} = \sqrt{2k_r d}.$$

Hence, $\nu_1 = 1$ as we obtained in the previous chapter using the master equation approach. To find the noise in n_2 we follow the same procedure for the second term of eq. (4.1.8) using also the obtained results for n_1. Taking the average of the square norm

$$\langle |\delta\hat{n}_2|^2 \rangle = \frac{\langle |\delta\hat{n}_1|^2 \rangle + \langle |\hat{\mu}_2|^2 \rangle + \langle \delta\hat{n}_1^* \hat{\mu}_2 \rangle + \langle \delta\hat{n}_1 \hat{\mu}_2^* \rangle}{\omega^2 + \gamma_p^2}.$$

Using the WK theorem we find from eq. (4.1.5) that the cross terms are zero. From eq. (4.1.4) we obtain $\langle |\hat{\mu}_2|^2 \rangle = q_2^2$. Putting together these results and using eq. (4.1.9) it yields

$$\langle |\delta\hat{n}_2|^2 \rangle = \frac{q_2^2}{\omega^2 + \gamma_p^2} + \frac{q_1^2}{(\omega^2 + \gamma_r^2)(\omega^2 + \gamma_p^2)}$$

Using eq (4.1.11), $q_2^2 = 2k_p k_r/\gamma_r$. Hence performing the inverse Fourier transform at $t = 0$ we get

$$\sigma^2(n_2)_s = \frac{2k_r}{2\pi} \left[\int_{-\infty}^{\infty} \frac{\mathrm{d}\omega}{(\omega^2 + \gamma_r^2)(\omega^2 + \gamma_p^2)} + \frac{k_p}{\gamma_r} \int_{-\infty}^{\infty} \frac{\mathrm{d}\omega}{\omega^2 + \gamma_p^2} \right].$$

After solving the integrals using residues, the result is

$$\sigma^2(n_2)_s = \langle p \rangle_s \left(\frac{b}{1 + \gamma_p/\gamma_r} + 1 \right), \tag{4.1.12}$$

with $b \coloneqq k_p/\gamma_r$, also consistent with the results of section 2.1.

4.2 Model circuit for the cascade

The calculations shown in this chapter are based in the work of J. M. Pedraza and A. van Oudenaarden in [3] and [18]. They build the model circuit and also tested the theoretical results experimentally.

We will consider a set of genes whose interactions are shown on figure 4.2.1 considering both intrinsic and global sources of noise. The intrinsic part refers to the inherent noise due to the low number of molecules and the nature of the reactions. This was the only source of noise consider on the previous chapter. The extrinsic part arises from another factors, such as environmental fluctuations or variations in intracellular concentrations due to sudden changes on cell volume. These factors causes fluctuations in every component of the cell and thus extrinsic noise is correlated among the different genes, while intrinsic noise is not [2].

Explain intrinsic/extrinsic noise in a section on concepts.

Explain plasmid and chromosomal DNA, and constitutive promoter.

Fig. 4.2.1 shows a genetic circuit built from four genes to be used on bacteria. Gene 0 is located in the chromosome while genes 1 to 3 are located in plasmids, hence, their expression is subjected to noise caused by plasmid number fluctuations altought we will neglect this source of noise [1]. Genes 0 codifies for *LacI* and 3 for the red fluorescent protein *rfp*. Both are regulated by a constitutive promoter, gene 1 has the promoter P_{lac}, which regulates the expression of *tetR* and cfp (cyan fluorescent protein). The transcription from P_{lac} is repressed by *LacI*. Also, the repressing effect of *LacI* is inhibited by IPTG.

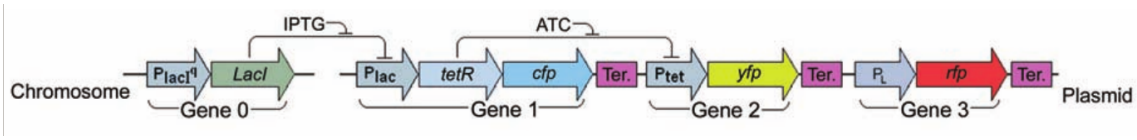

Figure 4.2.1: Circuit used in the Langevin model (from [3]).

[1] We will fix the plasmid number at 1. This makes calculations easier to follow and does not disrupt the analysis, the results with variable number of plasmids can be found in [3]

A similar effect happens on gene 2, the tetracycline promoter P_{tet}, which regulated the expression of the yellow fluorescent protein *yfp*. It is repressed by *tetR* and the repressing effect of *tetR* is also inhibited by ATC.

Therefore, both IPTG and ATC are environmental signals that are used to regulate the coupling between the different genes of the cascade. Notice that for instance, at higher quantities of IPTG, expression of gene 1 is low since the repression over it is reduced. This also causes that without ATC, the expression of gene 2 is high.

The fluorescent proteins are used to quantify experimentally via fluorescence microscopy the expressions of each gene, which will be proportional to the intensity of the flourescence on each color. *tetR* and *cfp* are transcribed bicistronically in order to be able to quantify *tetR* levels. We will assume that protein degradation is caused only by cell division (i.e. there is no active degradation). That allow us to use the same degradation constant for all proteins, and in particular, to assume that the behavior of *cfp* reproduces exactly the behavior of *tetR*.

We will label the concentrations of *LacI*, *tetR* (and *cfp*), *yfp* and *rfp* as x_0, x_1, x_2, and x_3, respectively.

4.3 Mathematical derivations

The differential equation for the mRNA will not be considered, we will write the equation for the proteins and include the effect of mRNA in the rate of creation k. The results of the previous chapter for the noise in proteins caused by mRNA will also be considered. The deterministic equation for the concentration of proteins of gene 0 is

$$\dot{x_0}(t) = k - \gamma x_0(t). \tag{4.3.1}$$

Where now k represents the average number of proteins created per unit time[2].

[2]For a better understanding of this point. If we assume $\gamma_p \ll \gamma_r$, we can treat the mRNA in steady state, hence eq. (2.1.2) becomes eq. (4.3.1) with $k \coloneqq k_p \langle n_1 \rangle_s = \frac{k_p k_r}{\gamma_r} = k_r b$.

We will use the Langevin approach following the same procedures that were explained on the previous section. Here we add two noise terms to the deterministic equations representing the intrinsic noise $\mu_0(t)$ and the global noise $\xi_0(t)$. Hence, the equation for the stochastic process $x_0(t)$ is

$$\dot{x_0}(t) = k - \gamma x_0(t) + \mu_0(t) + \xi_0(t). \tag{4.3.2}$$

The noise terms have zero average, i.e.

$$\langle \mu_0 \rangle(t) = \langle \xi_0 \rangle(t) = 0$$

and assuming white noise statistics for both sources, the autocorrelation functions are

$$\langle \mu_0(t)\mu_0(t+\tau) \rangle = q^2_{0,\text{int}}\delta(\tau) = 2\gamma(b_0+1)\bar{x}_0\delta(\tau), \tag{4.3.3}$$

$$\langle \xi_0(t)\xi_0(t+\tau) \rangle = q^2_{0,G}\delta(\tau) = 2\gamma\eta^2_G\bar{x}^2_0\delta(\tau). \tag{4.3.4}$$

where η_G is the strenght of the global noise, a parameter that is measured experimentally, and b_0 is the average number of protein produced per mRNA. In this section the bar denotes steady state average. Also, since both sources of noise are uncorrelated

$$\langle \mu_0(t)\xi_0(t+\tau) \rangle = 0. \tag{4.3.5}$$

Understand and explain more the constants and the assumption of white noise

We will derive the constant term in eq. (4.3.3). Notice that we can not use eq. (4.1.11) to find the factor q in the correlations since with the assumptions made in eq. (4.3.1), protein creation is not a one step Poisson process. From the results of section we obtain comparing with eq. (4.1.10)

$$\sigma^2(x_0) = \frac{q^2_{0,\text{int}}}{2\gamma}$$

and from eq. (4.1.12) we obtain

$$\bar{x}_0 \left(\frac{b_0}{1 + \gamma_p/\gamma_r} + 1 \right) = \frac{q^2_{\text{int}}}{2\gamma}.$$

The bar denotes steady state average. Since proteins are much more stable than mRNA, we can take $\gamma_p \ll \gamma_r$. With this approximation we get

$$q^2_{0,\text{int}} \approx 2\gamma \bar{x}_0 (b_0 + 1).$$

In fact, since intrinsic fluctuations depends only on the gene in question, the factor q has the same form for all genes, i.e.

$$q^2_{i,\text{int}} \approx 2\gamma \bar{x}_i (b_i + 1), \quad i = 0, 1, 2, 3.$$

In terms of $\delta x_0 \coloneqq x_0 - \bar{x}_0$, eq. (4.3.2) becomes

$$\dot{\delta x_0}(t) = -\gamma \delta x_0(t) + \mu_0(t) + \xi_0(t). \tag{4.3.6}$$

We will Fourier transform the equation, find its square norm and use the Wiener-Khinchin theorem (eq. (1.8.1)) to find the autocorrelations in terms of the power spectrum and to write the power spectrum of $\mu(t)$ and $\xi(t)$ in terms of their autocorrelations.

OJO

Fourier transforming and recalling that $[\mathscr{F}(\frac{\mathrm{d}x(t)}{\mathrm{d}t})](\omega) = i\omega \mathscr{F}(x(t))(\omega)$ for a function of time $x(t)$, we obtain after solving for $\hat{\delta x_0}$

$$\hat{\delta x_0}(\omega) = \frac{\hat{\mu_0} + \hat{\xi_0}}{\gamma + i\omega}. \tag{4.3.7}$$

Taking the square norm and averaging we get

$$\left\langle |\delta\hat{x}_0|^2 \right\rangle = \frac{\langle |\hat{\mu_0}|^2 \rangle + \left\langle \hat{\mu_0}^* \hat{\xi_0} \right\rangle + \left\langle \hat{\mu_0} \hat{\xi_0}^* \right\rangle + \left\langle |\hat{\xi_0}|^2 \right\rangle}{\gamma^2 + \omega^2}, \tag{4.3.8}$$

Using the Wiener-Khinchin theorem and eqs. (4.3.3) - (4.3.5)

$$\begin{aligned} \left\langle |\delta\hat{x}_0|^2 \right\rangle &= \frac{(2\gamma(b_0+1)\bar{x_0} + 2\gamma\eta_G^2 \bar{x_0}^2)\, \mathscr{F}(\delta(t))}{\gamma^2 + \omega^2} \\ &= \frac{2\gamma \bar{x_0}^2 \left({}^{(b_0+1)}/_{\bar{x_0}} + \eta_G^2 \right)}{\gamma^2 + \omega^2}, \end{aligned} \tag{4.3.9}$$

where the cross terms are zero by eq. (4.3.5). Applying the inverse Fourier transform at $t = 0$ we get

$$\langle \delta x_0^2 \rangle = 2\gamma \bar{x_0}^2 \left({}^{(b_0+1)}/_{\bar{x_0}} + \eta_G^2 \right) \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{d\omega}{\omega^2 + \gamma^2}$$

The integral can be easily solved by residues resulting in π/γ, therefore

$$\langle \delta x_0^2 \rangle = \bar{x_0}^2 \left(\frac{(b_0+1)}{\bar{x_0}} + \eta_G^2 \right)$$

And dividing by $\bar{x_0}^2$, we obtain the squared coefficient of variation

$$\boxed{\eta_0^2 = \frac{(b_0+1)}{\bar{x_0}} + \eta_G^2 = \eta_{0,\text{int}}^2 + \eta_G^2}, \tag{4.3.10}$$

where $\eta_{i,\text{int}}^2 \coloneqq \frac{(b_i+1)}{\bar{x_i}}$ for $i = 0, 1, 2, 3$.

Therefore the total squared noise for gene 0 is composed of the contributions from both the intrinsic and the global noise.

Now we will make the calculation for gene 1, which follows the equation

$$\dot{x_1}(t) = k_1(x_{0A}) - \gamma x_1 + \mu_1 + \xi_1 \tag{4.3.11}$$

The creation rate k_1 is a Hill equation for repression where the repressor is x_{0A}, the amount of x_0 that is unbound to IPTG. The statistics for the noise terms are analogous

to eqs. (4.3.3) - (4.3.5). We also need to know in this case the correlations between the noise terms corresponding to gene 0 and the ones corresponding to gene 1. As we have said, extrinsic sources of noise are uncorrelated

$$\langle \mu_0(t)\mu_1(t+\tau)\rangle = \langle \mu_0(t)\xi_1(t+\tau)\rangle = \langle \mu_1(t)\xi_0(t+\tau)\rangle = 0, \tag{4.3.12}$$

but the extrinsic parts of the noise of genes 0 and 1 are correlated. In analogy with eq. (4.3.4) we get

$$\langle \xi_0(t)\xi_1(t+\tau)\rangle = 2\gamma\eta_G^2\bar{x_0}\bar{x_1}\delta(\tau). \tag{4.3.13}$$

Also, understand and explain the q term here

We proceed in a similar way to gene 0. Defining $\delta x_1(t) \coloneqq x_1(t) - \bar{x_1}$, writing eq. (4.3.11) in terms of δx_1, δx_{0A}, and making a Taylor expansion of k_1 to first order in x_{0A} about $\bar{x}_{0A}$ we obtain.

$$\dot{\delta x_1} = k_1(\bar{x_{0A}}) + \left.\frac{dk_1(x_{0A})}{dx_{0A}}\right|_{\bar{x_{0A}}} \delta x_{0A} - \gamma(\delta x_1 + \bar{x_1}) + \mu_1 + \xi_1, \tag{4.3.14}$$

but from eq. (4.3.11), $\bar{x_1} = {}^{k_1(\bar{x_{0A}})}/_{\gamma}$, therefore

$$\dot{\delta x_1}(t) = c_1\delta x_{0A} - \gamma\delta x_1 + \mu_1\xi_1, \tag{4.3.15}$$

where $c_1 \coloneqq {}^{dk_1(x_{0A})}/_{dx_{0A}}\big|_{\bar{x_{0A}}}$ Fourier transforming and solving for $\hat{\delta x_1}$ we get

$$\hat{\delta x_1} = \frac{c_1\hat{\delta x_{0A}} + \hat{\mu_1} + \hat{\xi_1}}{\gamma + i\omega}.$$

Taking the square norm and averaging

$$\begin{aligned}\left\langle |\hat{\delta x_1}|^2 \right\rangle &= \frac{1}{\omega^2+\gamma^2}\left(c_1\hat{\delta x_{0A}} + \hat{\mu_1} + \hat{\xi_1}\right)\left(c_1\hat{\delta x_{0A}}^* + \hat{\mu_1}^* + \hat{\xi_1}^*\right) \\ &= \frac{1}{\omega^2+\gamma^2}\left(c_1^2\left\langle |\hat{\delta x_{0A}}|^2\right\rangle + c_1\left(\langle \hat{\delta x_{0A}}\hat{\xi_1}^*\rangle + \langle \hat{\delta x_{0A}}^*\hat{\xi_1}\rangle\right) + \left\langle |\hat{\mu_1}|^2\right\rangle + \left\langle |\hat{\xi_1}|^2\right\rangle\right)\end{aligned}$$

Using the Wiener-Khinchin theorem and the equations for the correlations we get

$$\left\langle |\hat{\mu_1}|^2 \right\rangle = 2\gamma(b_1+1)\bar{x_1},$$
$$\left\langle |\hat{\xi_1}|^2 \right\rangle = 2\gamma\eta_G^2 {\bar{x_1}}^2,$$

since the Fourier transform of the Dirac delta is 1. Usually the binding and unbinding of inducers such as $IPTG$ occurs at timescales that are much smaller than the timescales of transcription and translation. Therefore, time averaging makes those fluctuations negligible. With this in mind, we will assume that the noise in x_{0A} are the same as the noise in x_0. Then from eqs. (4.3.7) and (4.3.9) we get

$$\left\langle |\delta\hat{x_{0A}}|^2 \right\rangle = \frac{2\gamma {\bar{x_0}}^2 \left({}^{(b_0+1)}/_{\bar{x_0}} + \eta_G^2\right)}{\gamma^2+\omega^2},$$
$$\langle \delta\hat{x_{0A}}\hat{\xi_1}^* \rangle = \frac{1}{\gamma+i\omega}\left(\langle \hat{\mu_0}\hat{\xi_1}^* \rangle + \langle \hat{\xi_0}\hat{\xi_1}^* \rangle\right) = \frac{\langle \hat{\xi_0}\hat{\xi_1}^* \rangle}{\gamma+i\omega}$$
$$\langle \delta\hat{x_{0A}}^*\hat{\xi_1} \rangle = \frac{1}{\gamma-i\omega}\left(\langle \hat{\mu_0}^*\hat{\xi_1} \rangle + \langle \hat{\xi_0}^*\hat{\xi_1}^* \rangle\right) = \frac{\langle \hat{\xi_0}^*\hat{\xi_1} \rangle}{\gamma-i\omega}$$

Where the last step in the last two equations comes from the Wiener-Khinchin theorem and eq. (4.3.12). Replacing the previous equations in eq. (4.3) and taking the inverse transform we get for the variance

$$\begin{aligned}\langle \delta x_1^2 \rangle = {} & 2\gamma {\bar{x_0}}^2 c_1^2 \left({}^{(b_0+1)}/_{\bar{x_0}} + \eta_G^2\right)\frac{1}{2\pi}\int_{-\infty}^{\infty}\frac{d\omega}{(\omega^2+\gamma^2)^2} \\ & + 2\gamma\eta_G^2\bar{x_0}\bar{x_1}c_1\frac{1}{2\pi}\left(\int_{-\infty}^{\infty}\frac{d\omega}{(\gamma+i\omega)(\omega^2+\gamma^2)} + \int_{-\infty}^{\infty}\frac{d\omega}{(\gamma-i\omega)(\omega^2+\gamma^2)}\right) \\ & + 2\gamma {\bar{x_1}}^2\left({}^{(b_1+1)}/_{\bar{x_1}} + \eta_G^2\right)\frac{1}{2\pi}\int_{-\infty}^{\infty}\frac{d\omega}{\omega^2+\gamma^2}.\end{aligned}$$

Solving the integrals in the complex plane and rearranging

$$\langle \delta x_1^2 \rangle = \frac{c_1^2 \bar{x_0}^2}{2\gamma^2} \left({}^{(b_0+1)}/_{\bar{x_0}} + \eta_G^2 \right) + \frac{c_1 \eta_G^2 \bar{x_0} \bar{x_1}}{\gamma} + \bar{x_1}^2 \left({}^{(b_1+1)}/_{\bar{x_1}} + \eta_G^2 \right).$$

Dividing by $\bar{x}_1^2$ yields

$$\eta_1^2 = \frac{1}{2} \frac{c_1^2 \bar{x_0}^2}{\gamma^2 \bar{x}_1^2} \left(\frac{(b_0+1)}{\bar{x_0}} + \eta_G^2 \right) + \frac{c_1 \bar{x_0}}{\gamma \bar{x}_1} \eta_G^2 + \frac{(b_1+1)}{\bar{x_1}} + \eta_G^2. \tag{4.3.16}$$

Recall the definition of the logarithmic gain from eq. (3.4.1)

$$H_{10} = \frac{\partial \ln(\langle J_1^- \rangle / \langle J_1^+ \rangle)}{\partial \ln \langle x_0 \rangle}.$$

From eq. (4.3.11), we have $\langle J_1^+ \rangle = k_1(\bar{x}_{0A})$ and $\langle J_i^- \rangle = \gamma \bar{x}_1$. Then,

Aqui que con el promedio si la func es nonlinear

$$H_{10} = \frac{\partial \ln \left(\frac{\gamma \bar{x}_1}{k1(\bar{x}_{0A})} \right)}{\partial \ln \bar{x}_0} = - \frac{\frac{\gamma \bar{x}_1}{k_1(\bar{x}_{0A})} \partial \left(\frac{k_1(\bar{x}_{0A})}{\gamma \bar{x}_1} \right)}{\frac{\partial \bar{x}_{0A}}{\bar{x}_{0A}}} = - \frac{\bar{x}_{0A}}{k_1(\bar{x}_{0A})} \frac{\mathrm{d}k_1(\bar{x}_{0A})}{\mathrm{d}\bar{x}_{0A}} = - \frac{c_1 \bar{x}_{0A}}{\gamma \bar{x}_1}.$$

Since gene 2 has the same dependence on the number of active proteins of gene 1 (not bound to ATC), we have in general

$$H_{ij} = \frac{c_i \bar{x}_{jA}}{\gamma \bar{x}_i} = \frac{\bar{x}_{jA}}{\gamma \bar{x}_i} \frac{\mathrm{d}k_i(\bar{x}_{jA})}{\mathrm{d}\bar{x}_{jA}}, \quad \text{for} \quad (i,j) = \{(1,0),(2,1)\}. \tag{4.3.17}$$

Replacing this in eq. (4.3.16) we obtain

$$\begin{aligned} \eta_1^2 &= \eta_{1\text{int}}^2 + \frac{1}{2} H_{10}^2 \eta_0^2 + \eta_G^2 (1 - H_{10}) \\ &= \eta_{1\text{int}}^2 + \frac{1}{2} H_{10}^2 \eta_{0\text{int}}^2 + \eta_G^2 \left(1 + \frac{1}{2} H_{10}^2 - H_{10} \right), \end{aligned} \tag{4.3.18}$$

where $\eta_{1\text{int}}^2 = \frac{(b_1+1)}{\bar{x_1}}$ and η_0^2 is given by eq. 4.3.10.

This corroborates what was said at the end of chapter 3. The total noise in gene

one is given by the intrinsic part, the noise from gene 0 that is propagated to gene 1 (including both its intrinsic and global part) and the global noise that enters directly into gene 1. The factor of 1/2 represents the time averaging since we assumed equal degradation rates for both proteins.

For gene 2 we proceed similarly, with analogous statistics for the noise terms, the resulting noise is

$$\begin{aligned}\eta_2^2 &= \eta_{2\text{int}}^2 + \frac{1}{2}H_{21}^2\eta_{1\text{int}}^2 + \frac{3}{8}H_{21}^2H_{10}^2\eta_{0\text{int}}^2 \\ &\quad + \eta_G^2\left(1 + \frac{1}{2}H_{21} + \frac{3}{8}H_{21}^2H_{10}^2 - H_{21} - \frac{3}{4}H_{21}^2H_{10} + \frac{1}{2}H_{21}H_{10}\right), \\ &= \eta_{2\text{int}}^2 + \frac{1}{2}H_{21}^2\eta_1^2 + \frac{1}{8}H_{21}^2H_{10}^2\eta_0^2 + \eta_G^2\left(1 - H_{21} - \frac{1}{4}H_{21}^2H_{10} + \frac{1}{2}H_{21}H_{10}\right).\end{aligned} \tag{4.3.19}$$

Which contains the intrinsic noise of gene 2, the contribution from the total noise of gene 1, the contribution from the total noise of gene 0 that is transmitted first to gene 1 and then to gene 2 and the global noise that enters directly. Notice that the time average factor for a two step propagation of intrinsic noise is less than 1/2, verifying the conclusions of the previous section.

Since gene 3 has also a constitutive promoter, its noise is given by

$$\eta_3^2 = \eta_{3\text{int}}^2 + \eta_G^2. \tag{4.3.20}$$

This gene was used to measure the strenght of the global noise η_G since it is not connected with other components and thus it does not receives propagated noise. The correlations can be found in a very similar way. They are given by

$$
\begin{aligned}
C_{12} &= -\frac{1}{2}H_{21}\eta_{1\text{int}}^2 - \frac{3}{8}H_{21}H_{10}^2\eta_{0\text{int}}^2 + \eta_G^2\left(1 - \frac{1}{2}H_{21} - \frac{1}{2}H_{10} - \frac{3}{8}H_{21}H_{10}^2 + \frac{3}{4}H_{21}H_{10}\right), \\
C_{13} &= \eta_G^2\left(1 - \frac{1}{2}H_{10}\right), \\
C_{23} &= \eta_G^2\left(1 - \frac{1}{2}H_{21} + \frac{1}{4}H_{21}H_{10}\right).
\end{aligned}
\tag{4.3.21}
$$

From the results it is clear that the noise is strongly dependent on the details of the network and on how is the interaction between the different components of the genetic circuit. A detailed analysis of these results will be done in the next section.

4.4 Explicit expressions for the logarithmic gain

The repressions between the genes of the cascade are mathematically represented writing the rates of creation as Hill type functions. They are given by

$$
\begin{aligned}
k_1(y_{0A}) &= \alpha_1 + \frac{\beta_1}{1 + (\frac{y_{0A}}{K_{d1}})^{h_1}}, \\
k_2(y_{1A}) &= \alpha_2 + \frac{\beta_2}{1 + (\frac{y_{1A}}{K_{d2}})^{h_2}}.
\end{aligned}
\tag{4.4.1}
$$

The fraction of y_0 ($LacI$) and y_1 ($tetR$) that are unbound to ATC is also modeled as a Hill function i.e. $y_{0A} = k_0(IPTG)y_0$ and $y_{1A} = k_a(ATC)y_1$ where

$$
\begin{aligned}
k_0(IPTG) &= \alpha_0 + \frac{\beta_0}{1 + (\frac{IPTG}{K_{d0}})^{h_0}}, \\
k_2(y_{1A}) &= \alpha_a + \frac{\beta_a}{1 + (\frac{ATC}{K_{da}})^{h_a}}.
\end{aligned}
\tag{4.4.2}
$$

To find the logarithmic gains H_{21} and H_{10}, we use eq. (4.3.17) and eq. (4.4.1) to

obtain after differentiating for $(i,j) = \{(1,0),(2,1)\}$

$$H_{ij} = \frac{\bar{x}_{jA}}{\gamma \bar{x}_i} \frac{\mathrm{d}}{\mathrm{d}\bar{x}_{jA}} \left(\alpha_i + \frac{\beta_i}{1 + (\frac{\bar{x}_{jA}}{K_{di}})^{h_i}} \right) = \frac{h_i}{\beta_i k_i(\bar{x}_{jA})} \left(k_i(\bar{x}_{jA}) - \alpha_i \right) \left(\alpha_i + \beta_i - k_i(\bar{x}_{jA}) \right). \tag{4.4.3}$$

The values of the parameters that will be used in the following calculations and graphics can be found in github [3].

Replacing eqs. (4.4.1) - (4.4.3) we obtain explicit expressions for the logarithmic gains that, given the values of the parameters can be replaced in the eqs. for the noises and correlations. Fig. 4.4.1 shows the logarithmic gains as a function of [IPTG] when [ATC] = 0.

Figure 4.4.1: . Plot of H_{21} and H_{10} as a function of [IPTG] given by eq. (4.4.3) . The vertical line marks the maximum of H_{21}.

The maximum value in H_{21} represents the concentration of [IPTG] at which y_2 is more affected by changes in y_1. This corresponds to the point of the transfer function where the slope is greater.

Figure 4.4.2 shows the noises and correlations for certain set of values of the parameters as a function of the concentration of IPTG when there is no ATC.

Figure 4.4.2: . Plot of noises (eqs. (4.3.20) and (4.3.20)) and correlations ((4.3.21)) as a function of [IPTG] for genes 1 to 3 with [ATC]=0. The vertical line shows the IPTG concentration at which H_{21} reaches its greater value.

The noises and correlations behave in a nonintuitive manner. For instance, η_1 and η_2 are very different altough they are both regulated by an upstream component. The correlations are also very different for each pair of genes. Recalling that gene 3 is not

[3]The equations were solved and plotted using Wolfram Mathematica. The notebook, which also contains the values for the parameters can be found in `https://github.com/gutiloluis/61Monograph/blob/master/math/lan-analytic_solver.nb`. Most of the parameters were fitted experimentally by Pedraza and van Oudenaarden [3]

regulated in the cascade, it is expected that the correlations C_{13} and C_{23} are independent of IPTG. The graphics show that this is not the case. The noise in η_3 is constant as intuitively expected.

Another important aspect to notice is that altough there are components that do not interact directly, in the expression for the noises and correlations the interaction between them is clear. For instance, in eq. (4.3.19) can be observed that η_2 depends explicitely on H_{10} and on the intrinsic noise of gene 0. For this reason, noise in a regulation cascade can not be trivially added. All the upstream components contribute to the downstream components with a strenght that depends of the time average and the way the genes interact, measured by the logarithmic gain.

The vertical line in the graphic for η_2 shows that at the IPTG concentration at which H_{21} is larger, the maximum of the noise almost coincides with this level. A larger value of H indicates that x_2 is more sensitive to fluctuations in x_1. This explains the correspondent peak in η_2. The match is not exact due to the additional dependences in eq. (4.3.19).

From eqs. (4.3.18) and (4.3.19) it can be seen that the intrinsic propagated noise always depends on the square of the logarithmic gain while the global propagated noise has terms that depend on its sign. This is a consequence of the correlation of the global noise among all the genes. For instance, a sudden change in all the creation rates caused by global noise will, for instance, increase both x_1 and x_2 expressions. Since x_1 represses x_2 this will also decrease the expression of x_2 proportional to the coupling between them (H). In the hypothetical case where x_1 activates x_2 the effect would be that the expression of x_2 would be increased further. In this way the correlations determine how global noise is modulated. The sign dependence on the global propagated noise is also explained by this example.

The previous analysis also explains the non constant correlation C_{13} and C_{23}. Global fluctuations to gene 3 have only the direct component, while to gene 1 or 2 have both the direct and transmitted component that depends on the interactions with upstream

genes. Global noise will thus be reduced of increased in genes 1 and 2 with respect to the unregulated gene 3 in an amount that depends on the signs and strenghts of the log. gains. This yields correlations that varies with iPTG.

Fig. 4.4.3 shows each of the components of η_2. The intrinsic squared noise varies as the inverse of the mean. When [IPTG] is large, x_{0A}, making x_1 large and thus x_2 low. Therefore, the intrinsic noise should be larger at large IPTG concentration. The transmitted intrinsic noise corresponds approximately to the square of the logarithmic gain times the noise in the upstream components (including both genes 0 and 1). Its maximum corresponds roughly to the maximum in H_{21} (vertical dashed line). The global component depends also on the correlations as explained above. Comparing the total noise with the components it is clear how important are the network interactions in determining the noise.

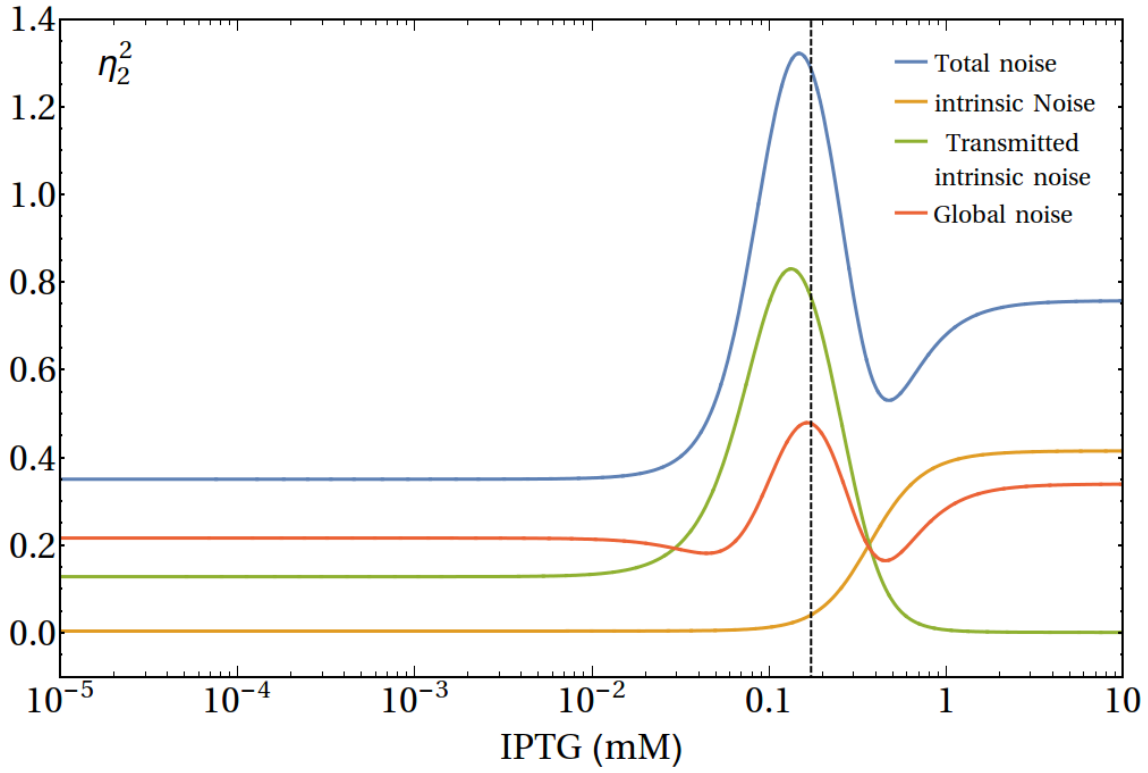

Figure 4.4.3: . Plot of each component of η_2^2 (eq. (4.3.19)) as a function of [IPTG] with [ATC]=0. The vertical line shows the IPTG concentration at which H_{21} reaches its greater value.

The coupling between genes 1 and 2 is regulated with ATC. If [ATC] is large, the coupling is low and viceversa. Fig. 4.4.4 shows how η_2 and C_{12} change as a function of IPTG for different concentrations of ATC.

Figure 4.4.4: . Plots of η_2 (eq. (4.3.19)) and C_{12} (eq. (4.3.21)) as a function of [IPTG] for different ATC concentration shown in the legend.

This reconfirms the high sensitivity of the noise to the coupling of the network. For small changes in concentration of ATC, the noise shows huge changes in its behavior. At a concentration of [ATC] = 1 mM (fig. 4.4.4, red curve), the only considerable contributions to η_2 are the intrinsic and the directly recieved global noise.

4.5 Conclusions and implications

Fig. 4.5.1 summarizes how noise propagates through a cascade of regulation. Each gene has its own intrinsic noise and recieves contributions to its total noise directly from correlated global sources, and indirectly from propagated global and intrinsic fluctuations of upstream genes. The effect of the propagated noise is modulated by the logarithmic gains between the directly connected genes. Altough not shown in the figure, the time average also regulates the transmitted noise.

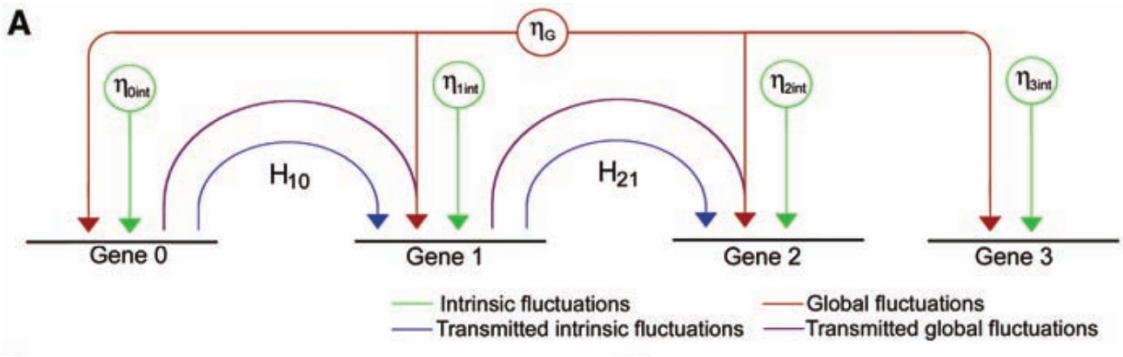

Figure 4.5.1: Different sources of noise and their propagation along the cascade of regulation (from [3]).

The limitations of the model are related to the ones for the ME approach. It is also a stationary and linearized model that does not account for the time dynamics of fluctuations. In spite of that, it is remarkable how the Langevin approach enabled the authors to write the expressions for the noises in a way that is very easy to interpret. Also, the match of the experimental results with the model is excellent [3] [18].

We have introduced the three basic approaches that have been used to model noise in genetic circuits: the master equation, the fluctuation-dissipation theorem and the Langevin equation. The following chapter will focus on analyzing how noise behaves when some additional factors are considered.

Chapter 5

Effects of bursting and senescence

There are many different phenomena that could have an effect on noise. Most of the models that have been based on assumptions and have made fits of the data obatined with fluorescent proteins according to those assumptions. Nevertheless, we will see in this chapter that noise coming from different mechanisms could have the same general behaviors, making it difficult to predict characteristics of the systems according to their noise.

We will use the Fluctuation-Dissipation theorem as we used it on section FILL to analyize the effects of bursting (the synthesis of several molecules per creation event) and senescence (the degradation of molecules in many exponential steps) on noise.

This chapter is based on the work done by J. M. Pedraza and J. Paulsson in [8].

TODO: Explain qualitatively/biologically how could bursting and senescence happen, include images.

5.1 mRNA bursts

Let the mRNA be produced with bursts of random size b, the degradation and protein creation is done one at a time with exponential waiting times (single-step Poisson processes). In this case the only modification with respect to the "standard model" (eqs.

(2.1.1) and (2.1.2)) is the D_{11} term of the matrix $\mathbf{D}$, which by definition is

$$D_{11} = \frac{1}{\langle n_1 \rangle^2} \sum_k (s_1^k)^2 r_k(\mathbf{n}). \tag{5.1.1}$$

All the possible k reactions include all the creation bursts and the reaction of degradation, which has rate $\langle n_1 \rangle / \tau_1$ and $s_1 = -1$, therefore

$$\sum_k (s_1^k)^2 r_k = \frac{\langle n_1 \rangle}{\tau_1} + \sum_{k'} (s_1^{k'})^2 r_{k'}. \tag{5.1.2}$$

Where now the index k' runs over all the synthesis reactions only. We can rewrite the second term as [1]

$$\sum_k (s_1^k)^2 r_k = \sum_k r_k \sum_k \left(\frac{r_k}{\sum_k r_k} \right) (s_1^k)^2,$$

where the sum over the term in parentheses results in 1. This term can be interpreted as the probability that the upcoming reaction turns out to be the k^{th} one. Writing it as ρ_k this yields

$$\sum_k (s_1^k)^2 r_k = \sum_k r_k \sum_k \rho_k (s_1^k)^2,$$

but s_1^k is the burst size for the k^{th} synthesis reaction. Therefore the inner sum of the previous equation is actually an average over all the possible burst sizes, hence

$$\sum_k (s_1^k)^2 r_k = \sum_k r_k \langle b^2 \rangle = \left(\langle b \rangle^2 + \sigma_b^2 \right) \sum_k r_k, \tag{5.1.3}$$

and using a similar trick we get

[1] For simplicity, we used the dummy index k on the following calculations, keeping in mind that it only runs along the synthesis reactions.

$$\sum_k r_k = \sum_k r_k s_1^k \frac{\sum_k r_k}{\sum_k r_k s_1^k} = \sum_k r_k s_1^k \left(\sum_k \left(\frac{r_k}{\sum_k r_k} \right) s_1^k \right)^{-1}$$
$$= \sum_k r_k s_1^k \left(\sum_k \rho_k s_1^k \right)^{-1} = \frac{1}{\langle b \rangle} \sum_k r_k s_1^k.$$

Assuming the system is in steady state, all the synthesis rates equal the degradation ones. Therefore

$$\sum_k r_k s_1^k = \frac{\langle n_1 \rangle}{\tau_1},$$

obtaining

$$\sum_k r_k = \frac{\langle n_1 \rangle}{\langle b \rangle \tau_1}.$$

Replacing this on eq. (5.1.3), and then on eq. (5.1.2) and (5.1.1) we get

$$D_{11} = \frac{1}{\langle n_1 \rangle^2} \left(\frac{\langle n_1 \rangle}{\tau_1} + \frac{\langle n_1 \rangle}{\langle b \rangle \tau_1} \left(\langle b \rangle^2 + \sigma_b^2 \right) \right) = \frac{1}{\tau_1 \langle n_1 \rangle} \left(1 + \langle b \rangle \left(1 + \frac{\sigma_b^2}{\langle b \rangle^2} \right) \right) \tag{5.1.4}$$

The matrices then become

$$\mathbf{D} = \begin{pmatrix} D_{11} & 0 \\ 0 & \frac{2}{\tau_2 \langle n_2 \rangle} \end{pmatrix}, \quad \mathbf{M} = \begin{pmatrix} \frac{1}{\tau_1} & 0 \\ -\frac{1}{\tau_2} & \frac{1}{\tau_2} \end{pmatrix}.$$

With D_{11} given by eq. (5.1.4) And solving the linear system $\mathbf{M}\eta + \eta\mathbf{M}^T + \mathbf{D} = 0$ (see sec. 3.4), we obtain the following expression for the noise in the proteins

$$\boxed{\eta_{22} = \frac{1}{\langle n_2 \rangle} + \frac{1}{\langle n_1 \rangle} \frac{\tau_1}{\tau_1 + \tau_2} \frac{\langle b \rangle \left(1 + \sigma_b^2 / \langle b \rangle^2 \right) + 1}{2}}. \tag{5.1.5}$$

TODO: Analysis of the equations, independence on details of the distribution

5.2 Arbitrary distribution of creation times

Suppose a creation event happened at $t = 0$, and let $f(t)$ be the probability density function of a creation event happening at time t (meaning a time t after the last event), i.e. $P(t \in [T, T + dt]) = f(T)dt$. According to that, the following properties are satisfied

$$P(n = 0|t = T) = P(t > T) = 1 - P(t < T) = 1 - F(T)$$

where n stands for the number of creation events and F is the cumulative distribution function associated to $f(t)$. Also, for one creation event to have happened before time $t = T$, there must be a creation on a time t_1 such that $0 < t_1 < T$ and no creation events on the remaining $(T - t_1)$ time. This leads to the following equation

$$\begin{aligned} P(n = 1|t = T) &= \int_0^T P(t = t_1)P(t > T - t_1)\mathrm{d}t_1 \\ &= \int_0^T f(t_1)(1 - F(T - t_1))\mathrm{d}t_1 = f * (1 - F)|_T, \end{aligned}$$

where the asterisk denotes the convolution product. Following a similar argument, we obtain for an arbitrary number of events

$$\begin{aligned} P(n = N|t = T) &= f * P(n = N - 1|t)|_T = f * f * P(n = N - 2|t)|_T \\ &= \cdots = \underbrace{f * \cdots * f}_{n \text{ times}} * P(n = 0, t)|_T = \underbrace{f * \cdots * f}_{n \text{ times}} * (1 - F)|_T. \end{aligned} \tag{5.2.1}$$

Since the convolutions are difficult to deal with, we will use the Laplace transform and solve on Laplace space. The property that $\mathcal{L}(f * g) = \hat{f} \cdot \hat{g}$, where $\hat{f} \coloneqq \mathcal{L}(f)$ will make the problem much simpler.

Aplying $\mathcal{L}$ to eq. (5.2.1) we get

$$\hat{P}(n,s) = \hat{f}^n(s)\mathcal{L}(1-F)(s).$$

It can be easily shown that $\hat{F} = \hat{f}/s$ and $\hat{1} = 1/s$ yielding

$$\hat{P}(n,s) = \frac{1}{s}\hat{f}^n(s)(1-\hat{f}(s)). \tag{5.2.2}$$

To find the moments and the noise, we will use the moment generating function, as defined on (1.6.1). It will be denoted as $G(z,s)$.

$$\hat{G}(z,s) = \sum_{n=0}^{\infty} z^n \hat{P}(n,s) = \frac{1}{s}(1-\hat{f}(s))\sum_{n=0}^{\infty}(z\hat{f})^n = \frac{1-\hat{f}(s)}{s(1-z\hat{f}(s))}. \tag{5.2.3}$$

Where the geometric series convereges in this case because $\hat{f}(s) \leq 1$ and we will evaluate z at 1. The first and second derivatives of G in Laplace space are given by

$$\langle\hat{n}\rangle(s) = \left.\frac{\partial\hat{G}(z,s)}{\partial z}\right|_1 = \frac{\hat{f}(s)}{s(1-\hat{f}(s))}. \tag{5.2.4}$$

$$\langle\hat{n}(\hat{n}-1)\rangle(s) = \left.\frac{\partial^2\hat{G}(z,s)}{\partial z^2}\right|_1 = \frac{2}{s}\left(\frac{\hat{f}(s)}{1-\hat{f}(s)}\right)^2. \tag{5.2.5}$$

It could also be proven that

$$\hat{f}(0) = 1, \quad \left.\frac{\mathrm{d}\hat{f}(s)}{\mathrm{d}s}\right|_0 = -\langle t\rangle, \quad \left.\frac{\mathrm{d}^2\hat{f}(s)}{\mathrm{d}s^2}\right|_0 = \langle t^2\rangle. \tag{5.2.6}$$

Therefore applying the inverse Laplace transform to eqs. (5.2.4) and (5.2.5), and using properties (5.2.6) we could obtain the moments.

$$\langle n\rangle(t) = \mathcal{L}^{-1}(\langle\hat{n}\rangle(s)) = \frac{1}{2i\pi}\oint e^{st}\frac{\hat{f}(s)}{s(1-\hat{f}(s))}\mathrm{d}s \tag{5.2.7}$$

The integral can be solved by residues. Since $\hat{f}(0) = 1$, there is a pole of order 2 in

$s = 0$. To find the residues of a pole c of order m of the function f, the residue is given by

$$\mathrm{Res}_c(f) = \frac{1}{(m-1)!} \lim_{z \to c} \frac{\mathrm{d}^{m-1}}{\mathrm{d}z^{m-1}}((z-c)^m f(z)). \tag{5.2.8}$$

Then

$$\begin{aligned}\langle n \rangle(t) &= \mathrm{Res}_0 \frac{e^{st}}{s} \frac{\hat{f}(s)}{1-\hat{f}(s)} = \lim_{s \to 0} \frac{\mathrm{d}}{\mathrm{d}s} \frac{s e^{st} \hat{f}(s)}{1-\hat{f}(s)} \\ &= \lim_{s \to 0} \frac{e^{st}}{(\hat{f}(s)-1)^2} \left[(1+st)(\hat{f}(s)-1)\hat{f}(s) + s\hat{f}'(s) \right].\end{aligned}$$

To find the limit we have to apply L'Hospital rule twice, after some algebra it yields

$$\langle n \rangle(t) = \frac{\hat{f}''(0)}{2(\hat{f}'(0))^2} - \frac{t}{\hat{f}'(0)} - 1 = \frac{t}{\langle t \rangle} + \left(\frac{\langle t^2 \rangle}{2\langle t \rangle^2} - 1 \right). \tag{5.2.9}$$

Now inverting eq. (5.2.5) we obtain

$$\begin{aligned}\langle n(n-1) \rangle(t) &= \frac{1}{2i\pi} \oint e^{st} \frac{2}{s} \left(\frac{\hat{f}(s)}{1-\hat{f}(s)} \right)^2 \mathrm{d}s = \mathrm{Res}_0 \frac{2}{s} \left(\frac{\hat{f}(s)}{1-\hat{f}(s)} \right)^2 \\ &= \lim_{s \to 0} \frac{\mathrm{d}^2}{\mathrm{d}s^2} e^{st} \left(\frac{s\hat{f}(s)}{1-\hat{f}(s)} \right)^2\end{aligned}$$

Where we used eq. (5.2.8) to find the residue of a pole of order 3. After doing the necessary algebra and the L'Hospital rule three times we get

$$\langle n(n-1) \rangle(t) = \frac{t^2}{\langle t \rangle^2} + \frac{4t}{\langle t \rangle} \left(\frac{\langle t^2 \rangle}{2\langle t \rangle^2} - 1 \right) + 2 \left(1 - \frac{\langle t^2 \rangle}{\langle t \rangle^2} + \frac{3\langle t^2 \rangle^2}{4\langle t \rangle^4} + \frac{\langle t^3 \rangle}{3\langle t \rangle^3} \right)$$

And combining with eq. (5.2.9) we obtain the variance

$$\sigma_n^2(t) = \frac{t}{\langle t\rangle}\left(\frac{\langle t^2\rangle}{\langle t\rangle^2} - 1\right) + \left(-\frac{\langle t^2\rangle}{2\langle t\rangle^2} + \frac{5\langle t^2\rangle^2}{4\langle t\rangle^4} - \frac{2\langle t^3\rangle}{3\langle t\rangle^3}\right). \tag{5.2.10}$$

Analysis, (compare with the exponential case) and explain why we ignore the other part

Ignoring the second terms in parentheses in eqs. 5.2.9 and 5.2.10 we get

$$\langle n\rangle = \frac{t}{\langle t\rangle}, \tag{5.2.11}$$

$$\sigma_n^2 = \frac{t}{\langle t\rangle}\left(\frac{\langle t^2\rangle}{\langle t\rangle^2} - 1\right). \tag{5.2.12}$$

Rearranging eq. (5.2.12) and using eq. (5.2.11) we get

$$\sigma_n^2 = \frac{t}{\langle t\rangle}\left(\frac{\langle t^2\rangle - \langle t\rangle^2}{\langle t\rangle^2}\right) = \frac{t}{\langle t\rangle}\eta_t^2 = \langle n\rangle\eta_t^2.$$

Hence

$$\eta_n^2 = \frac{1}{\langle n\rangle}\eta_t^2. \tag{5.2.13}$$

Now we will include the effect of bursts of creation. Let n be the number of creation events on a given time interval (meaning the number of bursts, not the total number of molecules created), b_i be burst size for the i^{th} events. Both n and b_i are random variables, and each b_i follows the same probability distribution. Consider the random variable

$$x \coloneqq \sum_{i=0}^{n} b_i \tag{5.2.14}$$

representing the total number of molecules created on the given time interval. (It is a sum of a random number of random variables). We will denote the probability mass function of x as $P_x(x)$.

We will use the properties of the characteristic function $\phi(s)$ and find the moments using its properties (see sec. 1.7). From its definition

$$\phi(s) \coloneqq \langle e^{xs} \rangle_x = \sum_{a=0}^{\infty} e^{xs} P_x(x=a).$$

Using the total probability theorem, we can write it as

$$\begin{aligned} \phi(s) &= \sum_{a=0}^{\infty} e^{xs} \sum_{n=0}^{\infty} P_x(x=a|n) P(n) = \sum_{n=0}^{\infty} \left(\sum_{a=0}^{\infty} e^{xs} P_x(x=a|n) \right) P(n) \\ &= \sum_{n=0}^{\infty} \langle e^{xs} \rangle_{x|n} P(n) \end{aligned} \tag{5.2.15}$$

Where $\langle \quad \rangle_x$ denotes average with respect to the distribution of x. Using eq. (5.2.14) we get for $\langle e^{xs} \rangle_{x|n}$

$$\langle e^{xs} \rangle_{x|n} = \langle e^{s \sum_{i=0}^{n} b_i} \rangle_b = \langle \prod_{i=0}^{n} e^{sb_i} \rangle_b,$$

Notice that the average is now taken with respect to the distribution of burst sizes (denoted by $\langle \quad \rangle_b$) since we wrote the function in terms of that variable. Assuming independence of the burst sizes we get

$$\langle e^{xs} \rangle_{x|n} = \prod_{i=0}^{n} \langle e^{sb_i} \rangle_b,$$

and since all the b_is follow the same distribution, the product is in fact independent of i, yielding

$$\langle e^{xs} \rangle_{x|n} = \prod_{i=0}^{n} \langle e^{sb} \rangle_b = \langle e^{sb} \rangle_b^n$$

.

Replacing this result in eq. (5.2.15) we obtain

$$\phi(s) = \sum_{n=0}^{\infty} \langle e^{sb} \rangle_b^n P(n) = \left\langle \langle e^{sb} \rangle_b^n \right\rangle_n$$

Where $\langle \quad \rangle_n$ denotes average with respect to the distribution of events P_n.

We proceed to obtain the moments, using the properties of the characteristic function (eqs. (1.7.2) - (1.7.2))

$$\langle x \rangle = \left. \frac{\mathrm{d}\phi(s)}{\mathrm{d}s} \right|_0 = \left. \frac{\mathrm{d}}{\mathrm{d}s} \left\langle \langle e^{bs} \rangle_b^n \right\rangle_n \right|_0 = \left\langle n \langle e^{bs} \rangle_b^{n-1} \langle b e^{bs} \rangle_b \right\rangle_n \Big|_0 = \langle n \langle b \rangle_b \rangle_n = \langle n \rangle_n \langle b \rangle_b, \quad (5.2.16)$$

which is intuitive: the average number of molecules produced is the average number of bursts times the average burst size. For the second moment we have

$$\begin{aligned} \langle x^2 \rangle &= \left. \frac{\mathrm{d}^2\phi(s)}{\mathrm{d}s^2} \right|_0 = \left. \frac{\mathrm{d}\phi(s)}{\mathrm{d}s} \left\langle n \langle e^{bs} \rangle_b^{n-1} \langle b e^{bs} \rangle_b \right\rangle_n \right|_0 \\ &= \left\langle n(n-1) \langle e^{bs} \rangle_b^{n-2} \langle b e^{bs} \rangle_b^2 + n \langle e^{bs} \rangle_b^{n-1} \langle b^2 e^{bs} \rangle_b \right\rangle_n \Big|_0 \\ &= \langle n^2 \rangle_n \langle b \rangle_b^2 - \langle n \rangle_n \langle b \rangle_b^2 + \langle n \rangle_n \langle b^2 \rangle_b. \end{aligned}$$

Using the previous result with eq. (5.2.16) to find the variance we differentiate twice

$$\begin{aligned} \sigma_x^2 &= \langle n^2 \rangle_n \langle b \rangle_b^2 - \langle n \rangle_n \langle b \rangle_b^2 + \langle n \rangle_n \langle b^2 \rangle_b - \langle n \rangle_n^2 \langle b \rangle_b^2 \\ &= \langle b \rangle_b^2 \left(\langle n^2 \rangle_n - \langle n \rangle_n^2 \right) + \langle n \rangle_n \left(\langle b^2 \rangle_b - \langle b \rangle_b^2 \right) \\ &= \langle b \rangle_b^2 \sigma_n^2 + \langle n \rangle_n \sigma_b^2. \end{aligned}$$

Dividing by $\langle x \rangle^2 = \langle n \rangle_n^2 \langle b \rangle_b^2$,

$$\eta_x^2 = \frac{\sigma_n^2}{\langle n \rangle_n^2} + \frac{\sigma_b^2}{\langle n \rangle_n \langle b \rangle_b^2} = \eta_n^2 + \frac{1}{\langle n \rangle_n} \eta_b^2.$$

Using eq. 5.2.13 we obtain

$$\eta_x^2 = \frac{1}{\langle n \rangle} \left(\eta_t^2 + \eta_n^2 \right) = \frac{\langle b \rangle \left(\eta_t^2 + \eta_n^2 \right)}{\langle x \rangle}, \quad (5.2.17)$$

where

$$\langle x \rangle = \langle b \rangle \frac{t}{\langle t \rangle}. \tag{5.2.18}$$

This result holds for a pure birth process.

5.3 Decay of molecules

We include the effects of decay of molecules considering a binomial partitioning during cell division. Let $P_{\mathrm{Dr}}(l|m)$ be the probability of finding l molecules in volume fraction r given that there are m molecules before division.

Assuming that each molecule segregates independently[2], and that the probability of arriving at a volume is proportional to it we obtain a binomial distribution

$$P_{\mathrm{Dr}}(l|m) = \binom{m}{l} r^l (1-r)^{m-l}. \tag{5.3.1}$$

Let $P_{\mathrm{Br}}(m)$ and $P_{\mathrm{Ar}}(m)$ be the probabilities of having m molecules before and after division, respectively, for a fixed volume fraction r. We have then

$$P_{\mathrm{Ar}}(l) = \sum_{m=0}^{\infty} P_{\mathrm{Dr}}(l|m) P_{\mathrm{Br}}(m) = \sum_{m=0}^{\infty} \binom{m}{l} r^l (1-r)^{m-l} P_{\mathrm{Br}}(m). \tag{5.3.2}$$

Multiplying by z^l and summing we get the moment generating function $G_{\mathrm{Ar}}(z)$

$$\begin{aligned} G_{\mathrm{Ar}}(z) &= \sum_{l=0}^{\infty} z^l \sum_{m=0}^{\infty} \binom{m}{l} r^l (1-r)^{m-l} P_{\mathrm{Br}}(m) \\ &= \sum_{m=0}^{\infty} \left(\sum_{l=0}^{\infty} (zr)^l (1-r)^{m-l} \right) P_{\mathrm{Br}}(m). \\ &= \sum_{m=0}^{\infty} (zr + 1 - r)^m P_{\mathrm{Br}}(m), \end{aligned} \tag{5.3.3}$$

[2]We will see in chapter 6 that segregation could be much more complicated than that.

where we used the binomial expansion formula on the last step.

The number of molecules of at the end of a growth stage (following P_{Br}) equals the number of molecules at the beginning (following P_{Ar}) plus the number of molecules created during the cycle (following $P_{x,\tau} \coloneqq P_x|_{t=\tau}$). Assuming both random variables as independent and recalling that the probability mass function of the sum of random variables is the convolution of the individual PMFs we have

TODO: Again, add a section in concepts about sum of random variables and convolution product

$$P_{\mathrm{Br}}(m) = P_{\mathrm{Ar}} * P_{x,\tau}(m)$$

Therefore,

$$G_{\mathrm{Br}}(z) = G_{\mathrm{Ar}}(z) G_{x,\tau}(z). \tag{5.3.4}$$

Finding the moments by using the properties of G and the previous equation we obtain

$$\langle n \rangle_{\mathrm{Br}} = \left.\frac{\partial G_{\mathrm{Br}}(z)}{\partial z}\right|_1 = G_{\mathrm{Ar}}(1) \left.\frac{\partial G_{x,\tau}(z)}{\partial z}\right|_1 + \left.\frac{\partial G_{\mathrm{Ar}}(z)}{\partial z}\right|_1 G_{x,\tau}(1) = \langle m \rangle_{x,\tau} + \langle m \rangle_{\mathrm{Ar}},$$

from eq. 5.3.3

$$\begin{aligned}
\left.\frac{\partial G_{\mathrm{Ar}}(z)}{\partial z}\right|_1 = \langle m \rangle_{\mathrm{Ar}} &= \left.\frac{\partial}{\partial z} \sum_{m=0}^{\infty} (zr + 1 - r)^m P_{\mathrm{Br}}(m)\right|_1 \\
&= \left.\sum_{m=0}^{\infty} m(zr + 1 - 1)^{m-1} r P_{\mathrm{Br}}(m)\right|_1 \\
&= r \sum_{m=0}^{\infty} m P_{\mathrm{Br}}(m) = r \langle m \rangle_{\mathrm{Br}}
\end{aligned}$$

Therefore

$$\langle m\rangle_{\mathrm{Br}} = \frac{1}{1-r}\langle m\rangle_{x,\tau} \qquad \langle m\rangle_{\mathrm{Ar}} = r\langle n\rangle_{\mathrm{Br}} = \frac{r}{1-r}\langle m\rangle_{x,\tau}. \tag{5.3.5}$$

The variances are obtained by taking the second derivative

$$\begin{aligned}\langle m(m-1)\rangle_{\mathrm{Br}} &= \left.\frac{\partial^2 G_{\mathrm{Br}}(z)}{\partial z^2}\right|_1 \\ &= G_{\mathrm{Ar}}(1)\left.\frac{\partial^2 G_{x,\tau}(z)}{\partial z^2}\right|_1 + \left.\frac{\partial^2 G_{\mathrm{Ar}}(z)}{\partial z^2}\right|_1 G_{x,\tau}(1) + 2\left.\frac{\partial G_{\mathrm{Ar}}(z)}{\partial z}\right|_1 \left.\frac{\partial G_{x,\tau}(z)}{\partial z^2}\right|_1 \\ &= \langle m(m-1)\rangle_{x,\tau} + \langle m(m-1)\rangle_{\mathrm{Ar}} + 2\langle m\rangle_{x,\tau}\langle m\rangle_{\mathrm{Ar}}.\end{aligned} \tag{5.3.6}$$

but from eq. 5.3.3

$$\begin{aligned}\langle m(m-1)\rangle_{\mathrm{Ar}} &= \left.\frac{\partial^2}{\partial z^2}\sum_{m=0}^{\infty}(zr+1-r)^m P_{\mathrm{Br}}(n)\right|_1 \\ &= \left.\sum_{m=0}^{\infty} m(m-1)(zr+1-r)^{m-2}r^2 P_{\mathrm{Br}}(n)\right|_1 = r^2\langle n(n-1)\rangle_{\mathrm{Br}}.\end{aligned} \tag{5.3.7}$$

For any random variable x, we can write $\langle x(x-1)\rangle = \sigma_x^2 - \langle x\rangle + \langle x\rangle^2$. Using this on eqs. 5.3.6 and 5.3.7 we get

See if it would be useful to give more detail in the next two steps

$$\begin{aligned}\sigma^2_{\mathrm{Br}} - \langle m\rangle_{\mathrm{Br}} + \langle m\rangle^2_{\mathrm{Br}} = &\left(\sigma^2_{x,\tau} - \langle m\rangle_{x,\tau} + \langle m\rangle^2_{x,\tau}\right) \\ &+ 2\left(r\langle m\rangle_{\mathrm{Br}}\right)\left[(1-r)\langle m\rangle_{\mathrm{Br}}\right] + r^2\left(\sigma^2_{\mathrm{Br}} - \langle m\rangle_{\mathrm{Br}} + \langle m\rangle^2_{\mathrm{Br}}\right)\end{aligned}$$

After some algebra we obtain

$$\sigma^2_{\mathrm{Br}} = \frac{1}{1-r^2}\sigma^2_{x,\tau} + \frac{r}{1+r}\langle m\rangle_{\mathrm{Br}} \tag{5.3.8}$$

Dividing by $\langle m\rangle_{\rm Br}^2$ and using eq. 5.3.5 we get

$$\begin{aligned}\eta_{\rm Br}^2 &= \frac{1}{1-r^2}\sigma_{x,\tau}^2\frac{(1-r)^2}{\langle m\rangle_{x,\tau}^2} + \frac{r}{1+r}\langle m\rangle_{\rm Br} \\ &= \frac{1-r}{1+r}\eta_{x,\tau}^2 + \frac{r}{1+r}\frac{1}{\langle n\rangle_{\rm Br}}\end{aligned} \tag{5.3.9}$$

Also, from eqs. 5.3.7 and 5.3.8 we get

$$\begin{aligned}\sigma_{\rm Ar}^2 - \langle m\rangle_{\rm Ar} + \langle m\rangle_{\rm Ar}^2 &= r^2\left(\sigma_{\rm Br}^2 - \langle m\rangle_{\rm Br} + \langle m\rangle_{\rm Br}^2\right) \\ &= r^2\left(\frac{1}{1-r^2}\sigma_{x,\tau}^2 + \frac{r}{1+r}\langle m\rangle_{\rm Br}\right) - r^2\langle m\rangle_{\rm Br} + r^2\langle m\rangle_{\rm Br}^2\end{aligned} \tag{5.3.10}$$

Using eq. 5.3.5 and after a little algebra we get

Do this algebra?

$$\sigma_{\rm Ar}^2 = \frac{r^2}{1-r^2}\sigma_{x,\tau}^2 + \frac{1}{1+r}\langle m\rangle_{\rm Ar}, \tag{5.3.11}$$

hence, dividing by $\langle m\rangle_{\rm Ar}^2$ and using eq. 5.3.5 we get

$$\begin{aligned}\eta_{\rm Ar}^2 &= \frac{r^2}{1-r^2}\sigma_{x,\tau}^2\left(\frac{1-r}{r\langle m\rangle_{x,\tau}}\right)^2 + \frac{1}{1+r}\frac{1}{\langle m\rangle_{\rm Ar}} \\ &= \frac{1-r}{1+r}\eta_{x,\tau}^2 + \frac{1}{1+r}\frac{1}{\langle m\rangle_{\rm Ar}}.\end{aligned} \tag{5.3.12}$$

Understand the intermediate steps here and complete

$$\boxed{\eta^2 = \frac{\langle b\rangle\left(\eta_t^2+\eta_b^2\right)+1}{2\langle n\rangle}} \tag{5.3.13}$$

Complete and analyze here also

5.4 Senescence of mRNA

Suppose that mRNAs are created at a constant rate λ_1 and that they senesce through a sequence of N steps labeled as $X_1, X_2, \ldots, X_N$. Thus, the states and their transitions are

$$\text{Transcription} \to X_1 \to X_2 \to \cdots \to X_N \to \text{Degradation} \tag{5.4.1}$$

The number of mRNA molecules in each step is labeled as x_i for $1 \leq i \leq N$. Also, assume that the rates of transcription β_1 per mRNA between the states are the same. The possible transtions in the state space are thus

$$\begin{aligned} x_1 &\xrightarrow{\lambda_1} x_1 + 1 \\ \{x_i, x_{i+1}\} &\xrightarrow{\beta_1 x_i} \{x_i - 1, x_{i+1} + 1\}, \quad \text{for} \quad 1 \leq i \leq N-1 \\ x_N &\xrightarrow{\beta_1 x_N} x_N - 1 \end{aligned} \tag{5.4.2}$$

Where the first line denotes transcription, the second denotes transitions between states and the third stands for degradation. Now, denote as x_{N+1} the number of proteins in a cell and suppose that independently of the current state of the mRNA molecules, they are translated with the same rate λ_2 per mRNA. Also, denoting the degradation rate per protein as β_2 the possible transitions for the proteins are

$$\begin{aligned} x_{N+1} &\xrightarrow{\lambda_2 \sum_{i=1}^{N} x_i} x_{N+1} + 1, \\ x_{N+1} &\xrightarrow{\beta_2 x_{N+1}} x_{N+1} - 1. \end{aligned} \tag{5.4.3}$$

Then, the average dynamics are

$$\begin{aligned}
\dot{\langle x_1 \rangle} &= \lambda - 1 - \beta_1 \langle x_1 \rangle, \\
\dot{\langle x_{i+1} \rangle} &= \beta_1 \left(\langle x_i \rangle - \langle x_{i+1} \rangle \right) \quad \text{for} \quad 1 \leq i \leq N, \\
\dot{\langle x_{N+1} \rangle} &= \lambda_2 \sum_{i=1}^{N} \langle x_i \rangle - \beta_2 \langle x_{N+1} \rangle.
\end{aligned} \tag{5.4.4}$$

At steady state we get

$$\begin{aligned}
\langle x_1 \rangle &= \frac{\lambda_1}{\beta_1}, \\
\langle x_{i+1} \rangle &= \langle x_i \rangle \quad \text{for} \quad 1 \leq i \leq N-1, \\
\langle x_{N+1} \rangle &= \frac{\lambda_2}{\beta_2} \sum_{i=1}^{N} \langle x_i \rangle.
\end{aligned} \tag{5.4.5}$$

Therefore,

$$\langle x_i \rangle = \lambda_1 \tau_1 \quad \text{for} \quad 1 \leq i \leq N, \tag{5.4.6}$$

where $\tau_1 \coloneqq {}^{1}/_{\beta_1}$.

Denoting the total mRNA as m, then $m = \sum_{i=1}^{N} x_i$ and taking the average

Complete this

Chapter 6

Effects of cell division

When the cell divides, all the components (organelles, proteins, genetic material, etc.) must be distributed among the daughter cells. Nevertheless, the assymetries on partition are an important source of noise even for components present at high numbers. In this chapter we will explore some general mechanisms of partition of molecules during cell division and how they can affect noise statistics.

This chapter follows the work done by D. Huh and J. Paulsson in [10].

6.1 Characterizing the noise arising from cell division

Let $x = l + r$ be the number of copies of some component (e.g. a certain protein) for a dividing cell, with l and r being the number of copies each daughter cell recieves. Also, let v be the number of molecules of some component that affects the partition such as vacuoles or spindles. On average, we expect the molecules to distribute symmetrically. Therefore

$$\langle l \rangle = \langle r \rangle = \frac{\langle x \rangle}{2}. \tag{6.1.1}$$

We will find the noise for l [1]. Using the law of total variance (eq. (1.4.6)), the variance of l is given by

$$\sigma^2(l) = \sigma^2\left(\langle l|x,v\rangle\right) + \left\langle \sigma^2(l|x,v)\right\rangle.$$

From eq. (6.1.1), $\langle l|x,v\rangle = \frac{x}{2}$, using this and dividing by $\langle l\rangle^2 = \frac{\langle x\rangle^2}{4}$ we get

$$\begin{aligned}\eta^2(l) &= 4\frac{\sigma^2\left(x/2\right)}{\langle x\rangle^2} + \frac{\langle\sigma^2(L|x,v)\rangle}{\langle L\rangle^2} \\ &= \eta^2(x) + Q_x^2,\end{aligned} \tag{6.1.2}$$

where we used eq. (1.4.3). The term Q_x, defined as

$$Q_x^2 \coloneqq \frac{\langle\sigma^2(L|x,v)\rangle}{\langle L\rangle^2}, \tag{6.1.3}$$

is the noise arising from cell division. Eq. (6.1.2) states that the noise after cell division is the sum of the noise before division and the noise arising at the division process (all squared).

The term Q_x can be interpreted in another way. From the definition of variance and eq. (6.1.1)

$$Q_x^2 = \frac{1}{\langle l\rangle^2}\left\langle \left(l - \langle l\rangle\right)^2 |x,v\right\rangle = \frac{4}{\langle x\rangle^2}\left\langle \left(l - \frac{x}{2}\right)^2 \middle| x,v\right\rangle$$

but $l - \frac{x}{2} = \frac{1}{2}(2l - (l-r)) = \frac{l-r}{2}$, then

$$Q_x^2 = \frac{4}{\langle x\rangle^2}\left\langle \left(\frac{l-r}{2}\right)^2 \middle| x,v\right\rangle = \frac{1}{\langle x\rangle^2}\left\langle (l-r)^2\right\rangle.$$

Check notation, they seem to skip to suddenly switch from conditional expectations to unconditional ones. Is it abuse of notation or is correct?

[1] It does not matter if we choose r or l since there is not preference for one of the daughter cells

Therefore, the term Q_x^2 can be interpreted as the average square deviation between the quantities of the molecules of each daughter cell. For a perfect division $l = r$, making $Q_x = 0$. On the contrary, the most noisy case is when one daughter recieves x moleucles and the other 0 molecule.

TODO: Analysis of the general framework and mock processes

6.2 Independent segregation

In the case of independent segregation, each molecule has an equal probability per unit time to switch from a cell half to another. Assuming there are l and $x - l$ molecules in each half, a process that can describe this statistic is given by

$$\begin{aligned} l &\xrightarrow{x-l} l+1 \\ l &\xrightarrow{l} l-1 \end{aligned} \tag{6.2.1}$$

Using the FDT, the jacobian and diffusion 1×1 matrices are given by

$$\begin{aligned} \mathbf{A} &= \partial_l \left((x-l) - l \right) = -2, \\ \mathbf{B} &= (x-l) + l = x. \end{aligned} \tag{6.2.2}$$

Solving for the covariance matrix, which in this case is the variance, we obtain in steady state

$$\sigma^2(l|x) = \frac{x}{4}, \tag{6.2.3}$$

and averaging and using eq. (6.1.3) we get

$$Q_x^2 = \frac{4}{\langle x \rangle^2} \left\langle \sigma^2(l|x) \right\rangle = \frac{4}{\langle x \rangle^2} \frac{\langle x \rangle}{4} = \frac{1}{\langle x \rangle}. \tag{6.2.4}$$

In the following sections, we will find the noise for some common division mechanisms and compare it to the case of independent segregation. The mechanisms that increase the noise with respect to it are included in what we call "disordered segregation", and we call those that decrease it "ordered segregation".

6.3 Disordered segregation

6.3.1 General case

First we will consider a general case in which the rate with which each molecules goes from a cell half to the other is proportional to the available space generated by the upstream component. For a fixed number v of molecules of the upstream component, there are n and $v-n$ available spaces in each daughter cell independently of x. Therefore, the process can be written as

$$\begin{aligned} l &\xrightarrow{n(x-l)} l+1 \\ l &\xrightarrow{(v-n)l} l-1 \end{aligned} \tag{6.3.1}$$

From the law of total variance we have

$$\sigma^2(l|x,v) = \left\langle \sigma^2(l|x,v,n) \right\rangle_{(n|v)} + \sigma^2\left(\langle l|x,v,n\rangle \right)_{(n|v)} \tag{6.3.2}$$

Where the subscript $(n|v)$ denotes that averages and variances are evaluated over the conditional PDF of n given x. Notice that taking averages over $(n|v)$ and over $(n|x,v)$ is the same in this case by the assumption that n is independent of x. Also, by symmetry we can assume that $\langle n|v\rangle = \frac{v}{2}$.

Therefore, by finding the first and second moments for $(l|x,v,n)$, we can use eq. (6.3.2) to find the variance for l given x and v. We will use the method of the moment generating function on $P(l|x,v,n)$. The master equation for this PDF is given by

$$\partial_t P(l|x,v,n) = n(x-(l-1))P(l-1|x,v,n) - (v-n)lP(l|x,v,n). \tag{6.3.3}$$

Writing the moment generating function as defined on eq. (1.6.1) we get

$$G(z) \coloneqq \sum_{l=0}^{x} z^l P(l|x,v,n). \tag{6.3.4}$$

The master eq. in terms of G is thus given by [2]

$$\partial_t G(z) = nxzG(z) - (v-n+nz)z\partial_z G(z)$$

At steady state we have

$$\partial_z G(z) = \frac{nxz}{(v-z+nz)z} G(z) \tag{6.3.5}$$

Solving with the boundary condition $G(1) = 1$, which follows from the normalization of the PDF, we find

$$G(z) = \left(1 + \frac{n}{v}(s-1)\right)^x = \sum_{l=0}^{x} \binom{x}{l} \left(\frac{n}{v}\right)^l \left(1-\frac{n}{v}\right)^{x-l} z^l. \tag{6.3.6}$$

Comparing with eq. (6.3.4) we get

$$P(l|x,v,n) = \binom{x}{l} \left(\frac{n}{v}\right)^l \left(1-\frac{n}{v}\right)^{x-l},$$

which is a binomial distribution. Its average and variance are given by [3]

$$\langle l|x,v,n\rangle = \frac{n}{v}x, \qquad \sigma^2(L|x,v,n) = \frac{n}{v}\left(1-\frac{n}{v}\right)x.$$

[2] A detailed derivation of this step is done on section 2.1 for a more complex master equation.

[3] In this case it was easy to find the distribution in terms of G, but we could have found the moments by differentiating and using the properties of G as well.

Taking the average of the conditional variance we get

$$\left\langle \sigma^2(l|x,v,n)\right\rangle_{(n|v)} = \left\langle \frac{n}{v}\left(1-\frac{n}{v}\right)x\middle|\, x,v,n\right\rangle_{(n|v)} = \left\langle \left(\frac{n}{v}-\frac{n^2}{v^2}\right)x\middle|\, x,v,n\right\rangle_{(n|v)}$$
$$= \left(\frac{\langle n|v\rangle}{v} - \frac{\sigma^2(n|v)+\langle n|v\rangle^2}{v^2}\right)x$$

Where we replaced $\langle n^2|v\rangle = \sigma^2(n|v)+\langle n|v\rangle^2$. Now since $\langle n|v\rangle = v/2$

$$\left\langle \sigma^2(l|x,v,n)\right\rangle_{(n|v)} = \left(\frac{1}{2} - \frac{\sigma^2(n|v)}{v^2} + \frac{1}{4}\right)x = \frac{x}{4}\left(1-Q_v^2\right), \tag{6.3.7}$$

where $Q_v^2 \coloneqq {}^{4\sigma^2(n|v)}/_{v^2}$. On the other hand, the variance of the conditional mean is given by

$$\sigma^2\left(\langle l|x,v,n\rangle\right)_{(n|v)} = \sigma^2\left(\frac{n}{v}x\right)_{(n|v)} = \frac{x^2}{v^2}\sigma^2(n|v) = \frac{x^2}{4}Q_v^2. \tag{6.3.8}$$

Replacing eqs. (6.3.7) and (6.3.8) on eq. (6.3.2) we get

$$\sigma^2(l|x,v) = \frac{x}{4}(1-Q_v^2) + \frac{x^2}{4}Q_v^2,$$

averaging and multiplying by ${}^4/_{\langle x\rangle^2}$

$$Q_x^2 = \frac{4}{\langle x\rangle^2}\left\langle \sigma^2(l|x,v)\right\rangle = \frac{4}{\langle x\rangle^2}\frac{1}{4}\left\langle x(1-Q_v^2)+x^2Q_v^2\right\rangle = \frac{1}{\langle x\rangle} - \frac{\langle Q_v^2 x\rangle}{\langle x\rangle^2} + \frac{\langle Q_v^2 x^2\rangle}{\langle x\rangle^2}. \tag{6.3.9}$$

We will use this equation to calculate the partitioning error Q_x^2 at different scenarios.

6.3.2 Random size and random accessible volume

Consider an available volume for the molecules that varies randomly. Let n be the fraction of available volume in one of the daughter cells and assume that each molecule is equally likely to occupy any volume unit. Hence, the probability per unit time of each

molecule leaving its cell half is proportional to the available volume in the other cell half. Therefore the process is

$$\begin{aligned} l &\xrightarrow{n(x-l)} l+1 \\ l &\xrightarrow{(1-n)l} l-1 \end{aligned} \tag{6.3.10}$$

This is the general case with $v = 1$. Assuming the volume variance is independent of x, eq. (6.3.9) becomes

$$Q_x^2 = \frac{1}{\langle x\rangle} - \frac{Q_v^2}{\langle x\rangle^2}\left(\langle x\rangle - \langle x^2\rangle\right) = \frac{1}{\langle x\rangle}\left(1 - \langle Q_v^2\rangle\right) + \langle Q_v^2\rangle\frac{\langle x^2\rangle}{\langle x\rangle^2}, \tag{6.3.11}$$

but $\frac{\langle x^2\rangle}{\langle x\rangle^2} = \frac{(\sigma^2(x)+\langle x\rangle^2)}{\langle x\rangle^2} = \eta_x^2 + 1$. Also, recalling the definition $Q_v^2 \coloneqq \frac{4\sigma^2(n|v)}{v^2}$. In this case $v = 1$ and it is fixed. Denoting it as Q_{vol}^2 we have $= Q_{\text{vol}}^2 = Q_v^2 = \langle Q_v^2\rangle = \frac{\sigma^2(n)}{\langle n\rangle^2}$, therefore

$$\boxed{Q_x^2 = \frac{1 - Q_{\text{vol}}^2}{\langle x\rangle} + Q_{\text{vol}}^2(\eta_x^2 + 1)}. \tag{6.3.12}$$

6.3.3 Clustered segregation

The clustering of molecules into vesicles could increase randomness in cell division. Let x and v be the total number of molecules and vesicles in a cell before division, respectively, and let x_i be the number of molecules in vesicle i, then $\sum_{i=0}^{v} x_i = x$. There are two processes that produce randomness: the migration of the molecules between vesicles and the partition of the vesicles into each daughter cells.

In the first part, a vesicle loses a molecule with a probability proportional to its number of molecules.

$$(x_1, \ldots, x_i, \ldots, x_j, \ldots x_v) \xrightarrow{x_i} (x_1, \ldots, x_i - 1, \ldots, x_j + 1, \ldots, x_v), \quad \text{for all } j \neq i. \tag{6.3.13}$$

In the second part, let n be the number of vesicles in one of the daughters, then similarly to the previous sections.

$$\begin{aligned} n &\xrightarrow{v-n} n+1 \\ n &\xrightarrow{n} n-1 \end{aligned} \tag{6.3.14}$$

If we assume both processes are independent, they could be done in any order to calculate the analytical expressions. i.e. it is the same to first distribute the molecules in each vesicle and then distribute the vesicles into each cell than to first distribute the empty vesicles between cells and then distribute the molecules. We will follow the second approach.

Let $x_1, \ldots, x_n$ be the number of molecules in each of the vesicles in one of the daughter cells and $x_{n+1}, \ldots, x_v$ be the number of molecules in the vesicles of the other daughter cell. As usual, let l be the number of molecules in one of the cells, then $l = \sum_{i=1}^{n} x_i$, and $r = x - l = \sum_{i=n+1}^{v} x_i$. Therefore, among all the possible transitions of eq. (6.3.13), the transitions coming from x_i and entering into x_j, for $i, j = 1, \ldots, n$, or $i, j = n+1, \ldots, v$, both with $i \neq j$ does not change the number of molecules. The net effect on the number of molecules in one of the daughter cells is

$$\begin{aligned} l &\xrightarrow{n(x_{n+1}+\cdots+x_v)} l+1 \\ l &\xrightarrow{(v-n)(x_1+\cdots+x_n)} l-1 \end{aligned} \tag{6.3.15}$$

TODO: Analisis

This can be simplified to

$$\begin{aligned} l &\xrightarrow{n(x-l)} l+1 \\ l &\xrightarrow{(v-n)l} l-1, \end{aligned} \tag{6.3.16}$$

which is the same as eq. (6.3.1). Hence, from eq. (6.3.9)

$$Q_x^2 = \frac{1}{\langle x \rangle} - \frac{\langle Q_v^2 x \rangle}{\langle x \rangle^2} + \frac{\langle Q_v^2 x^2 \rangle}{\langle x \rangle^2}. \tag{6.3.17}$$

Also, a correspondence can be established between eq. (6.2.1) and eq. (6.3.14) by switching $l \leftrightarrow n$ and $x \leftrightarrow v$. Using this correspondence on eq. (6.2.3) we get

$$\sigma^2(n|v) = \frac{v}{4}. \tag{6.3.18}$$

Recalling that $Q_v^2 \coloneqq \frac{4\sigma^2(n|v)}{v^2}$ we have in this case

$$Q_v^2 = \frac{4}{v^2}\frac{v}{4} = \frac{1}{v}, \tag{6.3.19}$$

and replacing on eq. (6.3.17) results in

$$\boxed{Q_x^2 = \frac{1}{\langle x \rangle} + \frac{1}{\langle x \rangle^2}\left(\left\langle \frac{x^2}{v} \right\rangle - \left\langle \frac{x}{v} \right\rangle\right)} \tag{6.3.20}$$

TODO: Explain the terms of the eq.

If x and v are independent we can write it as

$$\begin{aligned} Q_x^2 &= \frac{1}{\langle x \rangle} - \frac{1}{\langle x \rangle}\left\langle \frac{1}{v} \right\rangle + \frac{\langle x^2 \rangle}{\langle x \rangle^2}\left\langle \frac{1}{v} \right\rangle = \frac{1}{\langle x \rangle}\left(1 - \left\langle \frac{1}{v} \right\rangle\right) + \frac{\langle x^2 \rangle}{\langle x \rangle^2}\left\langle \frac{1}{v} \right\rangle \\ &\approx \frac{1}{\langle x \rangle} + \frac{\langle x^2 \rangle}{\langle x \rangle^2}\left\langle \frac{1}{v} \right\rangle = \frac{1}{\langle x \rangle} + \left(1 + \eta_x^2\right)\left\langle \frac{1}{v} \right\rangle, \end{aligned}$$

under the assumption that $\langle 1/v \rangle \ll 1$. This also allow us to do a Taylor expansion of $\langle 1/v \rangle$ around $\langle v \rangle$ obtaining

$$\left\langle \frac{1}{v} \right\rangle \approx \left\langle \frac{1}{\langle v \rangle} - \frac{v - \langle v \rangle}{\langle v \rangle^2} + \frac{(v - \langle v \rangle)^2}{\langle v \rangle^3} \right\rangle$$
$$= \frac{1}{\langle v \rangle} + \frac{\langle (v - \langle v \rangle)^2 \rangle}{\langle v \rangle^3} = \frac{1}{\langle v \rangle} \left(1 + \eta_v^2\right).$$

After replacing Q_x becomes

$$\boxed{Q_x^2 \approx \frac{1}{\langle x \rangle} + \frac{(1 + \eta_x^2)(1 + \eta_v^2)}{\langle v \rangle}}$$

When $x \ll 1$, $Q_x^2 \approx (1 + \eta_v^2)/\langle v \rangle$, meaning that the segregation of clusters is the more significant factor on partitioning erros.

Understand and explain the assumptions

Now assume that $\langle x|v \rangle = sv$ where s is a constant representing the average number of molecules per vesicle. The term in parentheses of eq. (6.3.3) becomes by definition of expected value

$$\left\langle \frac{x^2}{v} \right\rangle - \left\langle \frac{x}{v} \right\rangle = \sum_{x,v} \left(\frac{x^2}{v} - \frac{x}{v} \right) P(x, v) = \sum_v \frac{1}{v} \left[\sum_x \left(x^2 - x \right) P(x|v) \right] P(v)$$
$$= \sum_v \frac{1}{v} \left(\langle\, x^2 | v \rangle - \langle x|v \rangle \right) P(v) = \sum_v \frac{1}{v} \left(\sigma^2(x|v) + \langle x|v \rangle^2 - \langle x|v \rangle \right) P(v)$$
$$= \sum_v \frac{1}{v} \left(\frac{sv\sigma^2(x|v)}{\langle x|v \rangle} + s^2 v^2 - sv \right) P(v) = s \left\langle \frac{\sigma^2(x|v)}{\langle x|v \rangle} \right\rangle + s^2 \langle v \rangle - s \tag{6.3.21}$$

Defining $q \coloneqq \sigma^2(x|v)/\langle x|v \rangle$ and recalling that by the law of total expectation (eq. (1.4.5)) $\langle x \rangle = \langle \langle x|v \rangle \rangle = s\langle v \rangle$, we get

$$\left\langle \frac{x^2}{v} \right\rangle - \left\langle \frac{x}{v} \right\rangle = s\langle x \rangle + s(q - 1). \tag{6.3.22}$$

in eq. (6.3.3) we get

$$Q_x^2 = \frac{1}{\langle x \rangle} + \frac{1}{\langle x \rangle^2} \left(s\langle x \rangle + s(q-1) \right) \tag{6.3.23}$$

And if $(1-q)$ is very small (or 0 as in the case of a Poissonian), we can approximate it as

$$Q_x^2 \approx \frac{1}{\langle x \rangle} + \frac{s}{\langle x \rangle} \tag{6.3.24}$$

6.3.4 Upper limit of the partitioning error

Check the order of this subsection

The following contents are not revised

There is an upper bound for the partitioning error corresponding to the case when one daughter cells keeps all the molecules. There is an equal probability of each daughter to keep all of them, hence

$$\sigma^2(L|x) = \left\langle \left(l - \frac{x}{2} \right)^2 \right\rangle = \frac{1}{2} \left(0 - \frac{x}{2} \right)^2 + \frac{1}{2} \left(x - \frac{x}{2} \right)^2 = \frac{x^2}{4} \tag{6.3.25}$$

Therefore

$$Q_x^2 = \frac{4}{\langle x \rangle^2} \left\langle \sigma^2(l|x) \right\rangle = \frac{4}{\langle x \rangle^2} \langle x^2 \rangle 4 = \eta_x^2 + 1. \tag{6.3.26}$$

It only depends on the prior heterogeneity of the mother cells.

6.4 Ordered segregation

6.4.1 Self-volume exclusion

By analogy to eq. FILL making the correspondence FILL. If a molecule occupy a fixed volume fraction k of the total cell volume, we have

UNDERSTAND

$$\sigma^2(l|x) = \frac{1}{4}x(1-kx), \tag{6.4.1}$$

so that

$$Q_x^2 = \frac{4}{\langle x\rangle^2}\left\langle \sigma^2(l|x)\right\rangle = \frac{4}{\langle x\rangle^2}\frac{1}{4}\left(\langle x\rangle - k\langle x^2\rangle\right) = \frac{1}{\langle x\rangle} - k\frac{\langle x^2\rangle}{\langle x\rangle^2} = \frac{1}{\langle x\rangle} - k(\eta_x^2+1). \tag{6.4.2}$$

COMMENTS. It can be noticed that the reduction with respect to independent segregation gets bigger when the volume fraction occupied by eac molecule is bigger. This makes sense because it makes the exclusion bigger when there are more molecules, having the net effect of reducing the uneveness.

6.4.2 Binding to spindle sites

Suppose each dividing cells has a random nmber of binding sites v which are also distributed randomly between both daughter cells. Letting x be the (random) total number of molecules of some type which are going to bind the sites before division. We will separate cells in which $v < x$ and $v \geq x$. Assume also that the binding is such that all possible molecules of x are bound, that is, at equilibrium if $v < x$ all binding sites are occupied, and if $v > x$, all molecules are bound to sites.

Let n be the number of binding sites on a cell half, and suppose that it increases with a rate dependent on the number of binding sites in the other cell half, then

$$\begin{aligned} n &\xrightarrow{f(v-n)} n+1 \\ n &\xrightarrow{f(n)} n-1 \end{aligned} \tag{6.4.3}$$

where f is some function. Also, the rate at which a molecule leaves a cell half is proportional to the number of molecules in its cell half and the number of free sites in

the other cell half, obtaining

$$\begin{aligned} l &\xrightarrow{\lambda(n-l)(x-l)} l+1 \\ l &\xrightarrow{\lambda(v-n-x+l)l} l-1 \end{aligned} \tag{6.4.4}$$

UNDERSTAND AND EXPLAIN. SOLVE. Solving the FDR we get

$$\sigma^2(l|x,v) = \frac{1}{4}\left(x - \frac{x^2}{v} + Q_v^2 x^2\right), \quad \text{for } v \geq x, \tag{6.4.5}$$

with $Q_v^2 \coloneqq 4\sigma^2(n|v)/v^2$ as before. In the case $v < x$, the v copies of the molecule that are bound segregate along with n, and for the remaining copies suppose they segregate independently. The result is (COMPARE?)

$$\sigma^2(l|x,v) = \frac{1}{4}(x-v) + \sigma^2(n|v), \quad \text{for } v < x. \tag{6.4.6}$$

OJO, ERROR AQUI (CREO QUE YA CORREGIDO) Putting together both cases we get by definition of expectations

$$\begin{aligned} Q_x^2 &= \frac{4}{\langle x\rangle^2}\left\langle \sigma^2(l|x)\right\rangle = \frac{1}{\langle x\rangle^2}\sum_{x,v}\sigma^2(l|x)P(x,v) \\ &= \frac{1}{\langle x\rangle^2}\left[\sum_{v\geq x}\left(x - \frac{x^2}{v} + Q_v^2x^2\right)P(x,v) + \sum_{v<x}\left((x-v) + 4\sigma^2(n|v)\right)P(x,v)\right] \end{aligned}$$

,

notice that there is an x in both sums than can be taken out as a $\langle x\rangle$, also, by replacing $4\sigma^2(n|v) = v^2Q_v^2$ and separating the sums by first summing x and then over all vs we obtain

$$
\begin{aligned}
Q_x^2 &= \frac{1}{\langle x\rangle} - \frac{1}{\langle x\rangle^2}\sum_{v=0}^{\infty}\left[\sum_{x=0}^{v}\left(\frac{1}{v}-Q_v^2\right)x^2P(x,v) + \sum_{x=v+1}^{\infty}\left(v - v^2Q_v^2\right)P(x,v)\right] \\
&= \frac{1}{\langle x\rangle} - \frac{1}{\langle x\rangle^2}\sum_{v=0}^{\infty}\left[\left(\frac{1}{v}-Q_v^2\right)\sum_{x=0}^{v}x^2P(x,v) + \left(v - v^2Q_v^2\right)\sum_{x=v+1}^{\infty}P(x,v)\right].
\end{aligned}
\tag{6.4.7}
$$

To make interpretations easier, consider a special case in which v is fixed, each daughter cell recieves exactly $v/2$ binding sites, $\langle x\rangle = v$, and $P(x)$ is symmetric. With these assumptions, the previous eq. can be reduced

$$
Q_x^2 = \frac{1}{\langle x\rangle} - \frac{1}{\langle x\rangle^2}\left[\frac{1}{v}\sum_{x=0}^{v}x^2P(x) + v\sum_{x=v+1}^{2v}P(x)\right]
$$

where we made $Q_v^2 = 0$ because there is no uncertainty on n since each cell recieves exactly $v/2$ sites. Also, for the sum over v only survives the term corresponding to the fixed number v of binding sites. Writing $x^2 = (x-\langle x\rangle)^2 + 2x\langle x\rangle - \langle x\rangle^2$ on the first sum we get

$$
Q_x^2 = \frac{1}{\langle x\rangle} - \frac{1}{\langle x\rangle^2}\left[\frac{\sigma^2(x)}{v}\sum_{x=0}^{v}P(x) + \frac{2\langle x\rangle}{v}\sum_{x=0}^{v}xP(x) - \frac{\langle x\rangle^2}{v}\sum_{x=0}^{v}P(x) + v\sum_{x=v+1}^{2v}P(x)\right]
$$

evaluating $\langle x\rangle = v$ we get

$$
Q_x^2 = \frac{1}{\langle x\rangle} - \frac{1}{\langle x\rangle^2}\left[\frac{\sigma^2(x)}{v}\sum_{x=0}^{v}P(x) + 2\sum_{x=0}^{v}xP(x) - v\sum_{x=0}^{v}P(x) + v\sum_{x=v+1}^{2v}P(x)\right],
$$

and since $P(x)$ is symmetric about $x = v$ we have $\sum_{x=0}^{v}P(x) = \sum_{x=v+1}^{2v}P(x) = 1/2$

$$
Q_x^2 = \frac{1}{\langle x\rangle} - \frac{1}{\langle x\rangle^2}\left[\frac{\sigma^2(x)}{2v} + 2\sum_{x=0}^{v}xP(x)\right]. \tag{6.4.8}
$$

The absolute deviation $\langle|x-\langle x\rangle|\rangle$ can be written using the symmetry of $P(x)$ as

$$\langle|x-\langle x\rangle|\rangle = \sum_{x=0}^{\infty} |x-\langle x\rangle|\, P(x) = 2\sum_{x=0}^{v} (\langle x\rangle - x)\, P(x) = \langle x\rangle - 2\sum_{x=0}^{v} xP(x). \quad (6.4.9)$$

Thus, solving for the sum and replacing we get

$$Q_x^2 = \frac{1}{\langle x\rangle} - \frac{1}{\langle x\rangle^2}\left[\frac{\sigma^2(x)}{2v} + \langle x\rangle - \langle|x-\langle x\rangle|\rangle\right] = \frac{\langle|x-\langle x\rangle|\rangle}{\langle x\rangle^2} - \frac{\eta_x^2}{2v}$$

And using $v = \langle x\rangle$

$$\boxed{Q_x^2 = \frac{1}{\langle x\rangle}\left(\frac{\langle|x-\langle x\rangle|\rangle}{\langle x\rangle} - \frac{\eta_x^2}{2}\right)} \quad (6.4.10)$$

ANALYSIS, η_x cannot exceed 1 by the symmetry of $P(x)$?

6.4.3 Pair formation mechanisms

A mechanism of ordered segregation consists in the pair formation of the molecules to be segregated and then spindles are formed which separates each molecule forming the pair into each cell half.

Assume that in each cell there are z pairs of molecules and m independent molecules i.e. $x = m+2z$. Suppose also that the paired molecules do not interact with the unpaired ones, then

$$\begin{aligned}\left\langle\sigma^2(l|x)\right\rangle &= \left\langle\sigma_{\mathrm{p}}^2(L|2z)\right\rangle + \left\langle\sigma_{\mathrm{u}}^2(l|m)\right\rangle = \sum_{x,z}\left[\sigma_{\mathrm{p}}^2(l|2z) + \sigma_{\mathrm{u}}^2(l|m)\right] P(2z,x) \\ &= \sum_{x,z}\left[\sigma_{\mathrm{p}}^2(l|2z) + \sigma_{\mathrm{u}}^2(l|x-2z)\right] P(2z|x)P(x)\end{aligned} \quad (6.4.11)$$

where the subscripts 'p' and 'u' stand for 'paired' and 'unpaired' and $P(2z,x)$ is the PMF of having z pairs and a total of x molecules before division. The variances can be

added in that way because for independent random variables, the variance of a sum is the sum of the variances (IS THIS THE REASON?).

Now We will proceed to find each one of the variances. The unpaired molecules segregate independently, therefore, by comparison with eq. (6.2.3) we get

$$\sigma_{\rm u}^2(l|x-2z) = \frac{x-2z}{4}. \tag{6.4.12}$$

For the paired molecules, assume that each pair is split to separate daughters with probability p and to the same daughter with probability $1-p$. MORE COMMENTS

$$\begin{aligned}
(M,l,r) &\xrightarrow{pM} (M-2,l+1,r+1)\\
(M,l,r) &\xrightarrow{(1-p)M/2} (M-2,l+2,r)\\
(M,l,r) &\xrightarrow{(1-p)M/2} (M-2,l,r+2)
\end{aligned} \tag{6.4.13}$$

where the first line represents a succesfull split and the other two unsuccesful ones. EXPLAIN ALL THE FDR, AND THE MATRICES, ETC. ————————————- we get

$$\sigma_{\rm p}^2 = (1-p)z. \tag{6.4.14}$$

Replacing eqs. (6.4.12) and (6.4.14) in eq. (6.4.11) we get

$$\begin{aligned}
\left\langle \sigma^2(l|x)\right\rangle &= \sum_x \left[\sum_z \left((1-p)z + \frac{x-2z}{4}\right) P(2z|x)\right] P(x) = \sum_x \left(\frac{1-p}{2}\langle 2z|x\rangle + \frac{x-\langle 2z|x\rangle}{4}\right) P(x)\\
&= \frac{1}{4}\sum_x \left(2(1-p)\langle 2z|x\rangle + x - \langle 2z|x\rangle\right) P(x) = \frac{1}{4}\sum_x \left(x - (2p-1)\langle 2z|x\rangle\right) P(x)\\
&= \frac{1}{4}\left(\langle x\rangle - (2p-1)\langle 2z\rangle\right).
\end{aligned} \tag{6.4.15}$$

Hence, (POR QUE EL SOBRE 2, SUSTITUCION?)

$$\boxed{Q_x^2 = \frac{1-(2p-1)k}{\langle x \rangle}} \tag{6.4.16}$$

where $k \coloneqq \langle 2z \rangle / \langle x \rangle$ is the average fraction of molecules that are in pairs. If $k = 0$ there is independent segregation and $Q_x = 1/\langle x \rangle$ on the previous equation. For the segregation into pairs to be 'ordered', p must be greater than $1/2$, in the opposite case, the paired molecules have a higher chance of not being split, increasing segregation error with respect to the independent case.

Chapter 7

Conclusions

Bibliography

1. Kaern, M., Elston, T. C., Blake, W. J. & Collins, J. J. Stochasticity in gene expression: from theories to phenotypes. *Nat. Rev. Genet.* **6,** 451–464 (2005).

2. Elowitz, M. B., Levine, A. J., Siggia, E. D. & Swain, P. S. Stochastic gene expression in a single cell. *Science* **297,** 183–186 (2002).

3. Pedraza, J. M. & van Oudenaarden, A. Noise Propagation in Gene Networks. *Science* **307,** 2005 (1965–1969).

4. Paulsson, J. Models of stochastic gene expression. *Phys. Life Rev.* **2,** 157–175 (2005).

5. Alon, U., Surette, M. G., Barkai, N. & Leibler, S. Robustness in bacterial chemotaxis. *Nature* **397,** 168–171 (1999).

6. Arkin, A., Ross, J. & McAdams, H. H. Stochastic Kinetic Analysis of Developmental Pathway Bifurcation in Phage λ-Infected Escherichia coli Cells. *Genetics* **149,** 1633–1648 (1998).

7. Thattai, M. & van Oudenaarden, A. Intrinsic noise in gene regulatory networks. *Proc. Natl. Acad. Sci. U.S.A.* **98,** 8614–8619 (2001).

8. Pedraza, J. M. & Paulsson, J. Effects of Molecular Memory and Bursting on Fluctuations in Gene Expression. *Science* **319,** 339–343 (2008).

9. Huh, D. & Paulsson, J. Non-genetic heterogeneity from stochastic partitioning at cell division. *Nat. Genet.* **43,** 95–100 (2011).

10. Huh, D. & Paulsson, J. Random partitioning of molecules at cell division. *Proc. Natl. Acad. Sci. U.S.A.* **108,** 15004–15009 (2011).

11. Alberts, B. *et al. Essential Cell Biology* (Garland Science, New York, 2013).

12. Alon, U. *An Introduction to Systems Biology: Design Principles of Biological Circuits* (Chapman and Hall/CRC, Boca Raton, 2006).

13. Bertsekas, D. P. & Tsitsiklis, J. N. *Introduction to Probability* (Athena Scientific, Belmont, 2008).

14. Van Kampen, N. G. *Stochastic Processes in Physics and Chemistry* (North-Holland, Amsterdam, 1992).

15. Gardiner, C. W. *Handbook of Stochastic Methods for Physics, Chemistry and the Natural Sciences* (Springer, Berlin, 2003).

16. Gillespie, D. T. Exact stochastic simulation of coupled chemical reactions. *J. Phys. Chem.* **81,** 2340–2361 (1977).

17. Paulsson, J. Summing up the noise in gene networks. *Nature* **427,** 415–418 (2004).

18. Pedraza, J. M. *Signal and noise propagation in genetic circuits* PhD thesis (Massachusetts Institute of Technology, 2006).