UNIVERSITY OF SOUTHAMPTON

Dynamics of Molecular Fluctuations in Gene Regulatory Networks

by

Vijayanarasimha Hindupur Pakka

A thesis submitted in partial fulfillment for the
degree of Doctor of Philosophy

in the

Faculty of Engineering, Science and Mathematics
School of Electronics and Computer Science

December 2009

to ANNA and AMMA

ABSTRACT

The components that are central to cellular processing are proteins, whose production is regulated by other proteins known as transcription factors. Proteins are products of genes that regulate the expression of one another, thereby forming large gene regulatory networks that perform specific cellular functions. The complex connectivity between genes of a network could result in various behaviours that are interesting. The assumption then is that tracking subnetwork behaviour helps understand the characteristics of the larger networks they are embedded in. For example, the structure of a subnetwork could say a lot about its biological role. Theoretical models of such systems and their deterministic dynamical properties have been the focus of study in the past. However, the dynamics of transcriptional control involves small numbers of molecules and result in significant fluctuations in protein and mRNA concentrations. Hence the recent shift in focus has been towards stochastic modelling approaches. Experimentally, the issues regarding average molecular numbers over a cell population draw our attention towards single-cell techniques where these fluctuations in the numbers are captured. On understanding the fluctuation properties of the smaller networks, one could study or design a combination of these networks leading to more complex regulatory networks.

The objective of this thesis is to characterize small subnetworks of genes, based on the properties of their internal fluctuations. The correlations between these intrinsic fluctuations then offer, via the fluctuation dissipation relation, the possibility of capturing the system's response to external perturbations, and hence the nature of the regulatory activity itself. Therefore we do a stochastic analysis and derive time-dependent noise correlation functions between molecular species of the networks, and using these functions we study simple networks by varying three of its factors. One is the type of regulatory activity that is present between two genes or proteins, whose correlations we are interested in. We show that the regulatory mechanism of activation, repression either by monomers or dimers, produces different correlations. We also study the dependence of the correlations on the values of the rate constants for the ingredient processes. We demonstrate the influence of various rate constants on the protein correlations. Finally, we analyze regulatory networks of different motifs such as cascades and feedforward loops and explore the extent to which fluctuation correlations report on the network structure. The distinct correlated fluctuations could then possibly be used as signatures for identifying the regulatory mechanism present between two genes of a network. To that end, in this thesis we present analytical and numerical results on features such as the magnitudes and time delays in dynamic correlations between proteins within smaller networks, and the dependence of these features on rate constants and regulatory and network mechanisms.

Contents

List of Figures

List of Tables

Acknowledgements

First of all I would like to express my sincere gratitude to Dr. Srinandan Dasmahapatra, Lecturer of Science and Engineering of Natural Systems Research Group, School of Electronics and Computer Science, Southampton University, who has been my supervisor for the last four years. I am indebted to him for his excellent guidance, patience and care during the entire course of my Ph.D study. He has been a source of immense motivation and for the last four years has constantly encouraged and inspired me towards higher pursuits. He stands tall and solid as the Rock of Gibraltor, constantly reminding me that it takes lots of patience and determination to achieve a desired objective. I would cherish all the very interesting discussions we had relating to gene regulation and stochastic processes. The knowledge and research techniques that I learnt from him are invaluable. My sincere and heartfelt thanks to him for guiding me smoothly through an interesting and challenging topic of research.

I also wish to express my deep appreciation to Dr. Adam Prügel-Bennett of the ISIS group, School of Electronics and Computer Science for being my secondary supervisor and giving me many valuable suggestions and constructive advice. The discussions I had with him proved to be very useful. I am indebted to him for his constant encouragement without which this work would not have been possible.

Thanks also to Dr. Klaus-Peter Zauner, and Prof. Vladimiro Sassone for taking academic interest in this study. I also wish to express my cordial appreciation to Dr. Jason Noble, Dr. Seth Bullock and Dr. Richard Watson for making me feel an important part of the research group. I would greatly miss my lab mates Rob Mills and Effirul Ramlan who joined the PhD course along with me and have been excellent companions for the last four years. I would also invariably miss Gareth Jones, Chris Lovell and Simon Powers who created a relaxed and jovial environment at the workplace. I wish them the very best for their PhD studies. Sincere thanks to post-docs Dr. Soichiro Tsuda, Dr. Nicholas Geard and Dr. Alexandra Penn with whom I had some interesting discussions. I wish to thank the entire research group and the department for providing me a conducive environment to work in. A note of thanks to the Academicals Cricket Club with whom I enjoyed three seasons of excellent cricket that kept my body and mind fresh.

Special thanks to Dr. Sujala Singh for befriending me and making me feel at home in Southampton. My best wishes to her and little one Aashish.

My greatest appreciation goes to my parents, Sri. Ananda Murthy Hindupur Pakka and Smt. Shakuntala Hindupur Pakka, without whom nothing would have been possible. They kept me totally away from any family responsibilities, which allowed me to concentrate on my studies. In them, I find a source of motivation that is infinite in volume. I am lucky to have closely witnessed them overcome many hurdles in life, through which I constantly draw inspirations. I may never be able to express how much I love and respect them and also may never be able to thank them enough for all their support throughout my academic journey.

I would like to express my heartiest thanks to my brother-in-law Jayanth and sister Anupama for their unending kindness and affection, and for looking after me like a son. They never let me feel that I was away from my parents and country. Also, sincere regards to my brother-in-law Surendra and sister Jyothi Roopa for their great love and affection and constant encouragement. I wish her the very best in her PhD studies and shall try to reciprocate the encouragement that she gave me all these years.

Neither can I forget the precious moments that I spent with little ones Krishnaprasad, Sharanya and Shreya, which pulled me up in times of depression. I would like to express special thanks to my love Pannagashree. She helped me concentrate on completing this dissertation and, who I know will support me in my future research and academic works. Thanks to Anand, Naresh, CD, Madan, HS, Balaji, Manojav, my friends for life, for having stood behind me in times of difficulty. Finally, I would like to remember and thank everyone who has directly or indirectly helped me in completing this work. I look forward to their support and encouragement in my future research endeavours.

- Vijay Pakka.

Nomenclature

$\mathbf{A}$	The Jacobian matrix
$(A)_{ij},\ A_{ij}$	$(ij)^{th}$ element of $\mathbf{A}$
$\mathbf{A}^T$	Transpose of $\mathbf{A}$
$\mathbf{V}$	Matrix whose columns are the right eigenvectors of $\mathbf{A}$
$\mathbf{V}_i$	i^{th} right-eigenvector of $\mathbf{A}$ or i^{th} column of $\mathbf{V}$
$\mathbf{U}$	Matrix whose columns are the left eigenvectors of $\mathbf{A}$
$\mathbf{\Lambda}$	Diagonal matrix whose elements are the eigenvalues of $\mathbf{A}$
λ_i	i^{th} eigenvalue of $\mathbf{A}$
$\mathbf{C}$	Stationary covariance matrix
$\mathbf{S}$	Stoichiometric matrix
$\mathbf{Y}$	Propagator matrix
X_i	Number of molecules of the i^{th} species
x_i	Concentration of the i^{th} species
$\mathbf{X}$	Vector whose elements are X_i
$\mathbf{x}$	Vector whose elements are x_i
N	Number of species/variables $(s_1, \cdots, s_N)$ in a biochemical reacting system
M	Number of elementary reactions $(r_1, \cdots, r_M)$ in a biochemical reacting system
$\mathbf{B}$	Diffusion matrix
$\mathbf{R}$	Vector whose elements $\mathbf{R}_j$ are the flux of each r_j^{th} reaction
M_x	mRNA of gene labelled X
P_x	mRNA of gene labelled X
G_y	Gene Y
$k^-_{m_x}$	Rate constant of the decay process of M_x
p_α	General symbol for the system parameters such as $k^-_{m_x}$
t	Time
τ	Additional symbol for time
Cov	Covariance between two variables/molecular species
Corr	Correlation between two variables/molecular species
P	Probability distribution function
$\frac{\partial}{\partial t}$	Partial derivative
Ω	Volume of the system

Chapter 1

Introduction

The existence and growth of cells, whether those of bacteria or humans, is in essence sustained by molecular interactions that follow complex specific pathways in order to robustly perform various cellular functions. A pathway represents a set of closely related biochemical reactions describing molecular interactions between components of a cell. Depending on the nature of the cellular processes, pathways are broadly classified as metabolic, signalling or genetic. Just as metabolic networks describe the metabolism of cellular components such as sugars, amino acids and lipids and signal transduction networks describe the transfer of information from extracellular signals to the genetic regulatory system inside the nucleus through enzymatic messengers, the genetic networks describe all the molecular interactions that are specific to processes that regulate the quantities of proteins that participate in metabolic and signalling pathways. A classic example of interconnectivity that exists between the three networks is the JAnus Kinase-Signal Transducer and Activator of Transcription (JAK-STAT) pathway that involves activation of JAK through the binding of a ligand to an associated transmembrane receptor. The JAK protein further induces the phosphorylation of STAT proteins that travel into the cell nucleus and activate the transcription of certain genes.

Proteins are the most significant components of such complex cellular machinery and the genome is a template or a codebook that directs the production of these proteins. The sequencing of the genome is therefore viewed as a milestone in the quest for unravelling the secrets of cellular functioning. Though genome sequencing is significant, it alone does not explain the cellular functioning of an organism. One needs to understand the complex interactions between genes, their products the proteins, and other genetic components in order to dissect the numerous cellular processes that depend on these molecular interactions. The basic requirement for this is a model of the regulatory process. A model that captures the essential dynamics of the system would consist of interactions between DNA, RNA and protein molecules. A pictorial depiction of a simple model of gene expression, including the polymerases and ribosomes that initiate

the transcription and translation processes, is given in figure 1.1. A brief mathematical description of the transcription function generating out of these molecular interactions is discussed in section 1.4.1. Apart from these core set of molecular interactions, some of the supplementary events that are typical of a regulatory system are interactions between various proteins, formation of multimolecular complexes, regulatory proteins undergoing modifications such as phosphorylations, etc.

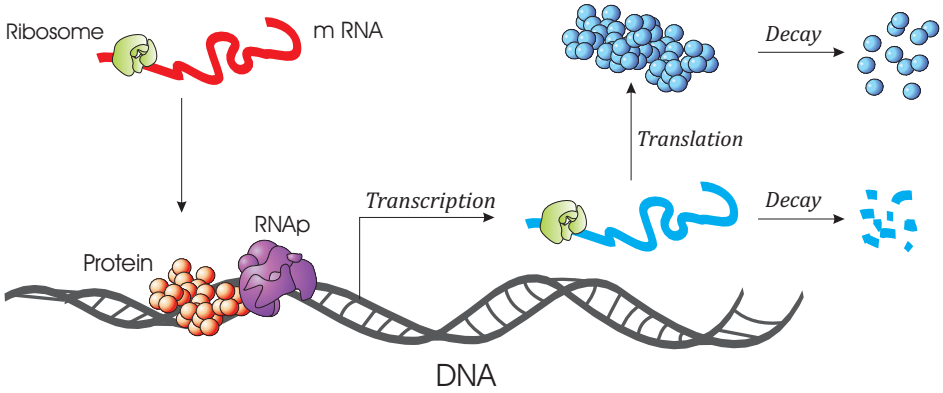

FIGURE 1.1: Transcription Factor protein (red) up-regulating the production of another protein (blue).

A Gene Regulatory Network (GRN) can then be viewed as a set of interacting components, which controls and regulates the expression of various genes that are responsible for a specific cellular function. Transcripts of a gene are used to produce particular protein complexes, which might then bind to various regulatory sequences of other genes including its own and control their rate of transcription, thereby acting as a Transcription Factor (TF). These genes in turn produce proteins, which may further act as TFs for other genes, thus forming a regulatory network. The regulatory action of a TF on another gene is loosely represented by an edge in the GRN.

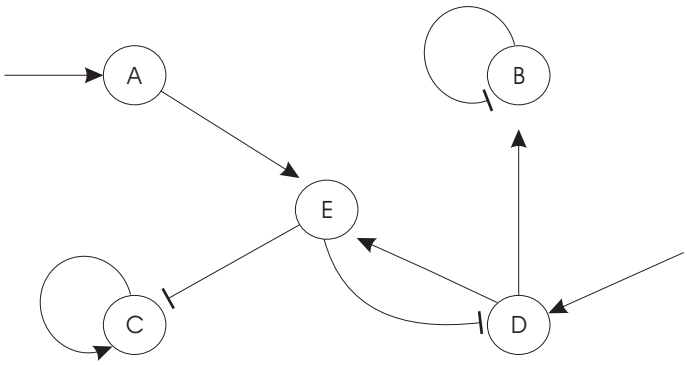

FIGURE 1.2: A sample gene regulatory network of five genes encoding the proteins *A*, *B*, *C*, *D*, *E*. Gene activation is denoted by an arrowhead and repression by a dashhead. TFs *A* and *D* respond to external stimuli.

1.1 Motivation

Retrieving the structure of a GRN is a challenging task due to its complexity and dynamic nature. Experimentalists have strived hard to accumulate the right type and amount of data required to learn more about these genetic processes and to retrieve the structure of GRNs. Recent years have seen great technological advances in data collection from cells. Microarray technologies that provide large scale genome-wide measurements of mRNAs have proved to be of immense value in inference of the connectivities and causal relationships between genes of a GRN. We shall call this the *top-down* approach. We shall elaborate a bit more on this in section 1.4.

Average concentration levels of molecular species taken from a population of cells could be quite different from those seen in individual cells Novick and Weiner (1957); Vilar et al. (2003). This is due to the random nature of the molecular interactions giving rise to fluctuations in the concentration levels of the species. Therefore due to averaging effects, the observed behaviour of GRNs in a population of cells could be different compared to that in individual cells. As an example consider the gene regulation function (GRF), which is the relation between the concentration of active TFs and the rate at which their downstream gene products are expressed. The shape of this function determines key features of cellular behaviour such as developmental cell-fate decisions or oscillations in the system. Rosenfeld et al. (2005) experimentally showed that the single-cell GRF cannot be represented by a single valued function, as it fluctuates dynamically in individual cells, thereby limiting the accuracy with which genetic circuits can transfer signals. Further, stochastic analysis reveals that the distribution of mRNA and protein levels over population of cells could vary greatly for changes in certain reaction parameters, while still maintaining constant levels of average expression. Hence, the average response might not necessarily estimate the true character of the gene regulation function and also the connections in a GRN. As a consequence of the above facts, the estimated correlations between any two genes by making use of the average expression levels over a population, might not be wholly true. Also, a stochastic description might in some cases lead to qualitatively different outcomes than that of population averages. Therefore the focus in recent years has been towards studying the cause and effects of these molecular fluctuations in GRNs. A stochastic description of the models is used to address this issue. A detailed discussion on the causes and consequences of the random nature of the genetic process is given in the next chapter.

Secondly, the GRN under investigation has to perform in perfect synchrony in all the cells so that its average behaviour when taken over the population is in close accordance with its behaviour in a single cell. Therefore the average expression levels from a population of cells might not necessarily estimate the true regulatory activity or the network structure of GRNs. Such issues need to taken into account while characterizing the regulatory link

between any two genes. In section 1.5 we shall briefly discuss the issue of synchronization and other issues related to population averages.

However, investigating any GRN that comprises of hundreds if not thousands of genes before knowing its components is doomed to failure. A GRN in reality is a complex system whose functioning depends on the concerted performance of numerous subsystems that in turn could have well defined characteristics. These subsystems could be simple regulatory networks consisting of two to three genes that have well defined properties. Since extensive knowledge about the *parts* would lead to a better understanding of the functional properties of the *whole* system, it is vital to have a clear understanding of these simple regulatory structures Sprinzak and Elowitz (2005). This forms the motivation for investigating molecular fluctuations in simple gene regulatory which is helpful in characterizing the regulatory activity between the genes of the network. This is a *bottom-up* approach that we are interested in Guido et al. (2006).

1.2 Objective of the Thesis

Due to the issues mentioned in the previous section:

> **The objective of this work is to design a framework, built upon the stochastic nature of the biological process, that estimates the presence of and further characterizes the regulatory activity between genes in small regulatory networks.**

Towards this objective, we require a stochastic description of the regulatory system which estimates the internal fluctuations in molecular numbers. Stationary as well as time-dependent statistical features extracted from such fluctuations would then act as signatures for characterizing the regulatory activity that gave rise to these fluctuations in the first place. Therefore different types of regulatory mechanisms such as activation/repression via monomers/dimers would be expected to have their respective unique signatures. Observing these signatures in simulated or real-time data would suggest the type of regulatory action present between two genes. Since the molecular fluctuations are captured in single-cell measurements, the ideal measurements for such an investigation would come from single cells and would be valuable and informative. Emerging technologies such as time-lapse spectroscopy provide us with such measurements. In this work, we propose a mechanism wherein the fluctuating expression levels of mRNAs and proteins at the single-cell level are sufficient to characterize the nature of regulatory action between any two genes. On the other hand, the analytical framework presented here would be valuable in studying the fluctuation properties of networks that exhibit unique behaviours. Therefore by studying the internal fluctuations in networks through variations in their structure and parameters one could gain insights into their biological

functioning. This is a step towards elucidating the network structure of bigger and more complex GRN, whose building blocks are the focus of investigation in this work.

Outline of the Thesis: In Chapter 2, we discuss the causes and effects of stochasticity in GRNs. Various models of gene regulation and their corresponding analytical formulations of noise are also discussed with examples. They form the basic aspects of our analysis of GRNs. In Chapter 3 we derive an analytical framework based on the statistical measures of a stochastic system. We derive the stationary as well as the time-dependent correlations between components of a GRN. We show that the dynamics of molecular fluctuations at the level of single cells acts as signatures in characterizing the regulatory mechanism and structure of simple regulatory networks. We also show the relation between these internal fluctuations and the response characteristics of the system. On the other hand, for well-known regulatory networks, whose response characteristics are unknown, we could utilize our framework to predict the same. In Chapter 4, we derive a framework for estimating the sensitivity of molecular fluctuations for changes to the reaction rate constants. In Chapter 5 we apply the time-correlation functions to simple two-gene networks having different regulatory mechanisms and discuss the results in detail. Finally, in Chapter 6, we investigate networks comprising of three genes that demonstrate interesting behaviour both in the deterministic response and in the internal fluctuations, for changes to their network structure and rate constants.

1.3 Single-Cell Measurements

Fluorescent reporters such as the green fluorescent protein (*gfp*[1]) Tsien (1998) and its variants the yellow and cyan fluorescent proteins have increased our ability to track protein levels in individual cells, which is a major shift from the microarray realm where bulk averages are the measured quantities. Technologies such as flow cytometry measure the relative fluorescence intensities of individual cells as they flow in a fluid stream thus enabling one to plot histograms of protein fluorescence levels Hooshangi et al. (2005); Pedraza and van Oudenaarden (2005). Ozbudak et al. (2002) used flow cytometry to observe variations in the protein distributions for changes in parameters values such as the transcription and translation rates by incorporating the *gfp* into the chromosome of *Bacillus subtilis*. Such analyses are however concerned with the stationary distribution in proteins and could tell little about the dynamics of the reactions involved. On the other hand, technologies such as time-lapse microscopy where fluorescently tagged proteins could be tracked over time in individual cells, facilitate our understanding of the dynamics of gene regulation Rosenfeld et al. (2005).

[1]The *gfp* gene is isolated from organisms such as the Pacific jellyfish, *Aequoria victoria*

As an example, time-lapse microscopy was used by Elowitz and Leibler (2000) who constructed a synthetic regulatory network consisting of three genes on a plasmid[2], where each of the genes repressed another in a cyclic order. To monitor the activity of one of the three genes TetR, they built another plasmid with the sequence of the *gfp* gene inserted into it. The *gfp* was placed under the regulatory action of *tetR* by constructing a promoter sequence $P_L tet01$ upstream to it. Consequently, when cells of *Escherichia coli* containing the two plasmids were grown to a stationary state, the activity of the TetR gene was observed *via* the fluorescence levels of expressed *gfp*. The fluorescence levels of each cell over time was plotted by manually tracking back in time the individual cell lineages in microcolonies of the culture. Similarly, promoter sequences, ribosome-binding sites, *gfp* sequence and transcriptional terminator sites from the plasmids could be integrated into the chromosome of a host organism. Elowitz et al. (2002) do the same by incorporating green and yellow fluorescent proteins, controlled by identical promoters, into the chromosome of *E.coli* and observe the stochastic effects of gene expression by noting the difference in fluorescence levels of the two proteins.

Time-lapse measurements of molecular species in single cells would be of immense help in estimating the presence of any regulatory activity between the corresponding genes. Due to rapid technological advancements, obtaining such fine measurements that are needed to track the causal dynamics as revealed by dynamic correlations, is fast becoming a reality. A promising step in this direction is from the work of Yu et al. (2006), where they track single molecules of yellow fluorescent protein (*yfp*) in living cells by assembling them into the inner membrane of *Escherichia coli* cells thereby slowing their diffusion and making their individual detection possible by fluorescence microscopy. Other single molecule techniques such as Cai et al. (2006) suggests that the analytical techniques presented in this thesis can be made relevant to experimental investigations.

Studies such as those by Vargas et al. (2005) and Raj et al. (2006) who use the FISH technique to track single-molecules of mRNA in individual cells devote much attention to the stationary distributions. Though such studies give immense amount of information regarding the processes of production of mRNAs, some unanswered questions could be better tackled when time-dependent behaviour of these species are studied. In this regard, variations in protein levels in human cells was observed by Sigal et al. (2006b), who tracked the fluorescently tagged proteins and concluded that the fluctuations in the protein levels varied slowly in time, in comparison to the cell-cycle times. It was also observed that genes of the same pathway showed correlations between them. The analytical framework presented in this work would then be a useful tool in predicting the type of regulatory activity or even the reaction structure between two genes. The values for the parameters could also be estimated within the framework of sensitivity analysis from such single-cell data. However, due to the underlying physical properties

[2]extrachromosomal self-replicating DNA molecule that contains a few genes

of the protein molecules being tampered with by tagging them with heavier fluorescent particles, learning the true parameter values from such data would be hard. Due to the high sensitivity of the correlations to certain reaction rate constants, we could also hope to extract more information about the underlying reactions or processes.

1.4 Background

1.4.1 Modelling gene expression

The pivotal work of Jacob and Monod (1961) gave us the first real glimpse of transcriptional regulation. They studied various mutations in order to determine how regulation of the *lac* operon in the bacteria *Escherichia coli* worked. They concluded that a repressor, which is the product of the *lacI* gene, negatively regulated the transcription of β-galactosidase. They also found that certain inducer molecules bind to the repressors thereby altering their binding ability to the operator site on the DNA. The influence of the repressor action on the amount of protein/enzyme produced out of the DNA is described conveniently by what is known as a transcription function or as the gene regulation function (GRF). Yagil and Yagil (1971) demonstrated that in the case of the *lac* operon, the transcription function was of sigmoidal shape. By elementary reaction kinetics, these functions can be shown to be nothing but *Hill* functions of the type $f(R) = \frac{\beta}{[1+(R/k_D)^n]}$ where β is the basal transcription rate in the absence of repressor binding the DNA, R being the amount of repressor. k_D is the concentration of repressor yielding half-maximal expression which is nothing but the dissociation constant of the binding and un-binding process. A *Hill* coefficient of $n = 1$ indicates a Michaelis-Menten kind of reaction mechanism, while $n > 1$ indicates cooperativity amongst the repressor molecules in binding on to the operator region of the DNA. The *Hill* coefficients are typically estimated by fitting the transcription functions with data of the amount of proteins and mRNAs. These functions are then used to solve for the time-evolution of the molecular species involved in gene regulation. One could also consider different forms for the transcription functions, whether linear or non-linear, depending on the nature of the regulatory activity between the TFs and the promoter.

Once the transcription functions are defined, the regulatory network is modelled by a set of differential equations that describe the time-evolution of the variables. For example, in a system with N genes or rather N variables having interactions with each other, the time-evolution of each variable is given by,

$$\frac{dX_i}{dt} = f_i(X_1, X_2, \cdots, X_N) \qquad (i = 1, 2, , \cdots, N) \tag{1.1}$$

where f's are the corresponding transcription functions that capture the form of interactions. Chen et al. (1999) use a protein-mRNA model involving first order differential

equations with a linear transcription function on the *rhs* of the above equation. Their model is

$$\frac{dM}{dt} = f(P) - \gamma_m M, \qquad \frac{dP}{dt} = T_p M - \gamma_p P \tag{1.2}$$

where M and P are the vectors containing the concentrations of all the N number of genes in the system and are functions of time t. While γ_m and γ_p are the $N-$dimensional diagonal matrices of the decay rates of mRNAs and proteins, T_p has translation rates of proteins as its diagonal elements. The transcription functions $f(P)$ were considered to be linear functions of the protein variables P, thereby enforcing regulatory connections between the various genes and also allowing for feedback regulation. However, regulatory links can also be learnt by considering either just the proteins or the mRNAs. As measurements from microarrays consist of only mRNA levels, Hoon et al. (2002) decide to use the mRNAs as the variables in their model and consequently obtain promising results in the case of *Bacillus subtilis* de Hoon et al. (2003). Whilst Sakamoto and Iba (2001) use nonlinear transcription functions, Gebert et al. (2007) use piecewise linear functions that are better in capturing the dynamical behaviour of different types of systems. The parameters are then estimated through various learning methodologies by utilizing the time-series data of the mRNA expression levels. Finally one obtains a model that best fits the given data and reveals the possible regulatory links between different genes.

1.4.2 Monitoring gene expression levels

To determine a gene regulatory network, one has to know the levels of mRNAs, which represent the activities of the respective genes in a specific cellular function. Microarrays help in this effort by measuring the mRNA levels corresponding to various genes that are transcribed.

Complementary DNA or cDNA microarray technology Schena et al. (1995) involves fabricating microscopic glass slides over which gene sequences are printed onto by first amplifying these sequences through polymerase chain reaction (PCR). In order to quantify the level of gene expression, the cells are subjected to varying environmental conditions. Simultaneously, another set of cultures is grown in normal conditions, known as the control sample. This allows for the comparison of gene expression of experimental (treatment) samples to normal (control) samples. Infected cells such as cancerous or tumour cells can also be considered as treatment samples. Once the cultures are grown, mRNAs from the cells are harnessed and can now be used as indicators of the expression levels of the corresponding genes under various stress conditions. As RNA is inherently unstable, its complementary copy known as cDNA is synthesized for use in the microarrays. In order to quantify the amount of cDNA hybridized on the microarray, cDNA is

coloured by adding dyes to the samples. The colouring represents the amount of cDNA hybridized, in turn giving an indication of the number of gene transcripts present in the treatment sample. The expression value is actually the ratio of intensity levels of the treatment and normal samples which is log transformed so that a positive value indicates an induced gene, and a negative value indicates a repressed gene. When such expression data is being collected, it is important to note the use of temporal data for any significant analysis being done over such datasets. This is the main principle behind gene expression profiling.

Oligonucleotide technology Fodor et al. (1993); Lipshutz et al. (1999) is one of the other promising types of arraying technologies to have emerged over the last decade. It encourages the monitoring of large number of genes in parallel (about 500,000 probes[3] in a microarray of size of about 1 cm^2). The probes are nucleotide sequences of length 20 bp taken from a short sequence (around 300 bp) of a gene or an Expresssed Sequence Tag[4] (EST). Since probes are designed to be complementary to the gene or EST sequence and independent of sequences of other genes, they attach with high specificity to the target[5] sequences of their own genes. Soukas et al. (2000) employed oligonucleotide arrays to analyze changes in gene transcription with obesity in humans.

Serial Analysis of Gene Expression or SAGE is another such powerful method for analysis of gene expression of thousands of genes Velculescu et al. (1995) and is based on the principle that a short nucleotide sequence of length 10 bp (known as tags) identifies a specific gene. Theoretically, as there are 4 nucleotides types, tags of length 10 bp can identify 4^{10} different genes. This technology is widely used for analyzing gene expressions as in the case of cancers Nacht et al. (1999) and cardiovascular tissues Patino et al. (2002).

1.4.3 Retrieving gene interactions

Once the data related to expression levels are obtained the immediate task is then to estimate the presence of regulatory links between genes. As a first step towards this, genes of similar biological function or those that are transcriptionally related, *i.e.,* co-regulated by the same set of regulators, are grouped together into clusters. The clustering is based on analyzing the expression profiles of their mRNA products over varying environmental conditions and over a specific time-period. Such clustering of genes involved in a common cellular function into one cluster, could lead to the functional annotation of an unknown gene or could even help in mapping the topology of the regulatory network representing the cellular process. Segal et al. (2003); Bar-Joseph et al. (2003) made use

[3]DNA sequences immobilized on the solid subtrate
[4]A short sub-sequence of a transcribed protein-coding or non-protein-coding DNA sequence
[5]DNA or RNA sequence from the experimental sample

of genome-wide location and microarray expression data in reconstructing the regulatory networks of gene *modules* that are sets of co-regulated genes having common transcription factors as their regulators. This is a step in the direction towards identifying the complex regulatory network behind the expression of these genes.

Since the process of gene expression is known to be random in nature and also the fact that such measurements of gene expression levels on a microarray involve noise characteristic of such experimental measurements, they treat the process as probabilistic by considering the expression level of each gene as a *random variable*. Works such as those by Friedman et al. (2000) and Kim et al. (2003) among others make use of the probabilistic and dynamic properties of Bayesian networks to learn the network structures GRNs. Perrin et al. (2003) and others use the framework of Dynamic Bayesian Network (DBNs) to learn the regulatory connectivities. They assume that these measurements of the mRNA expression levels are corrupted by the inherent biological noise and by the measurement noise. The parameters for the Gaussian noise term are then learnt by the well-known Expectation-Maximization algorithm Dempster et al. (1977). However, in section 1.5.2 we shall discuss a statistical issue raised by Chu et al. (2003) with regard to network inference using bulk measurements.

1.5 Challenges

1.5.1 Estimating regulatory links

Estimating the presence of any regulatory action between any two genes involves witnessing correlated variations in the concentration levels of mRNAs when cells are subjected to various environmental conditions. Here we shall give two examples where a simple statistic such as the correlation coefficient is used to estimate the presence of any regulatory link between two genes. While in one case the correlation value is as expected the results in the second case are contrary to what is expected. In *Saccharomyces cerevisiae* GCN20 is said to activate the production of GCN2 under starvation conditions de Aldana et al. (1995). We therefore expect positive correlations between the expression profiles of their respective mRNAs. Gasch et al. (2000) provide the necessary experimental data by subjecting a population of cells to varying environmental conditions. The correlation coefficient for the GCN20 $\rightarrow$ GCN2 link is $+0.0885$. On the other hand, let us consider FLO1 which is a protein involved in flocculation and is known to be activated by FLO8 that is a well-known transcription factor required for flocculation Kobayashi et al. (1996, 1999). Their mRNA expression levels are negatively correlated. This could be due to the fact that the genes FLO1 and FLO8 are active only under flocculation, whereas the data was acquired over a range of other conditions. Therefore functional annotation of

genes becomes a necessity, without which such statistical analyses could give contrary results.

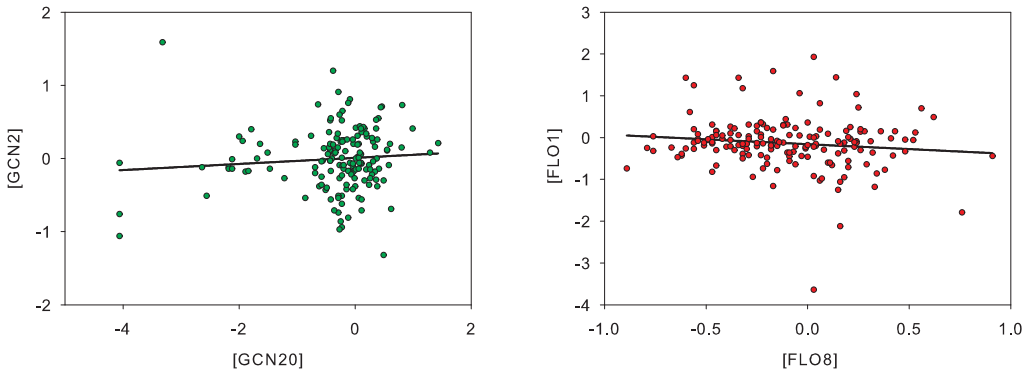

FIGURE 1.3: On the left is the regression plot between mRNAs of the genes GCN20 and GCN2, for which the correlation coefficient is +0.0885. Similarly, on the right, the correlation coefficient between the mRNAs of genes FLO1 and FLO8 is -0.126. The data for this is from Gasch et al. (2000).

1.5.2 A statistical problem for Inference

If regulatory networks are represented by equivalent directed graphs, one could estimate the causal relationships between the genes by evaluating their conditional probabilities. However, for this to be true the variables of the directed graph need to be the expression levels of the corresponding mRNAs or proteins taken from an individual cell. Such measurements would satisfy the notion of a random variable that is necessary for evaluating the conditional probabilities. However, a statistical problem is encountered when the causal relations are estimated from bulk measurements of a variable that take binary values. Danks and Glymour (2001) demonstrate this with the aid of an acyclic directed graph of 4 variables.

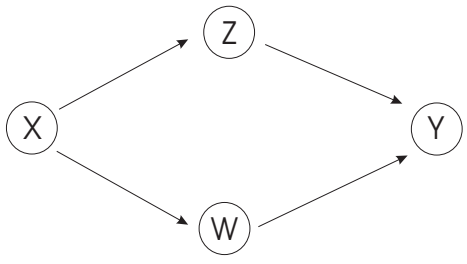

FIGURE 1.4: An acyclic directed graph with genes as variables/nodes.

From the graph of Figure 1.4 one can infer that the random variable X is independent of Y conditional on Z and W. However if measurements of the variables are from a population of n number of cells, the summation of these measurements $\sum_{i=1}^{n} X_i$ is not

independent of $\sum_{i=1}^{n} Y_i$ conditional on $\sum_{i=1}^{n} Z_i$ and $\sum_{i=1}^{n} W_i$. Therefore the causal relationship learnt from such bulk measurements do not satisfy the Markov factorization of $p(W, X, Y, Z) = p(Y|Z, W)p(Z|X)p(W|X)p(X)$. However they further show that there are exceptions such as singly connected graphs, where the conditional independencies hold for the summations as well. Chu et al. (2003) derive two conditions that are sufficient for the conditional independence to remain the same in the case of summation of variables. Therefore while inferring causal relationships in GRNs care has to be taken when the data is obtained from a population of cells.

1.5.3 Debates on Cell-Synchronization

Spellman et al. (1998) analyze and identify those genes whose mRNAs vary over the period of the cell-cycle in the yeast *Saccharomyces cerevisiae*. For this purpose they *synchronized* cell cultures so that each cell in the population is at the same point in time of the cell-cycle. Since the microarrays measure the mRNAs levels in population of cells, synchronization is necessary in order to estimate the true value of mRNA levels from a single-cell. Cells in the population are *arrested* at a particular time point in the cell-cycle, so that on withdrawal of these arresting conditions the cells are believed to grow in synchrony. The concentration levels of mRNAs are then profiled from such a synchronous population to identify those genes whose expression levels vary in accordance to the progression of the cell-cycle. They cluster the expression profiles together as was done by Eisen et al. (1998) based on their similarity of expression pattern over the cell-cycle. They further identify potential binding sites upstream to these genes for well-known regulators that might control their expression of these genes during the cell-cycle. Therefore such grouping of co-regulated genes and further confirmation that such genes share common promoter elements form a good foundation for inferring GRNs.

However, Shedden and Cooper (2002) have questioned the basis of the synchronization techniques used in the above experiments. Since whole-culture synchronization is done via starvation, inhibition of temperature arrest, they argue that the observed cyclic patterns in the mRNA expressions could in essence be the stress response to synchronization due to such perturbation of cells. Therefore they argue that the cyclic patterns might not be a true representation of the real dynamics behind the cell-cycle of an unperturbed or a normal growing cell. Cooper and Shedden (2003) point that such methods only *align* the cells with respect to a particular property such as equal amounts of DNA in G1-phase, but not necessarily *synchronize* them so that they mimic the cell-cycle of a normal unperturbed growing cell. In this regard, an interesting debate took place between two sets of researchers holding differing views on the validity of synchronization procedures — Cooper (2004a); Spellman and Sherlock (2004b); Cooper (2004b); Spellman and Sherlock (2004a). As a conclusion to these debates, Liu (2005) points towards

new and better techniques for synchronizing based on cell age Liu (July 2004). In addition to the above arguments Cooper and Shedden (2007) suggest that the variation in mRNA levels do not necessarily imply a corresponding variation in the respective proteins level, thereby questioning the very notion of mRNAs as the only indicators of cell-cycle control.

1.6 Summary

Irrespective of the above issues relating to synchronization, the ideal data that could reveal more information with regard to biological processes would come from a single cell that is growing normally in its natural environment. Further, such data would include not only the mRNA levels, but expression levels of proteins and other components involved in the regulatory process. Acquiring such time-series measurements are fast becoming a reality in the light of current technological advances in single-cell measurements Muzzey and van Oudenaarden (2009) and advances in tracking single molecules over time. Equipped with such data one can then move forward in the quest for retrieving and characterizing the regulatory activities between genes. The techniques and analyses done in this work would prove to be valuable in such a quest.

Chapter 2

Molecular Fluctuations in GRNs

Gene expression is an inherently stochastic process. As described in the previous chapter the process of gene expression and its regulation consists of biochemical interactions that are random events. This randomness greatly explains the inherent stochasticity that is particular to such biochemical processes. Adding to this is the low number of individual species which amplify the effects of stochasticity. Also, the reaction rates may fluctuate due to variations in RNA polymerase and ribosome numbers. Proteins being central to cellular processing, tracking fluctuations in their levels is important to understand the way in which GRNs regulate the amount of these proteins and timing of their expression. Various methods have been used in determining the source of these fluctuations, such as measuring the variance of protein distribution in a population of cells for changes in the reaction rates such as transcription or translation. Once fluctuations are quantified, models of gene expression are required to explain these fluctuations. If the model is assumed to include just the random birth-death processes, the variance should be equal to the mean concentration levels, which is the character of a Poissonian process. Such a strategy could be employed to identify or characterize the reaction steps that are the source of the fluctuations. In the case of two or more genes forming a regulatory network, the fluctuations in the levels of a protein species that regulates the expression of another gene, would have an impact on the protein noise of the regulated gene. Therefore, the *network structure* in a regulatory pathway too has a great influence on the protein fluctuations.

Once we have an extensive knowledge of the properties of these fluctuations and the processes causing them, a reverse engineering approach could be applied where these fluctuations provide an insight not only into the actual process of gene expression with its numerous reaction steps, but would also suggest possible network connectivities between various genes that gave rise to the observed fluctuations. Towards this goal,

researchers over the last decade have studied various simplified models of gene expression and simultaneously validated them by tracking molecules in individual cells of a population over time. In this respect:

The intriguing question is whether molecular fluctuations exhibit unique features corresponding to different regulatory systems. By different systems we mean those GRNs that are different in their network structures and/or values of kinetic parameters.

Throughout the thesis our aim is to answer the above question. Towards this objective, it is of paramount importance that we build a good understanding of the sources, consequences of these fluctuations, and the control strategies adopted by nature to utilize this inherent stochasticity for its advantage. Also we shall elaborate various issues at the level of single-gene expression that have generated huge interest in experimentalists as well as system modellers. On the way we also identify some significant experimental studies that claim to identify and assert the factors that determine the fluctuations in the expression of a single-gene.

2.1 Consequences of Molecular Fluctuations

Fluctuations or noise is believed to be used by nature to perform certain important regulatory functions. The most significant use of noise is in creating subpopulations of different phenotypes within an isogenic population of cells. Noise is shown to be the primal factor in the random lysis/lysogeny decision-making of bacteriophage - λ. Arkin et al. (1998) analyzed the effect of fluctuations in the expression rates and other molecular fluctuations on the phage λ-infected *Escherichia coli* cells. These cells consist of a decision circuit known as λ lysis/lysogeny, which is a regulatory circuit comprising mainly of four promoter regions and five genes. Arkin et al. (1998) simulate this regulatory network in detail, for which the reaction rates are known through a vast literature, and they showed how fluctuations in the concentration of Cro_2 and CI_2 leads to the division of the infected cells into lysis and lysogeny subpopulations. This brings to the forefront the significance of noise for the purpose of choosing specific pathways, leading to two different phenotypes.

Another interesting example of noise influencing cell fate decisions is seen in *Bacillus subtilis*. In response to stressful conditions, it was found that a minority of *Bacillus subtilis* cells became *competent* where they have the ability to take up DNA from the environment, while the majority of cells are in the *vegetative* state. This phenotypic variabilty is the result of a single-gene activator feedback loop formed by *comK* resulting in bistability, where one state has high numbers of *comK* corresponding to the competent state while the other state has low numbers of *comK* corresponding to the vegetative

state. *comK* is initially low in concentration due to the action of various repressors on its promoter region. As cells grow exponentially and approach stationary state, *comS* represses the degradation of *comK* which along with the fluctuations in *comK* numbers eventually leads to some cells transiting to the *competent* state.

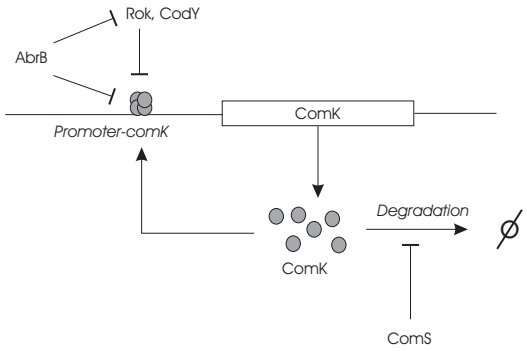

FIGURE 2.1: Competence induction network in *Bacillus subtilis*.

Maamar et al. (2007) observed that the cells that transit to competent state revert back to the incompetent state due to a relatively small decrease in *comK* transcription over a time period of approximately 2 hours. Such an interesting behaviour is definitely a result of a combination of small regulatory networks such as self-activators, repressors, etc. To verify the influence of fluctuations on the transition phenomenon, Maamar et al. (2007) increased the rate of transcription and also lowered the rate of translation thereby decreasing the amplitude of fluctuations while maintaining the mean levels of *comK* Thattai and van Oudenaarden (2001). These alterations in the rates did not allow fluctuations in many cells to cross the threshold in *comK* level which was necessary for the transition, thereby reducing the number of competent cells. On the whole, fluctuations in the level of *comK* is believed to be the cause for the transition between the two phenotypes.

Another example of noise influencing cell-fate decision can be found in the regulatory circuit that is responsible for colour vision in fruit fly (*Drosophila*). In this organism colour vision depends on the yellow(70%) and pale(30%) ommatidial subtypes Feiler et al. (1992). These are distributed randomly in the retina, with corresponding frequencies. The yellow subtype discriminates long wavelength light while the pale does so for the short wavelength. The cell-fate decision is made in cells containing colour-sensitive photoreceptor R7, which sends a signal to cells containing colour-sensitive photoreceptor R8. Wernet et al. (2006) showed that the cell-fate decision was due to the fluctuations in the level of the transcription factor *spineless*, large fluctuations of which lead to R7 cells committing to yellow fate. The 70% - 30% distribution of the ommatidia could be due to fluctuations in *spineless* being in a particular range.

Stochastic effects in transcription factor proteins affect the regulation of the genes whose *cis*-regulatory upstream sequence it binds to. One of the best examples for this was provided experimentally by Nachman et al. (2007) who demonstrated the variability in the timing of the onset of early meiosis genes due to fluctuations in the level of *Ime1* protein. They conducted experiments on diploid yeast cells by subjecting them to nutritional deprivation conditions that results in the cells going into the meiosis stage of development. While on the one hand, the Ime1 promoter is activated by depletion in nitrogen and glucose, the *Ime1* protein is a known regulator of several genes that are expressed during early meiosis that make up the meiosis initiation pathway. Therefore nutritional starvation induces these downstream genes through *Ime1*. However, on monitoring nearly 4000 individual cells over time using fluorecence microscopy, Nachman et al. (2007) observed that the early meiosis genes regulated by *Ime1* showed very large cell-to-cell variations in their onset times (the time taken for the protein concentration levels to cross a particular threshold). They demonstrated that this variation was mainly due to fluctuations in the level of *Ime1* protein in different cells, rather than other sources of variation such as nutrional history or cell-cycle phase. Though cell size was found to be another cause for the variability of the onset times its effect was shown to be through *Ime1*. Therefore fluctuations in transcription factors greatly influence the functioning of developmental pathways whose genes they regulate.

Noise in protein levels is also believed to affect the functioning of regulatory networks responsible for circadian rhythms in various organisms. Through a deterministic model Leloup and Goldbeter (1998) showed that circadian rhythms in *Drosophila* employ oscillating proteins that have specific periods and amplitudes. Further, when the mRNA and proteins molecules were low in number - typically tens and hundreds of molecules respectively, they demonstrated through stochastic simulations that the oscillatory behaviour is still retained with variations in the period and amplitude. The reasoning for the robustness was due to the network structure of the model. Barkai and Leibler (2000) illustrated the effectiveness of certain class of networks that retained the oscillatory behaviour in the face of noise. Their model comprised of two sets of interacting proteins - an activator activating its own expression as well as that of a repressor, while in turn being regulated by the repressor. In respect to the oscillations being influenced by stochasticity, Elowitz and Leibler (2000) demonstrated this by designing a synthetic network known as the Repressilator comprising of three repressors regulating each other's expression. Another case of stochastic influence on the functioning of a network was demonstrated by Gardner et al. (2000), where they constructed a synthetic toggle switch in *Escherichia coli* that was bistable in nature. They showed that the bifurcation point was blurred due to stochastic effects which lead to emergence of a bimodal distribution. The above cases show that certain network structures and certain range of parameter values enable the regulatory network to perform robustly in face of

stochasticity. Therefore it becomes even more important to study the relation between the molecular fluctuations and the structure and parameters of the network.

2.2 Sources of Molecular Fluctuations

It has for long been observed that protein and mRNA numbers fluctuate at the single-cell level, *i.e.,* the concentration levels of proteins and mRNAs vary from one cell to another in a population or over time in a single cell. This is due to the inherent random nature of the molecular interactions. Other factors such as cell-size, local environment, cell division, and cell-cycle also add to the stochasticity. To get a grasp on the understanding of these various factors, noise is usually classified as *intrinsic* and *extrinsic* to the process of gene regulation. The events of transcription and translation being random in nature cause fluctuations in the protein levels, which are labelled as *intrinsic*. The meaning is more physical in the sense that the source of this noise is believed to be from within the *system*, whereas extrinsic noise on the other hand is emerges due to factors outside the system. A major concern in such categorization is to what is precisely inclusive in this system. The general school of thought is that the events of promoter binding, transcription, translation and degradation of proteins need to be considered as the sources of intrinsic noise. This seems logical as each protein is believed to have its own set of operator regions and transcripts. Fluctuations in ribosomes, RNA polymerases, RNases, or even TFs regulating promoter regions are then grouped under the extrinsic noise category as their activity spans over all the genes. This is true if we go on and accept that the system excludes these molecular species. Experimentally, extrinsic noise affects two reporter proteins equally in any given cell but creates differences in two cells. Extrinsic noise is further classified by many into global noise caused by fluctuations in the reaction rates due to fluctuations in ribosomes, etc, and gene or pathway-specific noise caused by fluctuations in a specific transcription factor or stochastic events in a specific signal transduction pathway. These categorizations of noise are mainly used for experimental confirmations and less so for analytical purposes. However Swain et al. (2002); Thattai and van Oudenaarden (2001) derived analytic expressions for noise in protein/mRNA levels that can then be interpreted conceptually as intrinsic/extrinsic noise terms. Thattai and van Oudenaarden (2001) considered a basic model of single gene expression (boxed area in figure 2.2) to derive the expression for noise in protein levels in terms of the reaction rate constants. As the model excludes extrinsic variables the protein noise was of intrinsic nature.

To quantify the fluctuations in the protein levels Swain et al. (2002) use the following as an expression for protein noise,

$$\eta^2(t) = \frac{\langle p(t)^2 \rangle - \langle p(t) \rangle^2}{\langle p(t) \rangle^2},$$

where $p(t)$ is the protein concentration at time t and angled brackets denote averages over a population of cells. The protein noise is therefore variance over mean squared. Assuming that variables could be grouped as intrinsic or extrinsic, where time t is also considered as an extrinsic variable, the total protein noise measured over a population of identical cells is given by,

$$\eta^2_{tot} = \frac{\overline{\langle p^2 \rangle} - \left(\overline{\langle p \rangle}\right)^2}{\left(\overline{\langle p \rangle}\right)^2} = \frac{\overline{\langle p^2 \rangle - \langle p \rangle^2}}{\left(\overline{\langle p \rangle}\right)^2} + \frac{\overline{\langle p \rangle^2} - \left(\overline{\langle p \rangle}\right)^2}{\left(\overline{\langle p \rangle}\right)^2} \equiv \eta^2_{int} + \eta^2_{ext},$$

where overbar indicates the averaging over the extrinsic variables. Swain et al. (2002) use the above classifications of noise and derive expressions for the same in terms of the reaction rates for a detailed model comprising events such as binding and unbinding of RNA polymerase to the promoter region and formation of an open complex for transcription, binding and unbinding of ribosomes or degradosomes to the mRNA for translation or degradation respectively. In the case of a simplified model they show that the solutions to the intrinsic noise term is similar for the case of the detailed model. Such a theoretical framework for the fluctuations paved the way for the design of experiments using synthetic circuits that could identify the sources of noise by measuring the variability in protein levels in a population of cells. Elowitz et al. (2002) built strains of *Escherichia coli* incorporating cyan (*cfp*) and yellow (*yfp*) fluorescent proteins at a specific chromosomal loci. These proteins were placed under the control of identical promoters. The fluorescence levels were then measured in individual cells by a microscope providing an indication of the protein concentration levels. Intrinsic noise was the cause of variation in protein levels over time in a single cell and therefore its absence was indicated by equal presence of *cfp* and *yfp* in single cells. Such experiments by dual reporters provided a first direct glimpse of the effects of different noise sources in gene expression. While the above analyses was in the case of a single gene, intrinsic fluctuations in levels of its protein product act as a source of extrinsic fluctuations in the case of another gene whose expression it regulates. Therefore noise in an upstream genetic component trickles downstream and has an effect on the functioning of downstream genes in a cascade Pedraza and van Oudenaarden (2005).

To build a general model for noise in protein levels researchers performed genome-wide studies on organisms where they monitored protein levels in hundreds of genes and observed interesting properties of protein noise. Bar-Even et al. (2006) analyzed expression levels, at the cellular level, of 43 proteins in yeast under 11 different experimental conditions and observed that the noise in their levels, defined by variance over mean squared, increased for a proportional decrease in the average levels. They observed that for proteins having intermediate levels of expression, this scaling of noise is similar to the one derived through theoretical models Paulsson (2004, 2005) that not only consider the random birth-death of proteins, but also include the random birth-death of mRNAs and the random nature of gene activation-deactivation (as in Figure 2.2). However,

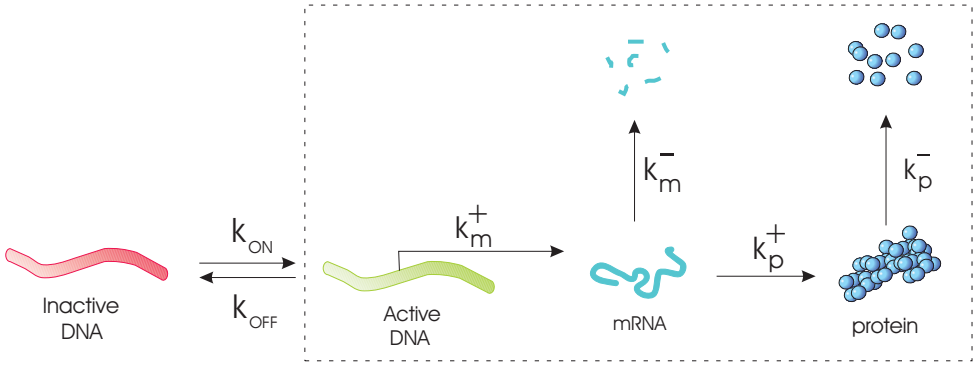

FIGURE 2.2: A simple model of single gene expression. Activation of DNA could be due to factors such as chromatin remodeling or due to activation by upstream TFs. Transcriptional bursting is explained by the switching of DNA between the active and inactive states. Thattai and van Oudenaarden (2001) consider the model [enclosed area] where mRNA transcripts are produced by a DNA that is constantly in its active state.

for proteins with high expression levels, the global factors might be the main source for noise. They also observed high noise in stress-related genes that may be due to variations in the expression level of a common regulator such as the MSN4 transcription factor that is known to regulate the expression of many of these genes. Similarly, Newman et al. (2006) observed in yeast that proteins belonging to different functional groups exhibited different noise behaviours. For example, proteins that responded to changes in the environment, such as those involved in stress-response, heat shock and amino acid biosynthesis were all found to have high levels of protein noise. On the other hand, ribosomal proteins and proteins involved in translation initiation and degradation exhibited low variation. This could be due to common factors that equally affected all the genes in the regulatory pathway. Such genome-wide analysis gave a broad picture of the generalized protein noise.

2.3 Translational and Transcriptional Bursts

An interesting aspect to stochastic gene expression, which has been observed in recent experimental studies, is that of bursting. It has been observed that proteins are produced in bursts from each translation event and that the average burst size is geometrically distributed Yu et al. (2006). The burst characteristics along with steady state protein distributions enable one to propose models of gene expression that yield the observed distributions. In section 2.5 we elabotare on ways of controlling the burst size and the steady state protein distributions. Thattai and van Oudenaarden (2001) derived the expression for protein noise (variance over mean) in terms of the protein burst size. For this, they considered the representation of the single-gene expression (boxed area in Figure 2.2) where m and p refer to the mRNA and protein molecules and $k_m^+, k_p^+, k_m^-, k_p^-$ represent the rate constants of the transcription, translation, and the decay processes.

Protein noise or fluctuations is typically quantified by the Fano factor as,

$$\frac{\sigma_p^2}{\langle p \rangle} = \frac{b}{1 + [k_p^- / k_m^-]} + 1 \approx b + 1, \qquad \text{if } (k_p^- \ll k_m^-) \tag{2.1}$$

where $b = k_p^+ / k_m^-$ is the *burst* size or the average number of proteins produced per single mRNA. One could directly visualize the effect of various rate constants on noise from such an expression. Since variance equals mean for a Poissonian process *i.e.,* the Fano factor is equal to 1, the protein noise in this case deviates from the Poissonian behaviour. The mRNA noise is however Poissonian in character due to simple random birth-death process with exponentially waiting times between events. Therefore $\sigma_m^2 / \langle m \rangle = 1$. Coming back to the protein noise, it can be seen that as the burst size increases, the protein noise along with the mean protein value $\langle p \rangle = k_m^+ . b / k_p^-$ increases. To prove this, Ozbudak et al. (2002) conducted experiments incorporating the green fluorescent protein (*gfp*) reporter gene into the genome of *Bacillus subtilis* and performed point mutations in the regulatory region of the promoter and the ribosome binding region to alter the transcriptional and translational efficiencies so that k_m^+ and k_p^+ are respectively varied. This resulted in larger bursts for low transcription and high translation rates, and smaller bursts for high transcription and low translation rates. We demonstrate the influence of reaction rates on the burst size through sample simulations in Figure 2.6, where the average burst size $b = \frac{15}{\ln(2)/7} \approx 150$ and in Figure 2.7 where $b = \frac{3.7}{\ln(2)/7} \approx 37$. These simulations and the experimental studies of Ozbudak et al. (2002) show the correspondence they have with the analytically derived expressions for protein noise.

The process of translational bursting could be best visualized by tracking single molecules of proteins in time in individual cells. Cai et al. (2006) did exactly that and detected single-molecules of proteins by monitoring fluorescence in single cells of *Escherichia coli* that are trapped in microfluidic chambers. They observed that β-galactosidase was produced in bursts and that the distribution of this burst size was exponential with an average of seven proteins per burst. Similarly Yu et al. (2006) too observed geometric distributions of burst sizes of a fluorescent protein Venus. They fused Venus with a membrane protein Tsr, in *Escherichia coli* cells thereby slowing down the diffusion of the proteins which now bind to the cell membrane. This allowed for the detection of single molecules of the protein. The average number of proteins per burst was 4.2. However, while single molecule measurements yielded geometric distributions of burst size in gene expression, it has been shown by Ingram et al. (2008) that these single-parameter geometric distributions that can be fit to the data cannot distinguish between various combinations of the rate constants used in the models. The protein burst size distribution was obtained by summing the distribution of the number of protein molecules produced by a single mRNA McAdams and Arkin (1997), and the distribution of the number of mRNAs produced during the active state of DNA. With the aid of a *standard* model for gene expression (Figure 2.2 plus ribosomes and RNA polymerases) Ingram

et al. (2008) demonstrated that the distribution of protein burst size is geometric in character and is therefore determined by a single parameter. Consequently many combinations of the rate constants of the standard model resulted in the same distribution making it impossible to ascertain the contributions of different reaction events towards protein noise. However, they further showed that for the combinations of rate constants that gave the same geometric distribution for the burst size, the steady state protein distributions were quite different. Therefore, the hope was that both the burst size distribution and the steady state distribution would together help in determining the reaction events that are responsible for protein noise. However, while the moments of the steady state distribution give more information with regard to the process behind it, being stationary in nature they may not account for the dynamics of the reaction events. This is particularly significant due to the inherent delays in transcription, translation and decay processes. Therefore we employ a time-dependent statistic that can extract more information regarding the processes.

The above issues are related to protein bursts. However, it has been widely observed that mRNAs are also produced in bursts. Golding et al. (2005) observed that transcription occured in bursts and that the mRNA burst size was geometrically distributed with time intervals between each burst being exponentially distributed. To monitor individual mRNA molecules they used *in vivo* tagging of mRNA in living *Escherichia coli* cells. It was observed that though the variance in mRNA levels was proportional to the mean value, the proportionality constant was around 4 which was more than that for a Poisson distribution ($\sigma_m^2/\langle m\rangle = 1$). To account for this deviation from the Poissonian statistics, it was proposed that the gene was randomly being switched into the active and inactive states. This switching is assumed to be at exponentially distributed time intervals and when the gene is in the ON state, transcription is a Poissonian process giving rise to geometrically distributed mRNA molecules being transcribed during the active state of the DNA. Therefore the DNA switching between active and inactive states would give rise to the experimentally observed behaviour in mRNA bursts. Becskei et al. (2005) and Raser and O'Shea (2004) also observed bursts in mRNA production but in eukaryotes. The reasons for the DNA switching between the ON and OFF states could be many. Chromatin remodelling has been widely suggested as the primary reason in eukaryotes. When the chromatin is in a condensed state, the DNA is said to be switched OFF and mRNA production ceases. However, the burst effect could also be simulated by considering the activation of DNA through binding of transcription factors to the upstream regulatory sequences.

Raj et al. (2006) monitored cases of transcriptional bursting by doing a molecular count of the mRNAs in single cells using Fluorescence In Situ Hybridization (FISH), which is a highly efficient technique Vargas et al. (2005). They integrated a reporter gene for a fluorescent protein, into whose $3'$ untranslated region of the coding sequence was

inserted 32 copies of probe-binding sequence, into Chinese hamster ovary cells by electroporation. The cells were then subjected to hybridization with fluorophores that bind to the mRNA molecules at the 32 probe-binding regions which made the detection of individual mRNAs easy. With the aid of their experimental results, Raj et al. (2006) argued that this chromatin remodelling might strongly correspond to gene inactivation and activation. They showed how these events are the principal reasons for transcriptional bursts by studying bursts from two reporter genes incorporated at different locations of the genome. They draw a fine line between gene activation by chromatin remodelling, which may be an inherently random event, and the actual transcription process which is regulated by transcription factors. They further showed that on decreasing the amount of transcription factors that activate a gene, thereby decreasing the average level of mRNA that is being transcribed, the noise properties of the mRNA does not alter. This suggests that the conventional model of gene activation by association of transcription factors does not fit smoothly in the case of higher eukaryotes. In support of the above observations, Raser and O'Shea (2004); Becskei et al. (2005) showed how chromatin condensation with histones acetyltransferases and decondensation by deacetyltransferases are the sources of transcriptional bursts. On the other hand, studies by Bar-Even et al. (2006); Newman et al. (2006) on yeast indicated that noise in genetic activity is very much independent of the histones. Irrespective of the actual causes for the gene activation, it seemed to be clear that variations in the switching characteristics of the gene affected the burst size. With regard to this, Raj et al. (2006) found that for variations in the transcriptional strength the average burst size was affected rather than the frequency of occurrence of these bursts. The transcriptional strength was increased either by increasing the number of operator sequences where the activator protein binds or by simply controlling the amount of this protein. However, the models corresponding to bacterial cells typically consider the gene to be activated by the binding of transcription factor to it.

In spite of arguments for the occurrence of bursts and the processes behind it, the fluctuations in protein and mRNA levels are in essence due to the inherent random nature of the reactions. The bursts are an additional feature of these fluctuations. To study the effect of reaction rate constants or each reaction step on the fluctuation properties of molecules, it makes sense to start with the simplest of models and only later include the additional processes that are required for bursting.

In Figure 2.3 we show the case where the gene is constantly in the active state and therefore mRNA does not experience bursting. However the random birth-death of mRNAs causes fluctuations in its level and consequently results in a Poissonian distribution in their numbers. The Fano factor for such a case is $\sigma_m^2/\langle m \rangle = 1$.

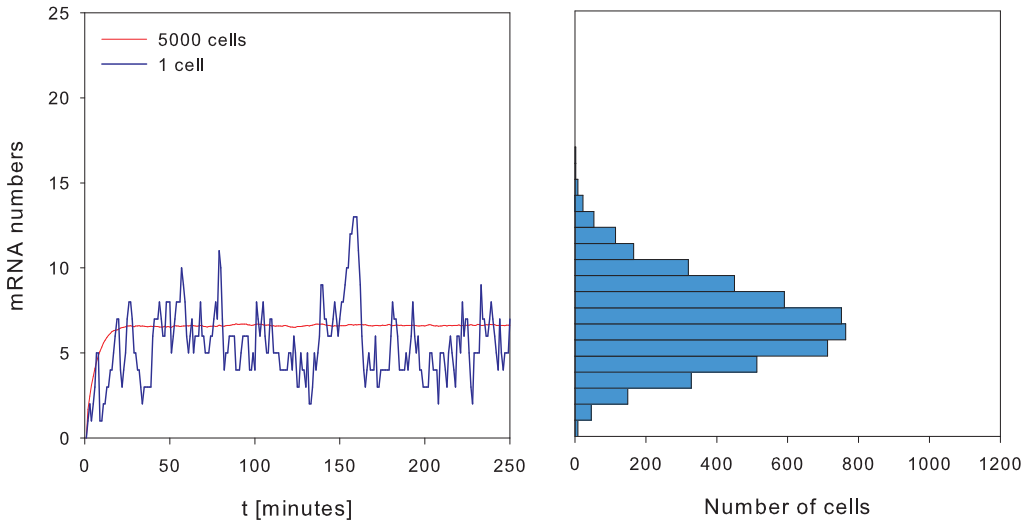

FIGURE 2.3: Time series of the mRNA species in the model of single-gene expression [boxed area of Fig 2.2]. Simulations are done using an algorithm that employs a simple Monte Carlo technique. 1 cell represents 1 run of the system. An ensemble of 5000 cells is simulated and the 5000 runs are averaged to obtain the deterministic solution to the model. The distribution of mRNA numbers around the mean value of 6.6 is given by a histogram. The inconspicuous bursts and the distribution without much skewness are reminiscent of a Poissonian process. This is due to DNA being constantly in the active state. Reaction rates are as follows: transcription rate $k_m^+ = 4\ min^{-1}$, translation rate $k_p^+ = 0.5\ min^{-1}$, mRNA and protein half-lives are 4 and 120 minutes respectively.

Now, let us consider the case where the gene switches between the active and inactive states. For such a case, the Fano factor (which we shall derive in a later chapter) is,

$$\begin{aligned}
\frac{\sigma_m^2}{\langle m \rangle} &= 1 + \left[\frac{k_m^+ k_{OFF}}{(k_{ON} + k_{OFF} + k_m^-)(k_{ON} + k_{OFF})} \right] \\
&\approx 1 + \left[\frac{k_m^+ k_{OFF}}{(k_{ON} + k_{OFF})^2} \right] && \text{if } (k_{ON} + k_{OFF} \gg k_m^-) \\
&\approx 1 + \left[\frac{k_m^+}{k_{OFF}} \right] && \text{if } (k_{OFF} \gg k_{ON}) \\
&\approx 1 && \text{if } (k_{OFF} \gg k_m^+)
\end{aligned} \tag{2.2}$$

From Figure 2.4 we instantly notice the bursts in mRNA level and also the heavy-tailed distribution. mRNA noise in this case is given by Equation 2.2. For higher values of k_{ON} and k_{OFF}, the approximations in Equation 2.2 apply and the mRNA noise tends to be Poissonian (Figure 2.5).

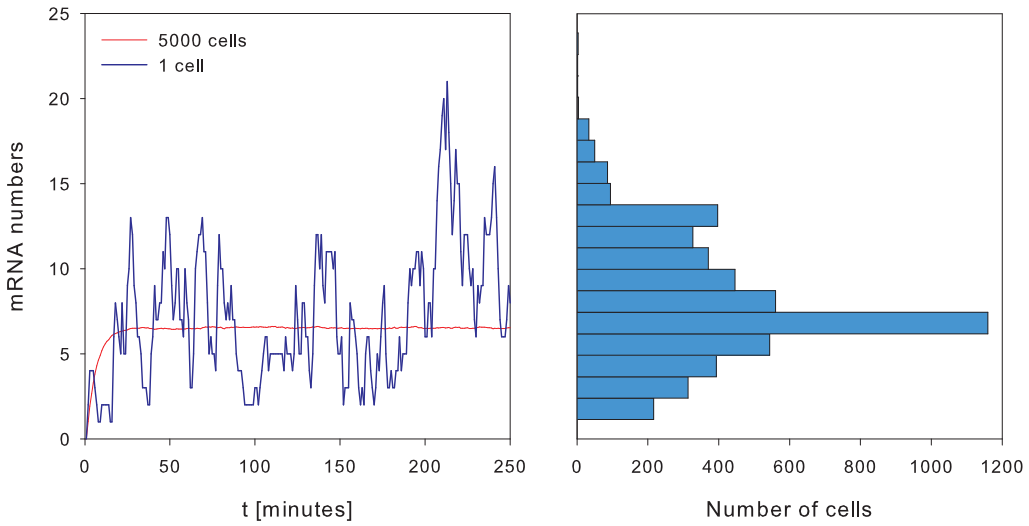

FIGURE 2.4: Time series of the mRNA species in the full model of single-gene expression [Fig 2.2]. The distribution of mRNAs at steady state is Gamma with a heavy tail and also bursts are prominent. Therefore the process is of non-Poissonian character and is due to DNA randomly switching between the inactive/OFF and the active/ON states. ON and OFF rates are $k_{ON} = 1\ min^{-1}$ and $k_{OFF} = 2\ min^{-1}$.

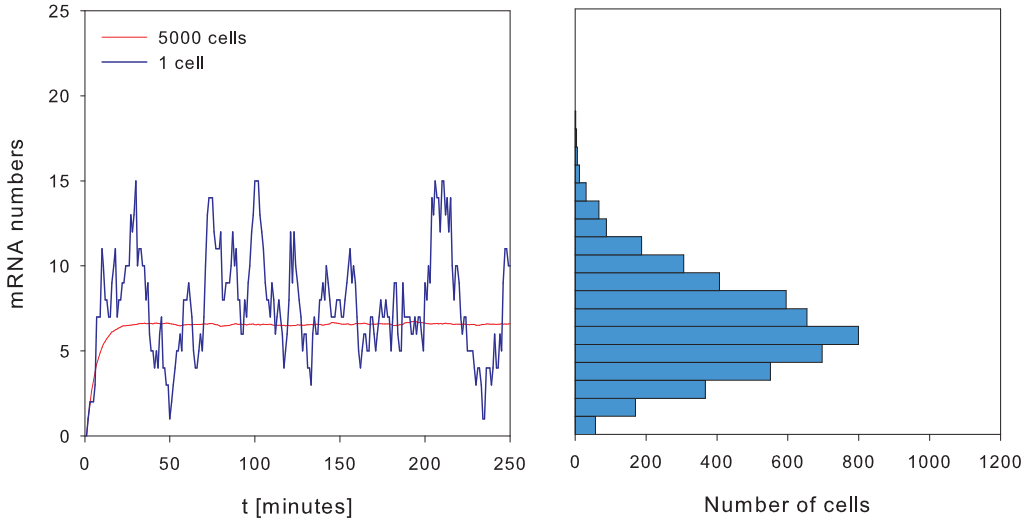

FIGURE 2.5: Time series of the mRNA species in the full model of single-gene expression [Fig 2.2]. Though the bursts are prominent, the distribution loses its heavy tail due to increase in the ON/OFF rates from the values in Figure 2.3 and due to approximation of Equation 2.2. ON and OFF rates are $k_{ON} = 10\ min^{-1}$ and $k_{OFF} = 25\ min^{-1}$.

2.4 Steady state distributions

In the case of a model where the gene transits between the active and inactive states, the mRNA production takes place during the ON state in a Poissonian manner. The steady state distribution of mRNAs in such a case is a gamma distribution as shown Figure 2.4. Note that the discrete analog of the gamma distribution is the negative binomial distribution which defines the probability $P(k)$ that the n^{th} success in a sequence of Bernoulli trials occurs on the $(n+k)^{th}$ trial. In the simple case where the gene is always in the ON state, the mRNA numbers are Poisson distributed as seen in Figure 2.3. In both the cases where the gene is constantly in the ON state and where it switches between the ON and OFF states, the protein numbers are gamma distributed as shown in Figure 2.6. While at lower levels of protein the probability distribution is long-tailed 2.6, higher levels are of near-normal distribution (Figure 2.8).

For the case of mRNAs distributions, Raj et al. (2006) and Golding et al. (2005) observed through experiments that mRNAs had distributions with long and heavy tails. Since the variance scaled according to the mean values but with the proportionaly constant far greater than 1, they proposed the gene activation-inactivation process as the cause for such distributions. In the case of protein distributions, while Cai et al. (2006) observed gamma-distributed β-galactosidase in living *Escherichia coli* cells, Sato et al. (2003) observed log-normal distributions of mutant *gfp*s in *E. coli*. The log-normal distribution indicates that a long-tailed distribution when transformed to the logarithmic scale is a near-normal distribution. Raj et al. (2006) too observed protein distributions which were log-normal-like in their experiments. Through the above experiments and analytical evaluations of relevant models we note that the shapes of the mRNA/protein distributions are worthy of being analyzed in order to gain more insight into the actual process behind them.

2.5 Effect of parameters and network structure on stationary fluctuations

The purpose of modelling noise in protein and mRNA levels is mainly to study the influence of various reaction mechanisms on it. Experimental verifications of such analytical expressions strengthen the justification for the study and modelling of simple genetic networks. One such example of experimental verification is the work of Ozbudak et al. (2002) where they show that the level of phenotypic variation in an isogenic population can be regulated by genetic parameters. They incorporated a green fluorescent protein gene as a reporter into the chromosome of *Bacillus subtilis*. A gene with low transcription but high translation rates produces bursts that are large, variable and infrequent,

resulting in large fluctuations. Conversely, a gene with high transcription and low translation rates produces bursts that are small and frequent, causing only small fluctuations in protein numbers and producing a smaller phenotypic variation in the population. We demonstrate the above point through simulations for the cases of large translation rate (Figure 2.6) and for low translation rate (Figure 2.7). An additional case is simulated (Figure 2.8) where only the protein decay rate is decreased so that the mean protein level increases and the steady state distribution is more Gaussian-like.

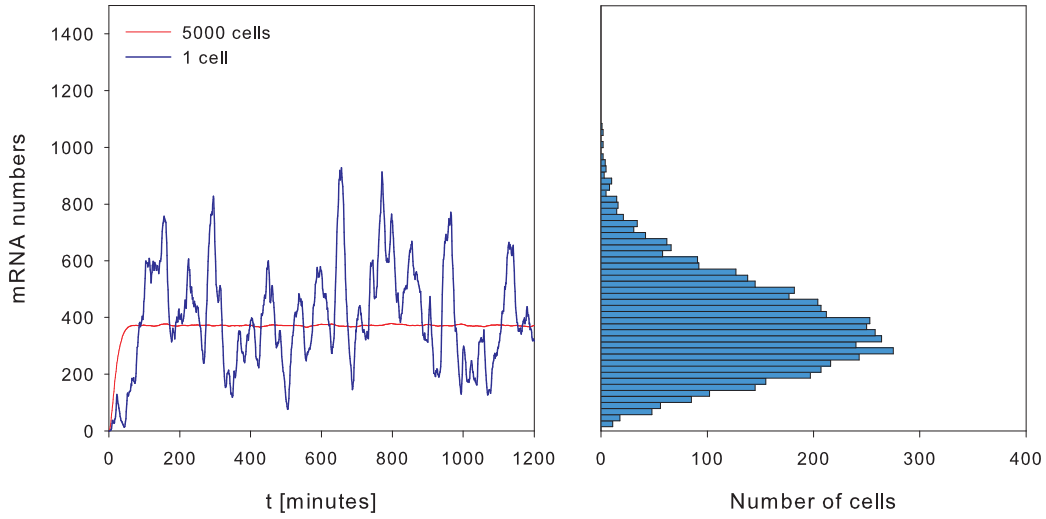

FIGURE 2.6: Protein numbers in a model of single-gene expression with parameters: $k_p^+ = 15\ min^{-1}$, $k_m^+ = 0.25\ min^{-1}$, Protein and mRNA half-lives are of 7 min each. Steady-state mean value of protein is $\langle p \rangle \approx 375\ nM$.

Ozbudak et al. (2002) suggest that several inefficiently translated regulatory genes have been naturally selected for their low-noise characteristics, even though efficient translation is energetically favourable. Inefficient translation is energetically unfavourable because high energy phosphate groups are hydrolyzed to drive the synthesis of unused or little-used transcripts. Similarly in the case where intrinsic noise is caused by promoter fluctuations, frequent promoter activation events followed by an inefficient transcription will result in less noise in mRNA levels than infrequent promoter fluctuations followed by efficient transcription Rosenfeld et al. (2005). Such simple analysis invariably shows that the process of gene expression has inherent mechanisms for controlling noise. Ozbudak et al. (2002) to a large extent have substantiated this by simple yet powerful experiments. Their experiments validated the work of Thattai and van Oudenaarden (2001) who adopted a simple analytic approach to prove that the mean and variance of protein number can be controlled by varying system parameters for single gene expression.

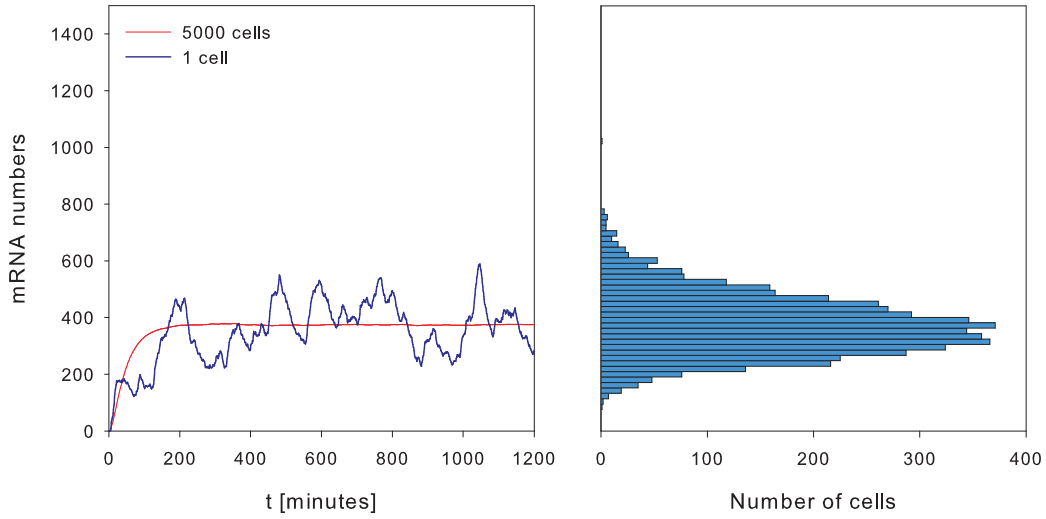

FIGURE 2.7: Protein half-life is increased from 7 *min* (Figure 2.6) to 28 *min* while the translation rate is reduced to $k_p^+ = 3.7\ min^{-1}$ from the values in Figure 2.6, so that mean protein value remains the same, while the average burst size b and the variance σ_p^2 reduce (Equation (2.1)).

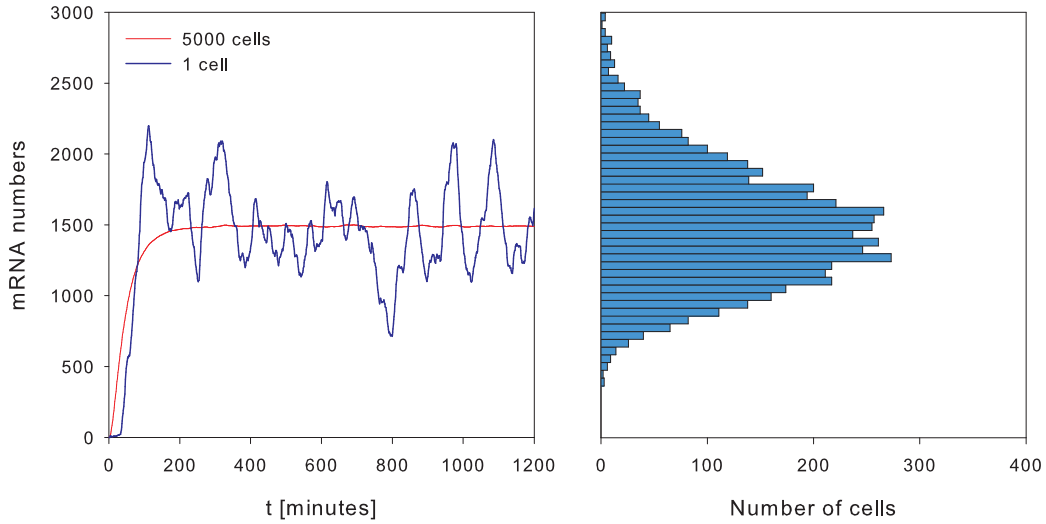

FIGURE 2.8: The only change in the parameters is to the protein half-life which is increased four-fold from 7 *min* (Figure 2.6) to 28 *min*. Consequently the mean protein value increases to $\langle p \rangle \approx 1500\ nM$ with a corresponding increase in the variance (Equation (2.1)). Notice the Gaussian-like distribution as compared to Figure 2.6.

Once protein noise is characterized in single genes the next step is to study the influence of regulating proteins on the expression of the regulated gene. In other words the interest now lies in analyzing the effect of the network structure on noise. Towards this end, Pedraza and van Oudenaarden (2005) designed a synthetic linear gene regulatory network of four genes, where three of them were monitored in single *Escherichia coli* cells by cyan, yellow and red fluorescent proteins. The response of single cells to various amounts of inducers such as isopropyl-β-D-thiogalactopyranoside (IPTG) was measured using fluorescence microscopy or flow cytometry. This allowed them to analyze the effect of noise in the upstream proteins on the downstream protein. A corresponding model was developed that considered noise in a gene to be determined by its intrinsic fluctuations, propagated noise from upstream genes, and global noise affecting all genes. Their model is based on the Langevin approach, where deterministic differential equations representing the system are modified by adding stochastic terms that represent both the intrinsic noise, as well as the global fluctuations in cellular components that change the reaction rates for all genes. These experiments clearly demonstrated how noise in a expression of a gene is affected by fluctuations in the levels of the transcription factors that regulate it expression, *i.e.,* how noise propagated in a network of genes.

Hooshangi et al. (2005) conducted similar experimental studies involving synthetic transcriptional networks. They analyzed a cascade of genes that repressed the production of the downstream gene. They arrived at many conclusions from their experiments. Firstly, they observed that the steady-state switching behaviour showed high sensitivity with increasing cascade length. Through simulations they showed that the Hill coefficients that represent the steepness of the switching functions were 2.8, 7.5, 11 and 29 for cascade lengths of one, two, three and seven respectively. The more interesting observation was that the longer cascades increased variations in the protein levels that is seen as variability in levels of cells in the population. Further, the time taken for the downstream components of the cascade to respond to signals at the top of the cascade, which is the response time, increased with increase in the length of the cascade. This was accompanied by loss of synchronization in these response times for longer cascades. This was again a problem in a population of cells in their ability to respond uniformly to a signal. The most important conclusion of their work was that as the cascade length increased, the ability of the system to filter out rapid fluctuations in protein levels *i.e.,* the system behaved as a low-pass filter. This was advantageous since the downstream component respond to only legitimate changes in the levels of the upstream component rather than responding to rapid and often unwanted fluctuations.

Regulation could also be effected through negative feedback loops where gene expression is repressed by its own product or by other TFs. It has been observed that negative feedback reduces noise while maintaining the mean levels in proteins. Becskei and Serrano (2000); Austin et al. (2006) analyzed autoregulated negative feedback loops in *Escherichia coli* cells where the protein product of the gene represses its own expression and demonstrated that indeed noise reduced for negative feedback. Similarly, Dublanche et al. (2006) analyzed the negative feedback experimentally and through simulations. They showed that not only is noise reduced in proteins which are auto-repressed but these proteins show stronger negative correlation with the downstream protein whose production they regulate.

2.6 Summary

In this chapter we have described in detail how analytical expressions that quantify the molecular fluctuations are useful in analyzing the various aspects of gene expression. Researchers have recently quantified molecular fluctuations through elegant and powerful analytical models Kepler and Elston (2001); Swain et al. (2002) and also substantiated through experimental studies. To understand the mechanisms by which molecular fluctuations arise it makes sense to study synthetic regulatory networks and also networks in live cells where noise is seen to have a profound effect on the overall functioning of the regulatory system. The numerous experimental studies described here and some significant single molecule techniques such as Sigal et al. (2006a) Yu et al. (2006), Raj et al. (2008) greatly enhance our knowledge of the microscopic world. As more useful data is churned out through such experiments, it is imperative that analytical formulations and numerical simulations are used beforehand to increase our understanding of GRNs by analyzing the inherent molecular fluctuations.

Chapter 3

Dynamic Correlation Functions

Given the stochasticity of the very processes that constitute regulatory responses to stimuli, it is likely that correlated fluctuations of the molecular species will illuminate the dynamics of regulatory interactions. In fact, this is the intuition behind the regression hypothesis by Onsager (1931) who said that *the average regression of fluctuations will obey the same laws as the corresponding macroscopic irreversible process* and which has been developed further in several fluctuation-dissipation theorems even away from equilibrium Keizer (1987); Speck and Seifert (2006). A simplest case demonstrating the Fluctuation Dissipation Relations (FDR) is Brownian motion, where the frictional force exerted by the medium in which the Brownian particle is present and the random force due to the collisions of other particles are interlinked. This relation is formalized into the Fluctuation-Dissipation Theorem (FDT) stating that the systematic part of the frictional force is determined by the correlations in the random force. Marconi et al. (2008) give an extensive review of FDR and demonstrate its relevance in many applications. Coming back to our problem, we consider correlated fluctuations in elementary fragments of GRNs to illustrate what they can tell us about the nature of the regulatory function enacted by the network. On the way, we shall also give a simple demonstration of the relation between the deterministic response to perturbation and the internal fluctuations of molecular species.

The analytical framework of this chapter builds upon the relations between the macroscopic dynamics and the fluctuation properties of the system. This is given by the FDT. In statistical physics, the FDT states that if a thermodynamic system responds linearly to an external perturbation, then the amount by which it responds is simply related to the fluctuation properties of the system. In other words, it proposes that there is an explicit relationship between the internal fluctuating force that is random in nature, and the observed macroscopic or deterministic response of the system to an external force that governs the dynamics of the *averages*. In the present context of biochemical reacting systems, this relation emerges naturally out of a system-size expansion of the

chemical master equation van Kampen (2007). This linear noise approximation (LNA) is equivalent to the Fokker-Planck equation for these processes Gillespie (2000). The distributions of the fluctuations around the average concentrations of the molecules, which is assumed to be Gaussian in nature, is then the solution to the LNA. In the next section, we start out by solving for the moments of the distribution of the molecular concentrations and go on to derive the dynamic correlations between the fluctuations of these molecular species. On the way, we observe the emergence of a relationship between the dynamics of the averages and the fluctuating part of the distributions through a common *source*.

In the previous chapter we demonstrated how noise has been quantified by simple formulations allowing one to study the effect of various network topologies, regulatory mechanisms and parameters on the second-order stationary statistics of GRNs Tao et al. (2007); Tomioka et al. (2004); Pedraza and van Oudenaarden (2005). Such observations made it possible for the quantification of noise to reveal the regulatory links between genes Austin et al. (2006); Cox et al. (2008). However, a crucial element missing in such studies was the temporal aspect. Since the entire process of gene expression is dynamic in nature, crucial insights could be gained by tracking the fluctuations in species numbers over time. For example, due to the inherent time-delay in the transcription and translation steps, a perturbation in the upstream process of gene activation say, would result in a delayed response in the protein fluctuations. Such a response could be evaluated by the temporal correlations that are more likely to be characteristic *of* the structure, regulatory mechanisms and parameter values of the system. Such studies are possible with advances in technology in tracking of mRNA/protein numbers over time in individual cells becoming a reality Raj et al. (2008). With the aid of time-lapse fluorescence measurements, recent reports suggest that the time dependent correlations of fluctuations in protein levels indicate the presence of regulatory activity in simple GRNs Dunlop et al. (2008); Sigal et al. (2006b). That is the object of our study.

Hence, our main focus in this work is to derive the time-correlation functions in a sum-of-exponentials form (Equation (3.14)) and use them for studying the dynamic correlations between two species in a GRN, such as proteins of the regulator and regulated genes. We expect distinct behavioural patterns in the correlated fluctuations for variations in the regulatory networks. We consider two-gene regulatory networks and shall introduce the variations in these networks by employing three strategies: (*i*) adopting different sets of values for the reaction rate constants (*ii*) adopting different regulatory mechanisms; such as simple activation and repression with or without dimerization, and (*iii*) introduction of an additional player *i.e.,* an additional gene into the model. They amount to varying the parameters, regulatory mechanisms and the network structure respectively, and hence form the basis for deriving qualitative relationships between the system characteristics and the fluctuation properties of the GRN. In chapters 4, 5 and

6, we demonstrate how the dynamic correlation function between protein species of two genes, changes its shape, magnitude and temporal *features* for changes in the values of the rate constants, the type of regulatory mechanism between the two genes and finally the structure of the regulatory network. Therefore, by tracking the species numbers in single cells over time, one could predict the type of regulatory activity present in a GRN and also the network structure. In this thesis, we show that this is possible at least in the simple cases of two-gene networks. In essence, the thesis aims at providing a firm theoretical foundation necessary for analyzing complex and larger regulatory networks and illustrates the usefulness of dynamic correlations in distinguising these networks. Some of the terms that we shall use consistently throughout are:

- The mRNAs and proteins are the *species* or *variables* of the regulatory network.
- A genetic unit comprising of the mRNA and protein of a gene is termed as a *node* of the regulatory network. If X is a gene-node then M_x would represent the mRNA of X.
- The rate constants of the reactions are the *parameters* of the gene regulatory system.
- *System* refers to a regulatory network that works independently and has 2 to 3 genes at most.
- By *model*, we mean a system having specific constituent species. Hence, the greater the number of types of constituent species, the more complex the *model*.
- The *regulatory mechanism* determines the form of the regulating function between two genes, which means that an activator system has a different *regulatory mechanism* to that of a repressor system. Similarly regulation via protein dimers is considered a different mechanism.
- A system built up of a set of elementary reactions that are specific to it, is termed to have a specific *structure* or *connectivity*.
- *Upstream* gene is one whose protein product is a transcription factor for a *downstream* gene and hence regulates its transcription.

The most common approach of mathematical modelling of the gene expression process, or in fact most biochemical processes, is by considering them to be of deterministic nature. The deterministic approach is based on the law of mass action, an empirical law deriving a simple relation between the concentrations of all variables and the reaction rates. Here, the dynamics or rather the time-evolution of all the variables in the system is described in a deterministic and continuous sense by first-order differential equations

better known here as Chemical Kinetics Equations. But as chemical reactions are probabilistic events invariably involving discrete, random collisions of reactant molecules, it is feasible that deterministic formalisms will leave out essential aspects of the behaviour of the network. Hence, one needs to resort to stochastic modelling in order to characterize such biochemical systems based on the properties of their internal fluctuations, which would otherwise be not possible by macroscopic analyses. Biochemical systems within individual cells occur typically in heterogeneous environments which have small volumes. This combined with the low copies of the involved molecules, accentuates the effect of fluctuations, which is better studied by adopting the stochastic approach as opposed to macroscopic methods. The biological significance of these fluctuations was discussed in detail in the previous chapter. These fluctuations have been studied experimentally and analytically in simpler ways. Since a biochemical system constitutes reacting species that are treated as stochastic random variables, they are described by probability distributions that evolve in time. The means or averages of these distributions evolve in time according to the system's deterministic dynamics. The Chemical Master Equation (CME) describes the time evolution of these probability distributions, whose *variance* also evolve in time. This would then be the source for deriving the dynamic covariances between any two species of the biochemical system.

3.1 The Chemical Master Equation

Let us consider a system with N reacting species $\{s_1, ..., s_N\}$ that react according to M reactions $\{r_1, ..., r_M\}$ within a small volume v at a constant temperature. The dynamical state of this system can be specified as $\mathbf{X}(t) \equiv (X_1(t), ..., X_N(t))$, where

$X_i(t) \equiv$ the number of s_i molecules in the system at time t.

$\mathbf{X}(t) =$ state of the system at time t.

For such a reacting system, let us consider an infinitesimal time interval dt within which the probability for two or more reactions to occur is negligible. We then have the following definitions:

$$
\begin{aligned}
a_j(\mathbf{X})dt \;\equiv\;& \text{the probability that one } r_j^{th} \text{ reaction will occur inside} \\
& \text{the system of volume } \Omega \text{ in the next infinitesimal} \\
& \text{time interval } [t, t+dt) \qquad (j = 1, \ldots, M) \\
\nu_{ij} \;\equiv\;& \text{the change in the number of } s_i \text{ molecules} \\
& \text{produced by one } r_j^{th} \text{ reaction} \qquad (j = 1, \ldots, M) \;\; (i = 1, \ldots, N)
\end{aligned}
$$

$a_j(\mathbf{X})$ is called the *propensity* function which has been shown to have solid microphysical basis rather than just a way of shochastization of deterministics chemical kinetics Gillespie (1976). It is the product of the reaction parameter k_j and the number of reactant combinations h_j for each r_j^{th} reaction, where

$$\begin{aligned} k_j dt &\equiv \text{average probability, to first order in } dt, \text{that a particular combination of} \\ &\quad r_j \text{ reactant molecules will react accordingly in the next time interval } dt. \\ h_j &\equiv \text{number of distinct molecules reactant combinations for reaction} \\ &\quad r_j \text{ found to be present in } v \text{ at time } t. \end{aligned}$$

The propensity function is then $a_j(\mathbf{X})dt \equiv h_j k_j dt$ indicating that the evolution of the state vector $\mathbf{X}(t)$ is a jump-type Markov process on a non-negative N-dimensional integer lattice. One can then easily establish that the evolution of such a system in time is given by the CME. This is shown below. If $P(\mathbf{X}, t \mid \mathbf{X}_0, t_0)d^M\mathbf{X}$, (which is the probability that $\mathbf{X}(t) = \mathbf{X}$, given $\mathbf{X}(t_0) = \mathbf{X}_0$), is the singly conditioned probability mass function of the random variable $\mathbf{X}$, we are then interested in the time evolution of this probability function. Then the probability of the system being in state $\mathbf{X}$ at time $t + dt$ is the sum of the probabilities of all mutually exclusive ways in which that can happen via zero or one reaction in $[t, t + dt)$. Therefore, the Chapman-Kolmogorov equation for such a case is:

$$\begin{aligned} P(\mathbf{X}, t + dt|\mathbf{X}_0, t_0) &= P(\mathbf{X}, t|\mathbf{X}_0, t_0) \times \Big[1 - \sum_{j=1}^{M} a_j(\mathbf{X})dt\Big] \\ &\quad + \sum_{j=1}^{M} [P(\mathbf{X} - \boldsymbol{\nu}_j, t|\mathbf{X}_0, t_0)a_j(\mathbf{X} - \boldsymbol{\nu}_j)dt] \end{aligned} \tag{3.1}$$

Now, since the master equation describes how the probability $P(\mathbf{X}, t \mid \mathbf{X}_0, t_0)$ for the state of the system evolves in time, it becomes obvious that the CME is the differential form of the Chapman-Kolmogorov equation. Therefore on taking the limit $(dt \to 0)$ in equation (3.1), the probability of the system being in state $\mathbf{X}$ at time $t + dt$ is defined by the CME as:

$$\begin{aligned} \frac{\partial}{\partial t} P(\mathbf{X}, t \mid \mathbf{X}_0, t_0) &= \sum_{j=1}^{M} \Big[P(\mathbf{X} - \boldsymbol{\nu}_j, t \mid \mathbf{X}_0, t_0)a_j(\mathbf{X} - \boldsymbol{\nu}_j) \\ &\quad - P(\mathbf{X}, t \mid \mathbf{X}_0, t_0)a_j(\mathbf{X})\Big], \end{aligned} \tag{3.2}$$

the initial condition being $P(\mathbf{X}, t_0) = \delta(\mathbf{X} - \mathbf{X_0})$. The CME thus describes how the joint probability distribution of all the species of a spatially homogeneous chemical system evolve in time. Analytical or numerical solutions to the CME are not available in general, *i.e.,* the CME is generally analytically intractable. However, van Kampen's Ω-expansion

provides a way of approximating perturbatively the Master Equation to the Linear Noise Approximation van Kampen (2007).

3.1.1 Numerical simulations

While the CME or its approximations are of prime importance in analytically describing a stochastic system, it is of equal interest to numerically simulate the time evolution of the molecular species within such a system. The Stochastic Simulation Algorithm (SSA) serves this purpose. The SSA allows one to numerically simulate the time evolution of all the molecular species in a chemical system such that it properly takes into account the inherent stochasticity of such systems. The SSA can be said to be an exact *equivalent* to the Chemical Master Equation as both are based on the same microphysical premise. Gillespie (1976) proposed a stochastic simulation algorithm based firmly on microphysical premises, and used simple Monte Carlo techniques to generate the randomness required to evolve the system through discrete time. Gillespie's algorithm was robust and simple to understand and immediately found acceptance by numerical simulators. Once again, the propensity function $a_j(\mathbf{X})dt$ is at the core of the SSA. The algorithm basically runs by generating a random pair (τ, ν) that control the *next-reaction probability density function* $P(\tau, \nu \mid \mathbf{X}, t)d\tau$, which is the probability at time t that the next reaction in v will occur in the differential time interval $[t+\tau, t+\tau+d\tau)$ and will be an r_μ reaction. The algorithm continues by advancing time t by τ and by changing the number of molecules of those species that are involved in the r_ν^{th} reaction. The details of the SSA are given in Appendix A. Efficient variations of this algorithm such as the τ-leap method Gillespie (2001) and the Next-Reaction method Gibson and Bruck (2000) also generated good interest among simulators. It has to be kept in mind that these algorithms do **not** derive numerical solutions to the Master Equation. Both are different in this aspect. In conclusion, we note that $a_j(\mathbf{X})dt$ acts as the basis of the stochastic formulation as exemplified in the Chemical Master Equation and also the Stochastic Simulation Algorithm.

3.2 Dynamic Correlation Functions

Coming back to the formulation of the Master equation, the means $\langle X_i \rangle$ and covariances $\mathrm{Cov}(X_i, X_j) = \langle X_i X_j \rangle - \langle X_i \rangle \langle X_j \rangle$ of the probability distribution are obtained by multiplying the CME (3.2) with X_i and $X_i X_j$ respectively and taking expectations resulting in

$$\frac{d\langle X_i \rangle}{dt} = \sum_{j=1}^{M} \langle \nu_{ij} a_j(\mathbf{X}) \rangle \tag{3.3}$$

and

$$\frac{d\,\mathrm{Cov}(X_i, X_j)}{dt} = \sum_{k=1}^{M} \Big[\langle X_i \nu_{jk} a_k(\mathbf{X}) \rangle + \langle X_j \nu_{ik} a_k(\mathbf{X}) \rangle + \langle \nu_{ik} \nu_{jk} a_k(\mathbf{X} - \boldsymbol{\nu}_k) \rangle \Big] \tag{3.4}$$

where the R.H.S of Equation (3.3) is the *jump moment* with the propensity functions a_j being equivalent to the transition probabilities per unit time and stoichiometries $\boldsymbol{\nu}$ equivalent to the step sizes of a general Markov process. We get an exact equation (in place of the approximate equation (3.5)) for the time-evolution of $\langle X_i \rangle$'s only when the propensity functions are linear in $\mathbf{X}$. Now, in terms of concentrations of the species $x_i = X_i/\Omega$, we see that the rate equations of the averages obey the mass action kinetics of the biochemical system,

$$\frac{d\langle x_i \rangle}{dt} = \sum_{j=1}^{M} \nu_{ij} R_j(\langle \mathbf{x} \rangle) \simeq \sum_{k=1}^{N} A_{ik} \langle x_k \rangle \tag{3.5}$$

where $R_j(\mathbf{x}) := \lim_{\Omega \to \infty} \frac{1}{\Omega} \langle a_j(\frac{\mathbf{X}}{\Omega}) \rangle$ are the deterministic rates of the M reactions and $\mathbf{A}$ is the Jacobian matrix whose elements are:

$$A_{ik} = \frac{\partial(\sum_{j=1}^{M} \nu_{ij} R_j(\langle \mathbf{x} \rangle))}{\partial \langle x_k \rangle} \qquad \text{and}$$

If the rates $R_j(\mathbf{x})$ were linear in $\mathbf{x}$, the terms in equation (3.4) do not involve higher order covariances and thus form a closed set. This yields the closed form equation for the covariances:

$$\frac{\partial \mathbf{C}}{\partial t} = \mathbf{A}\mathbf{C} + \mathbf{C}\mathbf{A}^T + \mathbf{B}\mathbf{B}^T \tag{3.6}$$

where elements of $\mathbf{C}$ are $\mathrm{Cov}(x_i, x_j)$ and $\mathbf{B}\mathbf{B}^T$ is the diffusion matrix:

$$\begin{aligned} \mathbf{B} &= \boldsymbol{\nu} \sqrt{\mathbf{diag}(\mathbf{R}(\langle \mathbf{x} \rangle))} \\ \mathbf{B}\mathbf{B}^T &= \boldsymbol{\nu}\ \mathbf{diag}(\mathbf{R}(\langle \mathbf{x} \rangle))\ \boldsymbol{\nu}^T. \end{aligned}$$

where, $\boldsymbol{\nu}$ is the stoichiometric matrix whose elements are ν_{ij}.

Before we proceed any further, we would like to mention that the above matrices and their inter-relationship can also be obtained through the *Fokker-Planck equation*. By considering $\mathbf{X}(t)$ to be real-numbered and the function $f_j(\mathbf{X}) \equiv a_j(\mathbf{X}) P(\mathbf{X}, t \mid \mathbf{X}_0, t_0)$ to be analytic in this variable, the Taylor's expansion of $f(\mathbf{X})$ in the order $\boldsymbol{\nu}$ is possible, which on substituting in the CME (3.2) and on passing to the limit $dt \to 0$ results in the *Chemical Kramers-Moyal equation*. The truncation of the Kramers-Moyal equation to the second order results in the linear multivariate Fokker-Planck equation Gillespie

(1980, 1996):

$$\frac{\partial P(\mathbf{X}, t \mid \mathbf{X}_0, t_0)}{\partial t} = -\sum_{i=1}^{N} \frac{\partial}{\partial X_i}[\mathcal{A}_i(\mathbf{X}, t) P(\mathbf{X}, t \mid \mathbf{X}_0, t_0)] + \sum_{i,i'=1}^{N} \frac{\partial^2}{\partial X_i \partial X_{i'}}[\mathcal{B}\mathcal{B}^T]_{ii'}(\mathbf{X}, t) P(\mathbf{X}, t \mid \mathbf{X}_0, t_0) \quad (3.7)$$

where,

$$\mathcal{A}_i(\mathbf{X}, t) = \sum_{j=1}^{M} \nu_{ji} a_j(\mathbf{X}, t), \qquad [\mathcal{B}\mathcal{B}^T]_{ii'}(\mathbf{X}, t) \equiv \sum_{j=1}^{M} b_{ij}(\mathbf{X}, t) b_{i'j}(\mathbf{X}, t)$$

and

$$b_{ij}(\mathbf{x}, t) = \nu_{ji} a_j^{1/2}(\mathbf{x})$$

The above formulation is an approximate description of Markov processes whose step-sizes are small. The solution $P(\mathbf{X}, t \mid \mathbf{X}_0, t_0)$ to Equation (3.7) is Gaussian with mean and covariances given by Equations (3.5) and (3.6). This is seen by multiplying equation (3.7) with X_i and integrating upon which we get (3.3). Similarly by multiplying equation (3.7) with $X_i X_j$ and integrating and by further considering the covariances $\langle X_i X_j \rangle - \langle X_i \rangle \langle X_j \rangle$ in place of the moments $\langle X_i X_j \rangle$ we get to equation (3.4). Equation (3.7) is generally valid only for *diffusion*-type Markov processes. In a diffusion process it is hypothesized that $P(\mathbf{X} + \boldsymbol{\xi}, t + dt \mid \mathbf{X}, t)$ describes, for vanishingly small dt, a Gaussian-distributed diffusion away from $\mathbf{X}$ with mean drift of $\mathcal{A}$'s and covariances given by $\mathcal{B}\mathcal{B}^T$'s. Note the relation between the $\mathcal{A}$ and the $\mathbf{A}$ matrices which is as follows:

$$\mathcal{A}_i = \sum_{k=1}^{N} A_{ik} \langle x_k \rangle = \sum_{j=1}^{M} \nu_{ij} R_j(\langle \mathbf{x} \rangle)$$

$$\therefore \quad A_{ik} = \frac{\partial \mathcal{A}_i}{\partial \langle x_k \rangle}$$

An equivalent formulation by approximating the Master equation is obtained through van-Kampen's Ω-expansion van Kampen (2007) where the Master equation is Taylor-expanded in the system volume Ω in powers of $\Omega^{-1/2}$, $\mathbf{X} = \Omega \boldsymbol{\phi}(t) + \Omega^{1/2} \boldsymbol{\xi}$, so that the fluctuations ξ are proportional to the square root of the volume. The expansion gives rise to the *Kramers-Moyal* form of the Master Equation, while a truncation of the expansion to first order returns the macroscopic equations similar to equation (3.5) but in terms of molecular concentrations $\boldsymbol{\phi}(t)$; and to second order gives rise to the *Linear Noise Approximation* (LNA) which has the form of the Fokker-Planck equation. The idea behind the expansion is that for constant average values, fluctuations in the species numbers vary proportional to the inverse of the square root of the volume whereas the macroscopic values of the species varies proportional to the inverse of the volume. The LNA provides analytical formulations that are locally valid around stationary points of

the chemical system, which quantify the fluctuations around macroscopic values of the molecular species. It also describes how these fluctuations are correlated with each other at the stationary state and over time. The LNA has the form of a Fokker-Planck equation as in equation (3.7) above, but is in terms of the new variables that are the fluctuating parts $\boldsymbol{\xi}$ of the macroscopic variables. The solution to this equation then predicts that the fluctuations $\boldsymbol{\xi}$ have Gaussian probability distributions around the macroscopic values $\boldsymbol{\phi}$.

Hence, as seen in the Fokker-Planck equation, the time-evolution of a species involves a drift term given by the $\mathcal{A}$ matrix that controls the dynamics of the *averages*, and a fluctuating part $\mathcal{B}\mathcal{B}^T$ that defines the *width* of the distributions. Since both are essentially defined by the stoichiometries $\boldsymbol{\nu}$'s and propensities a's, they both arise from the same source. This is the essence of the fluctuation-dissipation relations. In conclusion, the solution to the Fokker-Planck equation is a probability distribution which is Gaussian with mean given by the deterministic equations and the stationary Covariance matrix given by the solution to the Lyapunov equation (3.6). The stationary covariance matrix and also the time-dependent covariance matrix are derived next.

The contribution of our work lies in the form in which we derive the time-covariance matrix. The functional form of each of the elements of the matrix is derived in a simple sum-of-exponentials form (3.14). As we go along providing results for different regulatory networks, we demonstrate the benefit of such a functional form for the time-covariances in gaining valuable insight into the working of these regulatory networks. This insight in further useful in identifying or recognizing regulatory networks by their structure, regulatory mechanism and also the parameter values that they take.

Continuing with our derivation, to solve for the stationary covariance matrix $\mathbf{C}$, we employ the Lyapunov equation $\mathbf{AC} + \mathbf{CA}^T + \mathbf{BB}^T = 0$, which in general is not possible to solve explicitly, but can be solved easily by a set of transformations as shown in Elf and Ehrenberg (2003) and reproduced below.

Stationary Covariance Matrix: The stationary covariance matrix contains the second-order moments of the stationary Gaussian probability distribution obtained as a solution to the Fokker-Planck equation, and hence is significant as it reveals the correlations between various species of the chemical system. It is given by the Lyapunov Equation,

$$\mathbf{AC} \; + \; \mathbf{CA}^T \; + \; \mathbf{BB}^T \quad = \quad \mathbf{0} \tag{3.8}$$

Let $\mathbf{V}$ be the matrix whose columns are the right eigenvectors of the Jacobian matrix $\mathbf{A}$. $\mathbf{U}^T \equiv \mathbf{V}^{-1}$ is then the matrix whose rows are the left eigenvectors of $\mathbf{A}$. Now, pre

and post-multiplying the above equation by $\mathbf{U}^T$ and $\mathbf{U}$ respectively,

$$\begin{aligned}
\mathbf{U}^T\mathbf{A}\mathbf{C}\mathbf{U} + \mathbf{U}^T\mathbf{C}\mathbf{A}^T\mathbf{U} + \mathbf{U}^T\mathbf{B}\mathbf{B}^T\mathbf{U} &= \mathbf{0} \\
\mathbf{\Lambda}\mathbf{U}^T\mathbf{C}\mathbf{U} + \mathbf{U}^T\mathbf{C}(\mathbf{\Lambda}\mathbf{U}^T)^T + \mathbf{U}^T\mathbf{B}\mathbf{B}^T\mathbf{U} &= \mathbf{0} \\
\mathbf{\Lambda}\mathbf{U}^T\mathbf{C}\mathbf{U} + \mathbf{U}^T\mathbf{C}\mathbf{U}\mathbf{\Lambda}^T + \mathbf{U}^T\mathbf{B}\mathbf{B}^T\mathbf{U} &= \mathbf{0}
\end{aligned}$$

From the transformation of variables, $\mathbf{Z} = \mathbf{U}^T\mathbf{X}$ which is associated with the linearly transformed stoichiometric matrix $\underline{\boldsymbol{\nu}} = \mathbf{U}^T\boldsymbol{\nu}$ and also by defining $\underline{\mathbf{C}} = \mathbf{U}^T\mathbf{C}\mathbf{U}$ and $\underline{\mathbf{B}} = \mathbf{U}^T\mathbf{B}$, Equation (3.8) is transformed as:

$$\begin{aligned}
\mathbf{\Lambda}\underline{\mathbf{C}} + \underline{\mathbf{C}}\mathbf{\Lambda}^T + \underline{\mathbf{B}}\underline{\mathbf{B}}^T &= \mathbf{0} \\
(\underline{\mathbf{C}})_{ij} &= \frac{-(\underline{\mathbf{B}}\underline{\mathbf{B}}^T)_{ij}}{\mathbf{\Lambda}_{ii} + \mathbf{\Lambda}_{jj}} = \frac{-(\underline{\mathbf{B}}\underline{\mathbf{B}}^T)_{ij}}{\lambda_i + \lambda_j}
\end{aligned} \tag{3.9}$$

The original stationary covariance matrix $\mathbf{C}$ is backtransformed as $\mathbf{V}\underline{\mathbf{C}}\mathbf{V}^T$. Here, we should note that analytical solutions are possible. In fact, at this point one can do a detailed analysis and could hope to derive analytical relations between the terms of the covariance matrix and the structure of the network. This is a matter for further investigation. However, in all the gene regulatory networks that we consider here, we are able to evaluate C_{ij} analytically. In the next section, we specifically show this for the case of a single-gene system.

For deriving the time-covariances of fluctuations, firstly we note that the deviations $\delta\mathbf{x}(t) := \mathbf{x}(t) - \langle\mathbf{x}\rangle^s$ follow deterministic dynamics, where $\langle\mathbf{x}\rangle^s$ is the stationary solution that satisfies $\dot{\mathbf{x}} = 0$ in Equation (3.5). This is seen from linearization of Equation (3.5) around the steady state solutions $\dot{x_i} = 0$ resulting in

$$\left(\frac{d}{dt}\right)\delta\mathbf{x} = \mathbf{A}\delta\mathbf{x}, \tag{3.10}$$

where $\mathbf{A}$ is the Jacobian defining the deterministic dynamics. Now, since the means in the kinetic rates R_j are time-independent, and hence the elements of the $\mathbf{A}$ and $\mathbf{B}$ matrices are also time-independent that solution of the deterministic equation (3.5) is no longer of the form $\langle\mathbf{x}(t)\rangle = e^{t\mathbf{A}}\mathbf{x}(t_0)$, but has the general form $\langle\mathbf{x}(t)\rangle = \mathbf{Y}(t)\mathbf{x}(t_0)$, where $\mathbf{Y}(t)$ is the propagator matrix which is the solution of

$$\frac{d\mathbf{Y}(t)}{dt} = \mathbf{A}(t)\mathbf{Y}(t) \tag{3.11}$$

with $\mathbf{Y}(0) = 1$. Since the deviations $\delta\mathbf{x}$ are also controlled by $\mathbf{A}$ as shown in equation (3.10), it is clear that the propagator matrix governs $\delta\mathbf{x}$ and determines the time-dependence of covariances of fluctuations from the stationary covariances $\mathbf{C}$

$$\langle\ \delta\mathbf{x}(t+\tau)\ \delta\mathbf{x}(t)^T\ \rangle = \mathbf{Y}(\tau).\mathbf{C}. \tag{3.12}$$

Below we reduce this propagator matrix to the form $\mathbf{Y}(t) = \mathbf{V}e^{\mathbf{\Lambda}t}\mathbf{U}^T$, where $\mathbf{V}$ and $\mathbf{U}^T$ are the matrices comprising of the right and left eigenvectors of the Jacobian matrix $\mathbf{A}$ respectively and $\mathbf{\Lambda}$ is the diagonal matrix of the eigenvalues $(\lambda_1, \cdots, \lambda_N)$. They are related as $\mathbf{A} \,.\, \mathbf{V} = \mathbf{V} \,.\, \mathbf{\Lambda}$ and $\mathbf{U}^T \,.\, \mathbf{A} = \mathbf{\Lambda} \,.\, \mathbf{U}^T$.

Propagator Matrix: Suppose we have a system with N variables. The rate equations are:

$$\frac{d\mathbf{x}(t)}{dt} \;=\; \mathbf{A}\mathbf{x}(t)$$

resulting in a set of N-coupled differential equations, as $\mathbf{A}$ is a non-diagonal matrix (and also generally unsymmetric). Therefore we first need to uncouple these equations in the variables and then transform them back in terms of the old variables $\mathbf{x} = [x_1, x_2, ..., x_N]$. As the eigenvectors $\mathbf{V}_i$'s are independent, the vector $\mathbf{x}$ can be written as the sum of these N eigenvectors as

$$\mathbf{x}(t) \;=\; \sum_i z_i(t)\mathbf{V}_i \;=\; \mathbf{V}\mathbf{z}(t)$$

Since $\mathbf{U}^T = \mathbf{V}^{-1}$ the above is nothing but a transformation $\mathbf{z}(t) = \mathbf{U}^T.\mathbf{x}(t)$ to a new set of variables z_i's. Therefore the transformed variables are,

$$z_i(t) \;=\; \mathbf{U}_i^T\mathbf{x}(t) \;=\; \sum_j z_j(t)\mathbf{U}_i^T\mathbf{V}_j \;=\; \sum_j z_j(t)\delta_{ij}$$

which is due to the bi-orthogonality of the eigenvectors. Applying the transformation to the rate equations,

$$\begin{aligned}
\frac{d\mathbf{x}(t)}{dt} \;&=\; \mathbf{A}\mathbf{x}(t) \;=\; \mathbf{A}\sum_i z_i(t)\mathbf{V}_i \;=\; \sum_i z_i(t)\lambda_i\mathbf{V}_i \qquad (\because \mathbf{A}\mathbf{V}_i = \lambda_i\mathbf{V}_i) \\
\Rightarrow \quad \frac{d\sum_i z_i(t)\mathbf{V}_i}{dt} \;&=\; \sum_i z_i(t)\lambda_i\mathbf{V}_i \\
\Rightarrow \quad \sum_i \mathbf{V}_i\frac{dz_i(t)}{dt} \;&=\; \sum_i z_i(t)\lambda_i\mathbf{V}_i \\
\Rightarrow \quad \frac{dz_i(t)}{dt} \;&=\; z_i(t)\lambda_i
\end{aligned}$$

We now have a set of uncoupled equations in the new variables z_i's. These can now be solved easily to obtain,

$$z_i(t) \;=\; e^{\lambda_i t}z_i(0)$$

Writing the solution in the original variables $\mathbf{x}(t)$,

$$\mathbf{x}(t) \;=\; \sum_i z_i(t)\mathbf{V}_i \;=\; \sum_i e^{\lambda_i t}z_i(0)\mathbf{V}_i \;=\; \sum_i e^{\lambda_i t}\mathbf{U}_i^T\mathbf{x}(0)\mathbf{V}_i.$$

This means that the time evolution of each species in the system is a sum of a set of exponentials. Writing the above equation in matrix form,

$$\mathbf{x}(t) \quad = \quad \mathbf{Y}(t) \; . \; \mathbf{x}(0)$$

where $\mathbf{Y}(t) = \mathbf{V}e^{\Lambda t}\mathbf{U}^T$ is the propagator/evolution matrix. Each element of the propagator matrix is therefore,

$$Y_{ij}(t) \quad = \quad \sum_{k=1}^{N} e^{\lambda_k t} U_{kj}^T V_{ik} \tag{3.13}$$

Coming back to the equation for the time-covariance matrix $\langle\, \delta\mathbf{x}(t+\tau)\; \delta\mathbf{x}(t)^T \,\rangle = \mathbf{Y}(\tau).\mathbf{C}$ and re-writing it element-wise, we obtain the time-covariance functions between two molecular species x_i and x_j as:

$$\begin{aligned} \langle\, \delta x_i(t+\tau)\; \delta x_j(t) \,\rangle \quad &= \quad \sum_k Y_{ik}(\tau) \; . \; C_{kj} \\ &= \quad \sum_k \Big(\sum_l e^{\lambda_l \tau} U_{lk}^T V_{il} \Big) C_{kj} \\ &= \quad \sum_{l=1}^{N} e^{\lambda_l \tau} V_{il} \sum_{k=1}^{N} U_{lk}^T C_{kj} \end{aligned} \tag{3.14}$$

For effective comparison of results from different GRNs, we use dynamic correlations by normalizing Equation (3.14). Use of stationary auto-covariances for normalization helps in retaining the dynamic character along the τ-axis, though the magnitudes are rescaled between 0 and 1.

$$\text{Corr} \;\; [\delta x_i(t+\tau), \delta x_j(t)] \quad = \quad \frac{\langle\, \delta x_i(t+\tau)\; \delta x_j(t) \,\rangle}{\sqrt{\langle\, (\delta x_i(t+\tau))^2 \,\rangle \;\; \langle\, (\delta x_j(t))^2 \,\rangle}} \tag{3.15}$$

The form of the Equation (3.14) brings out the existing relation between the covariances and the deterministic characteristics of the system. The system Jacobian $\mathbf{A}$ and diffusion matrix $\mathbf{B}\mathbf{B}^T$ are derived through the stoichiometry $\boldsymbol{\nu}$ and reaction rates $\mathbf{R}$ that are in turn obtained from the deterministic rate equations. Therefore these rate equations that are responsible for the time-evolution of *averages* also influence the internal fluctuations of the system. The analytical framework presented above allows one to visualize the effect of various *attributes* of the system such as the rate constants, regulatory mechanism and the network structure on the fluctuation properties of the molecular species which are the mRNAs and proteins. Firstly, this helps in analyzing the sensitivity of the dynamic correlations *w.r.t* the rate constants thereby providing significant biological insights into the working of a GRN when stochastic effects are included. Secondly, this would be of help in drawing qualitative distinctions between different GRNs, as each

GRN is expected to generate dynamic correlations that are unique to its regulatory mechanism and network structure. In other words, this would make way for the possible identification of the type of regulation present between two genes and also in predicting the corresponding network structure, given the time-series of any two molecular species of a GRN. The dynamic correlations therefore act as *signatures* of the GRNs, and to demonstrate which, we consider the following regulatory and network mechanisms:

I Regulatory Mechanisms

(a) Elementary activation: where the protein of X is the activator or positive regulator of Y.

(b) Elementary repression: where the protein of X is the repressor or negative regulator of Y.

(c) Activation *via* dimerization: where dimers of the protein P_x are the activators of Y.

II Network Mechanisms

(a) Regulation through an Intermediary gene: Gene Y is indirectly activated or repressed by protein of X *via* another gene.

(b) Co-operative regulation: Gene Y is not only regulated by protein of X but also by proteins of other elements.

(c) Coherent and Incoherent FeedForward loops.

3.3 Time-covariance in a single gene

The best way to give an intuition of the time-covariance function is by applying it to the simplest of cases, that of a single gene. In this section we derive the time-covariance function between the mRNA and protein molecules in the case of a single gene. The model includes a single gene which spontaneously switches to the ON or active state G^* at a rate k_{on} and back to the OFF or inactive state G at the rate k_{off}. Therefore the stationary distribution of G^* is of the Binomial type. In the ON state the mRNA M is transcribed out of the gene at a rate of k_M^+ and is further translated to the protein P at a rate k_P^+.

Let $\langle G^* \rangle$ represent the average amount of active genes in concentration form and in units of nano-Molar. Then the rate of production of $\langle G^* \rangle$ follows the deterministic dynamics of the system given by,

$$\frac{d\langle G^* \rangle}{dt} = k_{on}\langle G \rangle - k_{off}\langle G^* \rangle$$

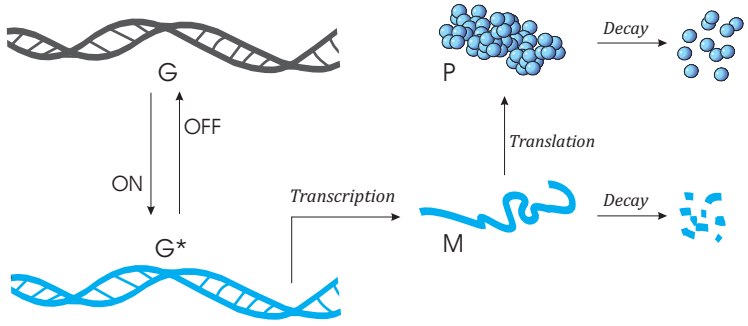

FIGURE 3.1: Schematic representation of the working of a single gene where the gene G switches between active and inactive states. M is transcribed from G^*.

Gene inactive/active	$G \underset{k_{off}}{\overset{k_{on}}{\rightleftharpoons}} G^*$	$k_{off} = 20\ min^{-1}$, $k_{on} = 5\ nM^{-1}min^{-1}$
Transcription	$G^* \xrightarrow{k_M^+} G^* + M$	$k_M^+ = 2\ min^{-1}$
Translation	$M \xrightarrow{k_P^+} M + P$	$k_P^+ = 1\ min^{-1}$
mRNA degradation	$M \xrightarrow{k_M^-} \phi$	$k_M^- = 0.1\ min^{-1}$
Protein degradation	$P \xrightarrow{k_P^-} \phi$	$k_P^- = 0.01\ min^{-1}$

TABLE 3.1: Elementary reactions in the case of a single-gene system where $x \rightarrow \phi$ denotes spontaneous decay of x. The decay rates are $k^- = \ln(2)/(halflife)$, where the half-lives of mRNAs and proteins are chosen to be 7 and 70 minutes respectively. Other rate constants are chosen to be biologically meaningful. nM is nano-Molar.

Due to the conservation of the gene molecules, $\langle G(t)\rangle = G(t_0) - \langle G^*(t)\rangle$, where $G(t_0)$ is the initial amount of the DNA molecule present in the medium. Using this substitution in the rate equation for G^* we have the following rate equations of the three variables describing their deterministic behaviour:

$$\begin{aligned}
\frac{d\langle G^*\rangle}{dt} &= k_{on}(G(t_0) - \langle G^*\rangle) - k_{off}\langle G^*\rangle \\
\frac{d\langle M\rangle}{dt} &= k_M^+\langle G^*\rangle - k_M^-\langle M\rangle \\
\frac{d\langle P\rangle}{dt} &= k_P^+\langle M\rangle - k_P^-\langle P\rangle
\end{aligned}$$

The expression for the mean steady state value of the variables are obtained by equating the above rate equations to zero, upon which we arrive at $\langle G^*\rangle = \frac{G(t_0)k_{on}}{k_{on}+k_{off}} = 0.2\ nM$, $\langle M\rangle = \frac{k_M^+\langle G^*\rangle}{k_M^-} = 4.04\ nM$ and $\langle P\rangle = \frac{k_P^+\langle M\rangle}{k_P^-} = 408\ nM$. Taking the vector of the system variables as $\mathbf{X} = [G^*, M, P]$, the corresponding Jacobian matrix derived from the rate equations is:

$$\mathbf{A} = \begin{pmatrix} (-k_{on} - k_{off}) & 0 & 0 \\ k_M^+ & -k_M^- & 0 \\ 0 & k_P^+ & -k_P^- \end{pmatrix}$$

Further, the vector of the deterministic rates and the stoichiometric matrix are also derived from the same rate equations as:

$$\begin{aligned}\mathbf{R} &= \left[k_{on}(G(t_0) - \langle G^*\rangle), k_{off}\langle G^*\rangle, k_M^+\langle G^*\rangle, k_P^+\langle M\rangle, k_M^-\langle M\rangle, k_P^-\langle P\rangle\right]^T \\ \boldsymbol{\nu} &= \begin{pmatrix} +1 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & +1 & 0 & -1 & 0 \\ 0 & 0 & 0 & +1 & 0 & -1 \end{pmatrix}\end{aligned}$$

The diffusion matrix which is given by $\mathbf{B}\mathbf{B}^T = \boldsymbol{\nu}\ \mathbf{diag}(\mathbf{R})\ \boldsymbol{\nu}^T$ is:

$$\mathbf{B}\mathbf{B}^T = \begin{pmatrix} (k_{on}(\alpha - \langle G^*\rangle) + k_{off}\langle G^*\rangle) & 0 & 0 \\ 0 & (k_M^+\langle G^*\rangle + k_M^-\langle M\rangle) & 0 \\ 0 & 0 & (k_P^+\langle m\rangle + k_P^-\langle P\rangle) \end{pmatrix}$$

Once the drift (Jacobian) and diffusion matrices are obtained, the next step is to solve for the stationary covariance matrix $\mathbf{C}$. This is done by solving the Lyapunov equation (3.8) through a set of transformations resulting in equation (3.9) or by simply solving for the elements in the stationary covariance matrix as we do here. The stationary covariance matrix for the three variable case can be written as:

$$\begin{aligned}\mathbf{C} &= \begin{pmatrix} C_{11} & C_{12} & C_{13} \\ C_{21} & C_{22} & C_{23} \\ C_{31} & C_{32} & C_{33} \end{pmatrix} \\ &= \begin{pmatrix} \mathrm{Var}[G^*(t), G^*(t)] & \mathrm{Cov}[G^*(t), M(t)] & \mathrm{Cov}[G^*(t), P(t)] \\ \mathrm{Cov}[M(t), G^*(t)] & \mathrm{Var}[M(t), M(t)] & \mathrm{Cov}[M(t), P(t)] \\ \mathrm{Cov}[P(t), G^*(t)] & \mathrm{Cov}[P(t), M(t)] & \mathrm{Var}[P(t), P(t)] \end{pmatrix}\end{aligned} \tag{3.16}$$

where Var and Cov represent the stationary auto-covariance and covariance terms between the variables of the system. It is important to note that the stationary covariance matrix is symmetric and due to its stationary nature is applicable at any time t during steady state conditions. Under steady state conditions, each element of the above matrix are equated to zero and solved systematically to obtain the stationary covariances between two variables. For example, the first element of $\mathbf{A}\mathbf{C} + \mathbf{C}\mathbf{A}^T + \mathbf{B}\mathbf{B}^T = \mathbf{0}$ is:

$$k_{on}\langle G^*\rangle + 2(-k_{on} - k_{off})C_{11} + k_{off}(\alpha - G^*) = 0$$

which on solving gives the auto-variance of G^*,

$$\mathrm{Var}[G^*(t), G^*(t)] = \frac{(k_{on} - k_{off})\langle G^*\rangle + G(t_0)k_{off}}{2(k_{on} + k_{off})} = \langle G^*\rangle\left[\frac{k_{off}}{k_{on} + k_{off}}\right] \tag{3.17}$$

Similarly, progressively solving for all the elements of the **C** matrix, we get the covariances between all the variables. **In these expressions of the covariances, we employ approximations to obtain simpler expressions. The approximations are that the decay rates are much smaller when compared to the ON and OFF rates of the gene $(\mathbf{k_{on}} + \mathbf{k_{off}}) \gg \mathbf{k^-_{(M,P)}}$ and that the OFF rate is much larger than the ON rate $\mathbf{k_{off}} > \mathbf{k_{ON}}$.** Both these approximations are valid for biologically plausible values of rate constants Bundschuh et al. (2003).

$$\begin{aligned}
\mathrm{Cov}[G^*(t), M(t)] &= \langle M\rangle \frac{k_M^- k_{off}}{(k_{on}+k_{off}+k_M^-)(k_{on}+k_{off})} \\
&\approx \langle M\rangle \frac{k_M^- k_{off}}{(k_{on}+k_{off})^2} \\
&\approx \langle M\rangle \frac{k_M^-}{k_{off}} \qquad (3.18)\\
\mathrm{Var}[M(t), M(t)] &= \langle M\rangle + \langle M\rangle \left[\frac{k_M^+ k_{off}}{(k_{on}+k_{off}+k_M^-)(k_{on}+k_{off})}\right] \\
&\approx \langle M\rangle + \langle M\rangle \left[\frac{k_M^+}{k_{off}}\right] \\
&\approx \langle M\rangle \qquad (3.19)\\
\mathrm{Cov}[M(t), P(t)] &= \langle M\rangle \frac{k_P^+}{k_M^- + k_P^-}\left[1 + \frac{k_M^+ k_{off}}{(k_{on}+k_{off})(k_{on}+k_{off}+k_M^-)}\left(1 + \frac{k_M^-}{(k_{on}+k_{off}+k_P^-)}\right)\right] \\
&\approx \langle M\rangle \frac{k_P^+}{k_M^- + k_P^-}\left[1 + \frac{k_M^+ k_{off}}{(k_{on}+k_{off})^2}\left(1 + \frac{k_M^-}{(k_{on}+k_{off})}\right)\right] \\
&\approx \langle M\rangle \frac{k_P^+}{k_M^- + k_P^-}\left[1 + \frac{k_M^+}{k_{off}}\left(1 + \frac{k_M^-}{k_{off}}\right)\right] \\
&\approx \langle M\rangle \frac{k_P^+}{k_M^- + k_P^-}\left[1 + \frac{k_M^+}{k_{off}}\right] \\
&\approx \langle M\rangle \frac{k_P^+}{k_M^- + k_P^-} \qquad (3.20)\\
\mathrm{Var}[P(t), P(t)] &= \langle P\rangle + \frac{k_P^+}{k_P^-}\mathrm{Cov}[M(t), P(t)] \\
&\approx \langle P\rangle + \langle P\rangle \frac{k_P^+}{k_M^- + k_P^-} \qquad (3.21)
\end{aligned}$$

As mentioned in the previous chapter it has been observed in various real biological systems that the noise in the mRNA distribution which is nothing but its variance over mean squared has sometimes the characteristics of a Poissonian process while in other cases is far from it. A Poisson process being reflected in the fact that the distribution resulting out of such a process has its variance equal to its mean. Therefore in the case where the gene is constantly in the ON state, the transcription process results in mRNA molecules that are Poissonian distributed, while in the above described gene ON/OFF model the distribution of mRNAs and proteins are skewed and are non-Poissonian in character. However, with the approximation of $(k_{on} + k_{off}) \gg k^-_{(M,P)}$ and $k_{off} > k_{on}$

the expressions for the stationary covariance terms are exactly as would be obtained in the model where the gene is constantly in the ON state and the mRNAs are being produced by random birth-death events leading to a Poissonian distribution. Previous works Thattai and van Oudenaarden (2001); Paulsson (2005) have involved investigating the stationary auto-covariance which is, as seen above, a function of the system parameters. However, as our aim is to include the additional factor of time into the analyses, the dynamic covariance between the mRNA and the protein variables are employed as representative of the system's internal fluctuations. In order to determine the time-covariance between the mRNA and protein $\mathrm{Cov}[M(t), P(t+\tau)]$ we need the eigenvectors of the system. The right eigenvectors of the Jacobian are the column vectors of the following matrix:

$$\mathbf{V} = \begin{pmatrix} 0 & \frac{(k_{on}+k_{off}-k_M^-)(k_{on}+k_{off}-k_P^-)}{k_M^+ k_P^+} & 0 \\ \frac{-(k_M^- - k_P^-)}{k_P^+} & \frac{-(k_{on}+k_{off}-k_P^-)}{k_P^+} & 0 \\ 1 & 1 & 1 \end{pmatrix}$$

and by the bi-orthogonal property of the right and left eigenvectors $\mathbf{U}^T\mathbf{V} = \mathbf{I}$, we get the left eigenvectors as:

$$\mathbf{U}^T = \begin{pmatrix} \frac{-k_M^+ k_P^+}{(k_{on}+k_{off}-k_M^-)(k_M^- - k_P^-)} & \frac{-k_P^+}{(k_M^- - k_P^-)} & 0 \\ \frac{k_M^+ k_P^+}{(k_{on}+k_{off}-k_M^-)(k_{on}+k_{off}-k_P^-)} & 0 & 0 \\ \frac{k_M^+ k_P^+}{(k_{on}+k_{off}-k_P^-)(k_M^- - k_P^-)} & \frac{k_P^+}{(k_M^- - k_P^-)} & 1 \end{pmatrix}$$

The corresponding eigenvalues of the $\mathbf{A}$ matrix are:

$$\boldsymbol{\lambda} = [-k_M^-, \ -(k_{on}+k_{off}), \ -k_P^-]$$

Therefore the time-covariance between the mRNa and protein species is:

$$\begin{aligned} \mathrm{Cov}[M(t), P(t+\tau)] &= \mathrm{Cov}[X_2(t), X_3(t+\tau)] \\ &= \sum_{l=1}^{3} e^{\lambda_l \tau} V_{3l} \left(\sum_{k=1}^{3} U_{lk}^T C_{k2} \right) \\ &= e^{-k_P^- \tau} \left[\frac{k_M^+ k_P^+}{(k_{on}+k_{off}-k_M^-)(k_M^- - k_P^-)} C_{12} + \frac{k_P^+}{(k_M^- - k_P^-)} C_{22} \right] \\ &\quad - \ e^{-k_M^- \tau} \left[\frac{k_M^+ k_P^+}{(k_{on}+k_{off}-k_M^-)(k_M^- - k_P^-)} C_{12} + \frac{k_P^+}{(k_M^- - k_P^-)} C_{22} + C_{32} \right] \\ &\quad + \ e^{-(k_{on}+k_{off})\tau} \left[\frac{k_M^+ k_P^+}{(k_{on}+k_{off}-k_M^-)(k_{on}+k_{off}-k_P^-)} C_{12} \right] \end{aligned} \quad (3.22)$$

Since $(k_{on} + k_{off})$ is assumed to be much larger than the decay rates, we neglect the fast decaying exponential $e^{-(k_{on}+k_{off})\tau}$ and its allied term leading to,

$$\begin{aligned}\mathrm{Cov}[M(t), P(t+\tau)] \;\approx\; & e^{-k_P^-\tau}\left[\frac{k_M^+ k_P^+}{(k_{on}+k_{off}-k_M^-)(k_M^- - k_P^-)}C_{12} + \frac{k_P^+}{(k_M^- - k_P^-)}C_{22}\right] \\ & - \; e^{-k_M^-\tau}\left[\frac{k_M^+ k_P^+}{(k_{on}+k_{off}-k_M^-)(k_M^- - k_P^-)}C_{12} + \frac{k_P^+}{(k_M^- - k_P^-)}C_{22} + C_{32}\right]\end{aligned} \tag{3.23}$$

On substituting for the $\mathbf{C}$ terms and doing approximations $(k_{on} + k_{off} \gg k^-_{(M,P)}, k_{off} > k_{on})$, we get

$$\mathrm{Cov}[M(t), P(t+\tau)] \;\approx\; \langle M\rangle \frac{k_P^+}{(k_M^- - k_P^-)}\left[e^{-k_P^-\tau}\left(\frac{2k_M^-}{k_M^- + k_P^-}\right) - e^{-k_M^-\tau}\right] \tag{3.24}$$

This function is a sum or rather difference of two exponentials which in all possibility could be a non-monotonic function. To find if this is truly the case, we differentiate the above covariance term partially *w.r.t* τ and equate to 0, yielding the τ at which the covariance reaches peak magnitude. We denote the corresponding τ as τ^* and is obtained on solving the following equation:

$$\begin{aligned}\frac{\partial}{\partial t}\mathrm{Cov}[M(t), P(t+\tau)] &= 0 \\ e^{-k_M^-\tau^*} - e^{-k_P^-\tau^*}\left(\frac{2k_p^-}{k_M^- + k_P^-}\right) &= 0 \\ \tau^* &= \frac{\ln\left[\frac{k_M^- + k_P^-}{2k_p^-}\right]}{k_M^- - k_P^-}\end{aligned} \tag{3.25}$$

and is equal to 19.1 minutes for the decay rates given in Table 3.1. The plot of the dynamic covariance is shown to the left in Figure 3.2. The shape of the covariance function signifies the fact that there is an inherent delay in the response of the downstream variable to perturbations in the upstream one. Talking in terms of the systems' internal fluctuations, the shape of the dynamic covariance simply means that any fluctuation in the mean value of the upstream variable at time (t), which in this case is the mRNA, is highly correlated to the fluctuations in the downstream variables that is the protein at time $(t + \tau)$. This time delay τ is shown to be a function of the decay rates and therefore the temporal character of the covariances can be controlled via these parameters. The interesting aspect of this time-delay is that it is related to the time delay in the response of the mean value of the regulated variable to a perturbation in the mean value of the regulator. This relation between the internal fluctuations and the deterministic response is what the FDT is about. Instead of going into cumbersome details of the relation between the fluctuations and the dissipation parts of the system, we shall simply show here that they are both in essence interlinked via common factors. For this, we focus on the temporal aspect of the deterministic response. Suppose the

system is at steady-state equillibrium conditions before say time $t = 0$, and the mean value of the mRNA is denoted by $\langle M(t_{ss})\rangle$, where t_{ss} is the steady state time-period before $t = 0$. Now, let us introduce a perturbation at time $t = 0$ in this mean value by an amount $\triangle M$ and then allow $M(t)$ to fall back freely to its initial condition which is the mean steady-state value $\langle M(t_{ss})\rangle$. This induces a deterministic response in $\langle P(t)\rangle$ which follows the deterministic rate equation but now, with the initial concentration of M being equal to $\langle M(t_{ss})\rangle + \triangle M$. On solving for the rate equation for the protein, we get:

$$\langle P(t)\rangle = \langle P(t_{ss})\rangle + \frac{k_P^+(\Delta M - M(t_{ss}))}{k_M^- - k_P^-}\left[e^{-k_P^- t} - e^{-k_M^- t}\right] \tag{3.26}$$

The plots of the mean steady-state mRNA and protein levels are given on the right side of the Figure 3.2. The response in the protein level also has the characteristic peak after a time-delay of (t_{resp}), which in this case is 26 minutes. The expression for t_{resp} is given by equating the time-evolution of $\langle P\rangle$ to zero,

$$\begin{aligned} \frac{\partial\langle P(t)\rangle}{\partial t} &= 0 \\ k_M^- e^{-k_M^- t_{resp}} - k_P^- e^{-k_P^- t_{resp}} &= 0 \\ t_{resp} &= \frac{\ln\left[\frac{k_M^-}{k_P^-}\right]}{k_M^- - k_P^-} \end{aligned} \tag{3.27}$$

This is smallest for the case where $k_M^- = k_P^-$ and is evaluated using l'Hopital's rule. Letting $k_M^-/k_P^- = x$, the expression for t_{resp} is now,

$$\begin{aligned} \lim_{x\to 1} t_{resp} &= \lim_{x\to 1} \frac{1}{k_P^-}\frac{\ln(x)}{(x-1)} \\ &= \lim_{x\to 1} \frac{1}{k_P^-}\frac{1/x}{1} \\ &= \frac{1}{k_P^-} \end{aligned}$$

In the present case, where $k_M^- \neq k_P^-$, t_{resp} is given by Equation (3.27) and is equal to 25.6 minutes for the decay rates given in Table 3.1. This delay which is greater than τ^* by about 7 minutes is self-explainatory. This is because $\langle M\rangle$ after being perturbed externally was allowed to go back to its original state by decaying freely. This means that at every infinitesimal time instant after the perturbation M is virtually still at a perturbed value. This is almost equivalent to many perturbations in M which sustains the response in P for a longer duration. Intuitively the duration for which the decay in M has a perturbation type of effect on P should be around the value of its half-life. Evaluating the difference between t_{resp} and τ^*, we find that this indeed is the case. This difference in the time-delays would be close to zero had the perturbation in M lasted

only for a very short time-interval after which it is forced back to its original value.

$$\begin{aligned} t_{resp} - \tau^* &= \frac{\ln\left[\frac{k_M^-}{k_P^-}\right]}{k_M^- - k_P^-} - \frac{\ln\left[\frac{k_M^- + k_P^-}{2k_P^-}\right]}{k_M^- - k_P^-} \\ &= \frac{\ln\left[\frac{2k_M^-}{k_M^- + k_P^-}\right]}{k_M^- - k_P^-} \end{aligned} \tag{3.28}$$

For large half-life of the protein in comparison to that of the mRNA molecule *i.e.,* for $(k_P^- \ll k_M^-)$ the time difference $t_{resp} - \tau^* \approx \frac{\ln(2)}{k_M^-} = \tau_{1/2_m}$, which in the present case is equal to 7 minutes, which is what is observed in Figure 3.2.

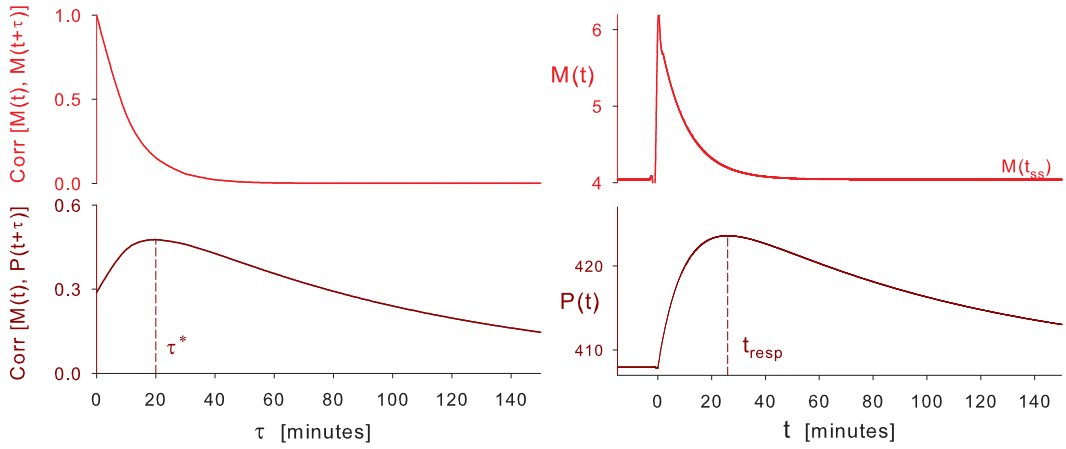

FIGURE 3.2: On the left is the time-correlation function between M and P. The time at which the correlation attains maximum value is $\tau^* = 19$ minutes. To the right we show the deterministic response in P for a perturbation in the steady state value of M from a value of $\langle M(t_{ss})\rangle = 4.04\ nM$ to a value of $\langle M(t_{ss})\rangle + \Delta M = 6.04\ nM$. The time at which the response in $P(t)$ attains its maximum value is $t_{resp} = 26$ minutes.

3.4 Summary

In conclusion, the time-covariance of the deviations $\delta\mathbf{x}$ are fully described by the Jacobian matrix whose elements are in turn functions of the average values $(\langle x_1\rangle, \cdots, \langle x_N\rangle)$ of the molecular species. Simply put the molecular fluctuations of a stochastic system can be estimated by the knowledge of the deterministic dynamics of the system. Since the deterministic behaviour of the system can be observed in the form of a response to a perturbation, in effect we have shown here that the fluctuation dissipation relation holds true in such systems. This relation says that the internal fluctuations and the deterministic response to an external perturbation arise from the same source which is described by the deterministic dynamics of the system.

Chapter 4

Sensitivity Analysis

In the previous chapter, for the case of a single gene, we derived the expression for τ^* which was the time required for the covariance between the mRNA and the protein species to reach the maximum value of Cov*. This was given in equation (3.25). This time-delay τ^* is the same amount of time taken by the dynamic correlations as well to reach a corresponding peak value of Corr*, since the dynamic correlations are simply the normalized form of the dynamic covariances and hence retain their temporal character. Coming back to the case of the single gene, where the expression for the τ^* corresponding to the dynamic correlations between the mRNA and the proteins, given by equation (3.25) involves the decay rates of these species. It is therefore obvious that this τ^* is sensitive to changes in these parameters. The question now is by how much it is sensitive. This question is important for a couple of reasons at least. One use of these sensitivities would be in inferring the values of these parameters through the experimentally observed correlation plots. However, whether one is able to easily do such an inference or not, the sensitivities basically improve our understanding about the biological processes that are responsible for the expression of τ^* that we obtained earlier. Therefore, in the case where the set of biochemical reactions are known beforehand, the sensitivity analyses would turn the focus of biologists onto a selected few reactions whose rate constants greatly affect the fluctuation properties represented by features such as τ^*. This would then provide a means of controlling the internal fluctuations of the reacting species by altering the most sensitive rate constants. Therefore it is important to evaluate the sensitivities of not only of τ^* but also of the term Corr* *w.r.t* each of the reaction rate constants of the system. This is made particularly easy by the functional form of the covariance that we derived earlier in equation (3.14).

Therefore, our goal here is to monitor changes in the dynamic correlation functions for changes in the values of the rate constants. Instead of monitoring the entire correlation plot, we focus on its defining *features* such as Corr*, which is the peak value and the time taken to achieve this peak which is τ^*. Therefore the *feature* vector that *defines*

the major features of the covariance function is $[\text{Corr}^*, \tau^*]$. If p_α are the rate constants or rather the parameters of the system, then the sensitivites that we are interested in are defined as $\frac{\partial\, \text{Corr}^*}{\partial p_\alpha}$ and $\frac{\partial \tau^*}{\partial p_\alpha}$. Since Corr^* is nothing but the normalized form of Cov^*, we first derive $\frac{\partial\, \text{Cov}^*}{\partial p_\alpha}$. Now, for the sake of convenience, let us denote the covariance $\langle \delta x_i(t+\tau)\ \delta x_j(t)^T \rangle$ between the species x_i and x_j, as $\text{Cov}(\tau, \mathbf{p})_{ij}$. This is because the covariances are basically functions in the variable τ and $\mathbf{p}$, where $\mathbf{p}$ is the vector whose M elements are the parameters p_α. Hence,

$$\text{Cov}(\tau, \mathbf{p})_{ij} \quad = \quad \sum_{l=1}^{N} e^{\lambda_l \tau} V_{il} \Big(\sum_{k=1}^{N} U_{lk}^T C_{kj} \Big) \tag{4.1}$$

where, N is the number of reacting species, λ_l is the l^{th} eigenvalue, V_{il} is the i^{th} element of the l^{th} right eigenvector, U_{lk}^T is the k^{th} element of the l^{th} left eigenvector and finally, C_{kj} is the $(k,j)^{th}$ element of the stationary covariance matrix $\mathbf{C}$. The elements of this stationary covariance matrix are derived through the Lyapunov equation (3.8). The Jacobian matrix and the stationary covariance matrix being made up of the parameters p_α, it is obvious that all the elements on the RHS of the above expression are functions of the system parameters. Therefore any change in the value of the parameters, *i.e.,* the reaction rate constants has an obvious effect on the dynamic covariances. Therefore from the above expression it is clear that the dynamic covariance is basically a function in $M+1$ variables, one being:

- τ - the time difference between the two variables δx_i and δx_j at which we measure their covariance, while the other M are the elements of,
- $\mathbf{p}$ - the M-dimensional vector of the system parameters p_α, which in the case of a biochemical reacting system are the rate constants.

A simple illustration of the use of sensitivities is given below. Suppose one is interested in quantifying the change in the correlations between mRNAs and proteins in the case of the single-gene system, for small changes in the protein decay rate k_P^-. The plots of the correlations for the two values of k_P^- are given in Figure 4.1.

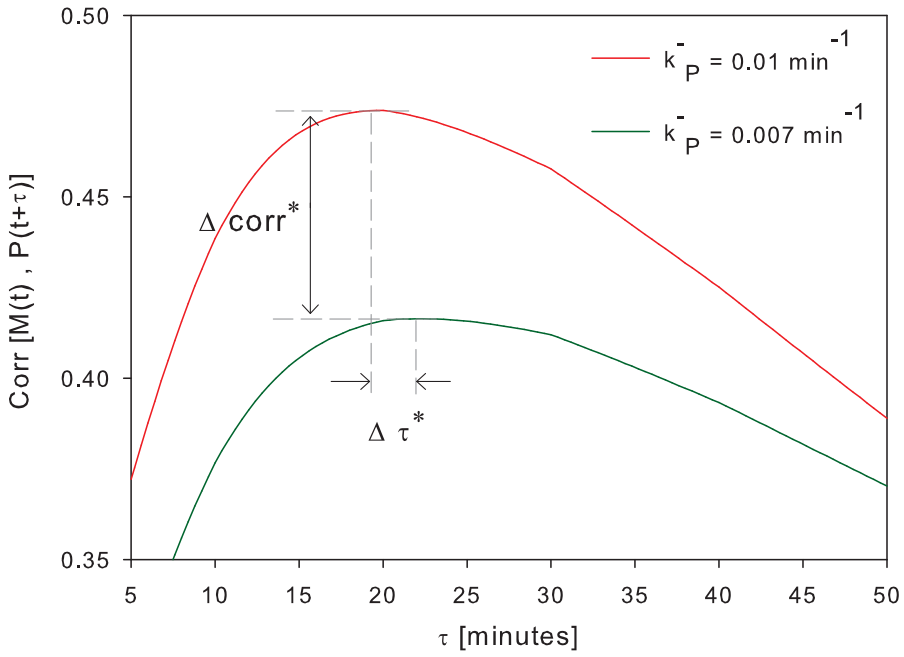

FIGURE 4.1: The plot of the dynamic correlation function $\mathrm{Corr}[M(t), P(t+\tau)]$ in the case of the single-gene system, for two different values of the protein decay rate k_P^-. k_P^- is reduced from its original value of 0.01 min^{-1} to 0.007 min^{-1} for which there is not only a decrease in the correlation peak Corr^* but also the time-delay of this peak τ^* increases.

As k_P^- is decreased from 0.01 min^{-1} to 0.007 min^{-1}, the dynamic correlation decreases which can be seen as a change in its *features* Corr^* and τ^*. Denoting this change by $\triangle\ \mathrm{Corr}^*$ and $\triangle\ \tau^*$, they can be evaluated via their sensitivites at the original value of k_P^- as follows:

$$\triangle\ \overset{*}{\mathrm{Corr}} \simeq \left.\frac{\partial\,\mathrm{Corr}^*}{\partial k_P^-}\right|_{k_P^-=0.01} \times \triangle\ k_P^- \tag{4.2}$$

$$\triangle\ \tau^* \simeq \left.\frac{\partial \tau^*}{\partial k_P^-}\right|_{k_P^-=0.01} \times \triangle\ k_P^- \tag{4.3}$$

which would hold true for small changes in k_P^-. For large $\triangle\ k_P^-$ these equations might not hold. This is because the derivative might be highly sensitive in the rate constants implying that it needs to be re-evaluated for small changes in them and is correspondingly termed as *non-linearly* sensitive in these parameters, *i.e.,* the sensitivities are classified as being *non-linear* in character if the parameters (p_α) induce different sensitivities for larger changes in their values, implying that $\frac{\partial\,\mathrm{Corr}^*}{\partial k_P^-}$ and $\frac{\partial \tau^*}{\partial k_P^-}$ have to be re-evaluated if the change in the value of those parameters is significantly large, say by more than $\pm 5\%$ of their original value. However, if the expressions in equation (4.2) hold true approximately, then the sensitivities are termed as *linear* in those parameters. In the next section we derive the term $\frac{\partial\,\mathrm{Cov}^*}{\partial k_P^-}$, which is in fact applicable at any Cov, *i.e.,* for

the covariance at any time instant τ and not just at $\tau = \tau^*$.

4.1 Sensitivity of Covariance Amplitude *w.r.t* the system parameters

Differentiating the covariance term of equation (3.14) *w.r.t.* the parameters p_α partially, and noting that,

$$\frac{\partial e^{\lambda_l \tau}}{\partial p_\alpha} = \frac{\partial e^{\lambda_l \tau}}{\partial \lambda_l}\frac{\partial \lambda_l}{\partial p_\alpha} = \tau e^{\lambda_l \tau}\frac{\partial \lambda_l}{\partial p_\alpha}$$

we have,

$$\begin{aligned}\frac{\partial \operatorname{Cov}(\tau, \mathbf{p})_{ij}}{\partial p_\alpha} &= \sum_{l=1}^{N} \tau e^{\lambda_l \tau}\frac{\partial \lambda_l}{\partial p_\alpha} V_{il} \sum_{k=1}^{N} U_{kl} C_{kj} + \sum_{l=1}^{N} e^{\lambda_l \tau}\frac{\partial V_{il}}{\partial p_\alpha} \sum_{k=1}^{N} U_{kl} C_{kj} \\ &+ \sum_{l=1}^{N} e^{\lambda_l \tau} V_{il} \sum_{k=1}^{N} \frac{\partial U_{kl}}{\partial p_\alpha} C_{kj} + \sum_{l=1}^{N} e^{\lambda_l \tau} V_{il} \sum_{k=1}^{N} U_{kl}\frac{\partial C_{kj}}{\partial p_\alpha}\end{aligned} \tag{4.4}$$

Firstly we shall obtain the derivatives of the eigenvalues and eigenvectors *w.r.t* the parameters p_α as follows.

4.1.1 Derivatives of Eigenvalues and Eigenvectors

Eigenvalues and eigenvectors are fundamental to the dynamic behaviour of the system. Since these quantities are functions of the parameters of the system, any variation in these parameters lead to changes in the dynamic response of the system. The response in our case is measured by the time-covariance functions. Hence, we expect any change in the values of the parameters to result in changes in the time-covariance functions. Since the time-covariance function is of the sum-of-exponentials form and involves the eigenvalues and eigenvectors, we deduce the following derivatives to estimate numerically the change in the time-covariance function *w.r.t* changes in the parameter values.

Let $\mathbf{A}$ be a N^{th} order Jacobian matrix of the system. The matrix eigenproblem can be written down as:

$$\mathbf{A}\ \mathbf{V}_i = \lambda_i\ \mathbf{V}_i \tag{4.5}$$

All eigenvalues are assumed to be distinct. The corresponding adjoint problem is:

$$\mathbf{U}_i^T\ \mathbf{A} = \lambda_i\ \mathbf{U}_i^T \tag{4.6}$$

The matrix $\mathbf{A}$ and hence the eigenvectors $\mathbf{V}_i$, $\mathbf{U}_i^T$ and the eigenvalues λ_i are functions of the parameter vector $\mathbf{p}$ of the system, whose elements are denoted as p_α. Hence our objective now is to obtain the derivatives of the eigenvectors and eigenvalues *w.r.t.* these parameters. Differentiating Equation (4.5) *w.r.t* p_α,

$$\frac{\partial \mathbf{A}}{\partial p_\alpha}\mathbf{V}_i + \mathbf{A}\frac{\partial \mathbf{V}_i}{\partial p_\alpha} = \frac{\partial \lambda_i}{\partial p_\alpha}\mathbf{V}_i + \lambda_i \frac{\partial \mathbf{V}_i}{\partial p_\alpha} \tag{4.7}$$

Pre-multiplying throughout by $\mathbf{U}_i^T$ and using Equation (4.6), we arrive at,

$$\frac{\partial \lambda_i}{\partial p_\alpha} = \frac{\mathbf{U}_i^T \frac{\partial \mathbf{A}}{\partial p_\alpha}\mathbf{V}_i}{\mathbf{U}_i^T\mathbf{V}_i} \tag{4.8}$$

Since the eigenvalues are assumed to be distinct, the set of eigenvectors forms a basis for the N-space and the first derivatives of eigenvectors can be expressed in terms of the eigenvectors as,

$$\frac{\partial \mathbf{V}_i}{\partial p_\alpha} = \sum_{j=1}^{N} g_{ij_\alpha}\mathbf{V}_j \quad \text{and} \quad \frac{\partial \mathbf{U}_i}{\partial p_\alpha} = \sum_{j=1}^{N} h_{ij_\alpha}\mathbf{U}_j \tag{4.9}$$

Now, the calculation of the first derivatives of the eigenvectors reduces to the evaluation of the co-efficients g_{ij_α} and h_{ij_α}. Pre-multiplying Equation (4.7) by $\mathbf{U}_j^T$ where $j \neq i$,

$$\mathbf{U}_j^T\frac{\partial \mathbf{A}}{\partial p_\alpha}\mathbf{V}_i + \mathbf{U}_j^T\mathbf{A}\frac{\partial \mathbf{V}_i}{\partial p_\alpha} = \mathbf{U}_j^T\frac{\partial \lambda_i}{\partial p_\alpha}\mathbf{V}_i + \mathbf{U}_j^T\lambda_i\frac{\partial \mathbf{V}_i}{\partial p_\alpha} \tag{4.10}$$

Now, substituting the expansions (4.9) and using Equation (4.6), and the bi-orthogonal property of $\mathbf{U}_i^T\mathbf{V}_j = 0$ *iff* $i \neq j$, we obtain,

$$g_{ij_\alpha} = \frac{\mathbf{U}_j^T \frac{\partial \mathbf{A}}{\partial p_\alpha}\mathbf{V}_i}{(\lambda_i - \lambda_j)\mathbf{U}_j^T\mathbf{V}_j}, \quad i \neq j \tag{4.11}$$

Proceeding in a similar fashion, after differentiating Equation (4.6) *w.r.t* p_α,

$$h_{ij_\alpha} = \frac{\mathbf{U}_i^T \frac{\partial \mathbf{A}}{\partial p_\alpha}\mathbf{V}_j}{(\lambda_i - \lambda_j)\mathbf{U}_j^T\mathbf{V}_j}, \quad i \neq j \tag{4.12}$$

The above expressions for g_{ij_α} and h_{ij_α} were obtained by Rogers (1970). Since $\mathbf{U}_i^T\mathbf{V}_i = 1$ which is the normalization condition due to bi-orthogonality,

$$\frac{\partial \lambda_i}{\partial p_\alpha} = \mathbf{U}_i^T\frac{\partial \mathbf{A}}{\partial p_\alpha}\mathbf{V}_i \tag{4.13}$$

$$g_{ij_\alpha} = \frac{\mathbf{U}_j^T \frac{\partial \mathbf{A}}{\partial p_\alpha}\mathbf{V}_i}{(\lambda_i - \lambda_j)}, \quad i \neq j \tag{4.14}$$

$$h_{ij_\alpha} = \frac{\mathbf{U}_i^T \frac{\partial \mathbf{A}}{\partial p_\alpha}\mathbf{V}_j}{(\lambda_i - \lambda_j)}, \quad i \neq j \tag{4.15}$$

It can be observed that,

$$h_{ij_\alpha} \;=\; -g_{ji_\alpha}\frac{\mathbf{U}_i^T\mathbf{V}_i}{\mathbf{U}_j^T\mathbf{V}_j} \;=\; -\,g_{ji_\alpha} \tag{4.16}$$

To calculate the diagonal terms g_{ii_α} and h_{ii_α}, Nelson (1976), used the index, m_j such that,

$$\mid V_{m_jj} \mid \;=\; \max_i \big(\mid V_{ij} \mid\big) \tag{4.17}$$

which gives the diagonal elements of g as,

$$g_{kk_\alpha} \;=\; -\sum_{\substack{l=1\\ l\neq k}}^{N} g_{kl_\alpha} V_{m_kl} \tag{4.18}$$

Equations (4.9) and (4.13) give the required first-order derivatives of the eigenvectors and the eigenvalues *w.r.t.* each of the parameters p_α of the system. The derivative of the Jacobian $\frac{\partial \mathbf{A}}{\partial p_\alpha}$ has to be evaluated beforehand.

Coming back to the expression of $\frac{\partial\,\mathrm{Cov}(\tau,\mathbf{p})_{ij}}{\partial p_\alpha}$ given by the equation (4.4) we now require the derivative of the stationary covariance terms *w.r.t* p_α. Firstly, the stationary Covariance Matrix $\mathbf{C}$ is given by,

$$\mathbf{C} \;=\; \mathbf{V}\,.\,\widetilde{\mathbf{C}}\,.\,\mathbf{V}^T \tag{4.19}$$

where $\widetilde{\mathbf{C}}$ is the covariance matrix for a new set of variables $\delta\widetilde{\mathbf{X}} = \mathbf{V}^{-1}\delta\mathbf{X}$, *i.e.,* fluctuations as normal modes. The other transformations are, $\widetilde{\boldsymbol{\nu}} = \mathbf{V}^{-1}\boldsymbol{\nu}$ which is the new Stoichiometric Matrix, $\widetilde{\mathbf{B}} = \mathbf{U}^T\mathbf{B}\mathbf{U}$ which is the new Diffusion Matrix. These transformations are necessary to solve the Lyapunov Matrix Equation $\mathbf{A}\mathbf{C} + \mathbf{C}\mathbf{A}^T + \mathbf{B}\mathbf{B}^T = 0$ for the covariance term $\mathbf{C}$ analytically Elf and Ehrenberg (2003). Rewriting the above equation element-wise,

$$\begin{aligned} C_{ij} \;&=\; \sum_{m=1}^{N} V_{im}\big(\widetilde{\mathbf{C}}\,.\,\mathbf{V}^T\big)_{mj} \\ &=\; \sum_{m=1}^{N} V_{im}\sum_{n=1}^{N}\widetilde{C}_{mn}V_{jn} \end{aligned}$$

whose partial derivative *w.r.t* p_α is,

$$\begin{aligned}\frac{\partial C_{ij}}{\partial p_\alpha} &= \sum_{m=1}^{N}\frac{\partial V_{im}}{\partial p_\alpha}\sum_{n=1}^{N}\widetilde{C}_{mn}V_{jn} \\ &\quad + \sum_{m=1}^{N}V_{im}\sum_{n=1}^{N}\frac{\partial \widetilde{C}_{mn}}{\partial p_\alpha}V_{jn} \\ &\quad + \sum_{m=1}^{N}V_{im}\sum_{n=1}^{N}\widetilde{C}_{mn}\frac{\partial V_{jn}}{\partial p_\alpha}\end{aligned} \tag{4.20}$$

The expression for $\widetilde{C}_{mn}$ as derived in the previous chapter (equation (3.9)) is

$$\widetilde{C}_{mn} = -\frac{\widetilde{B}_{mn}}{\lambda_m + \lambda_n} \tag{4.21}$$

whose derivative *w.r.t* p_α is,

$$\begin{aligned}\frac{\partial \widetilde{C}_{mn}}{\partial p_\alpha} &= \frac{\partial \widetilde{C}_{mn}}{\partial \widetilde{B}_{mn}}\frac{\partial \widetilde{B}_{mn}}{\partial p_\alpha} + \frac{\partial \widetilde{C}_{mn}}{\partial \lambda_m}\frac{\partial \lambda_m}{\partial p_\alpha} + \frac{\partial \widetilde{C}_{mn}}{\partial \lambda_n}\frac{\partial \lambda_n}{\partial p_\alpha} \\ &= -\frac{\partial \widetilde{B}_{mn}/\partial p_\alpha}{(\lambda_m+\lambda_n)} + \frac{\widetilde{B}_{mn}}{(\lambda_m+\lambda_n)^2}\Big(\frac{\partial \lambda_m}{\partial p_\alpha} + \frac{\partial \lambda_n}{\partial p_\alpha}\Big)\end{aligned} \tag{4.22}$$

Now, this expression contains the term $\frac{\partial \widetilde{B}_{mn}}{\partial p_\alpha}$ which needs to be evaluated first. The elements of $\widetilde{\mathbf{B}}$ are,

$$\widetilde{B}_{mn} = \sum_{q=1}^{N}U_{qm}\sum_{r=1}^{N}B_{qr}U_{rn}$$

whose partial derivative *w.r.t* p_α is,

$$\begin{aligned}\frac{\partial \widetilde{B}_{mn}}{\partial p_\alpha} &= \sum_{q=1}^{N}\frac{\partial U_{qm}}{\partial p_\alpha}\sum_{r=1}^{N}B_{qr}U_{rn} \\ &\quad + \sum_{q=1}^{N}U_{qm}\sum_{r=1}^{N}\frac{\partial B_{qr}}{\partial p_\alpha}U_{rn} \\ &\quad + \sum_{q=1}^{N}U_{qm}\sum_{r=1}^{N}B_{qr}\frac{\partial U_{rn}}{\partial p_\alpha}\end{aligned} \tag{4.23}$$

Once more, to evaluate this expression, we need another derivative term which is $\frac{\partial B_{qr}}{\partial p_\alpha}$, where

$$\begin{aligned} B_{qr} &= \left(\boldsymbol{\nu} \,.\, \mathbf{diag}(\mathbf{R}) \,.\, \boldsymbol{\nu}^T\right)_{qr} \\ &= \sum_{s=1}^{M} \boldsymbol{\nu}_{qs} \sum_{t=1}^{M} (diag(R))_{st} \boldsymbol{\nu}^T_{tr} \\ &= \sum_{s=1}^{M} \boldsymbol{\nu}_{qs} (diag(R))_{ss} \boldsymbol{\nu}_{rs} \quad , \quad \text{since } \mathbf{diag}(\mathbf{R}) \text{ is a diagonal matrix} \\ &= \sum_{s=1}^{M} \boldsymbol{\nu}_{qs} R_s \boldsymbol{\nu}_{rs} \end{aligned} \tag{4.24}$$

$\boldsymbol{\nu}$ is the stoichiometry matrix and $\mathbf{diag}(\mathbf{R})$ is the diagonal matrix whose diagonal elements are the elements of the vector $\mathbf{R}$. $\mathbf{R}$ is a vector whose elements are the deterministic rates or fluxes of each reaction. The derivative of B_{qr} *w.r.t* p_α is therefore,

$$\frac{\partial B_{qr}}{\partial p_\alpha} = \sum_{s=1}^{M} \boldsymbol{\nu}_{qs} \frac{\partial R_s}{\partial p_\alpha} \boldsymbol{\nu}_{rs} \tag{4.25}$$

Since the R_s terms are readily available from the deterministic rate equations, it is straight forward to obtain $\frac{\partial R_s}{\partial p_\alpha}$. For example, let R_s be say $k_M^- \langle M \rangle$, where $\langle M \rangle$ is the expression for the mean steady state value of the mRNA, which is turn is a function of the parameters p_α. Therefore R_s is wholly a function in the parameters and is therefore readily differentiated *w.r.t* these parameters. Finally, evaluating all the derivatives in the reverse order, *i.e.,* starting with $\frac{\partial R_s}{\partial p_\alpha}$ and then $\frac{\partial B_{qr}}{\partial p_\alpha}$ [equation (4.25)], $\frac{\partial \widetilde{B}_{mn}}{\partial p_\alpha}$ [equation (4.23)], $\frac{\partial \widetilde{C}_{mn}}{\partial p_\alpha}$ [equation (4.22)], we arrive at the expression for $\frac{\partial C_{ij}}{\partial p_\alpha}$ [equation (4.20)]. This brings us back to $\frac{\partial \operatorname{Cov}(\tau,\mathbf{p})_{ij}}{\partial p_\alpha}$ of equation (4.4) where all the terms on the R.H.S are now available. Note that the derivatives of the eigenvalues and the eigenvectors are obtained from equations (4.13) and (4.9) respectively. Representing $\operatorname{Corr}(\tau, \mathbf{p})$ simply as $\operatorname{Corr}(\tau)$, the dynamic correlations are nothing but,

$$\operatorname{Corr}(\tau)_{ij} = \frac{\operatorname{Cov}(\tau)_{ij}}{\sqrt{\operatorname{Cov}(0)_i \operatorname{Cov}(0)_j}} \equiv \frac{\operatorname{Cov}(\tau)_{ij}}{\sqrt{\operatorname{Var}(0)_i \operatorname{Var}(0)_j}} \tag{4.26}$$

their sensitivity *w.r.t* the parameters are given by the expression,

$$\frac{\partial \operatorname{Corr}(\tau)_{ij}}{\partial p_\alpha} = \frac{\partial \operatorname{Cov}(\tau)_{ij}/\partial p_\alpha}{\sqrt{\operatorname{Var}(0)_i \operatorname{Var}(0)_j}} - \frac{\operatorname{Cov}(\tau)_{ij}}{2} \left[\frac{\partial \operatorname{Var}(\tau)_i/\partial p_\alpha}{\sqrt{\operatorname{Var}(0)_i^3 \operatorname{Var}(0)_j}} + \frac{\partial \operatorname{Var}(\tau)_j/\partial p_\alpha}{\sqrt{\operatorname{Var}(0)_i \operatorname{Var}(0)_j^3}} \right] \tag{4.27}$$

4.2 Sensitivity of τ^* w.r.t system parameters

In this section we deduce the sensitivity of τ^* *w.r.t.* the parameters of the system. As the expression for τ^* involves the parameters, it is dependent on $\mathbf{p}$. This sensitivity which is denoted as $\frac{\partial \tau^*}{\partial p_\alpha} = \frac{\partial \tau}{\partial p_\alpha}\Big|_{\tau=\tau^*}$ is our object of interest. The expression for τ^* is obtained via differentiating $\mathrm{Cov}(\tau, \mathbf{p})$ *w.r.t* τ (at τ^*) and equating to zero,

$$\frac{\partial \,\mathrm{Cov}(\tau, \mathbf{p})}{\partial \tau}\bigg|_{\tau=\tau^*} \;=\; 0$$

For the sake of simplicity denoting $\frac{\partial \,\mathrm{Cov}(\tau,\mathbf{p})}{\partial \tau}$ as $\digamma(\tau, \mathbf{p})$, the expression for τ^* is obtained from,

$$\digamma(\mathbf{p})\Big|_{\tau=\tau^*} \;=\; \sum_l \lambda_l e^{\lambda_l \tau^*} V_{il}\Big(\sum_k U_{lk}^T C_{kj}\Big) \;=\; 0$$

Therefore $\digamma(\tau, \mathbf{p})$ at time $\tau = \tau^*$ can be said to be a function only in the M independent variables $\mathbf{p}$, while the dependent variable is τ^*. Differentiating $\digamma$ *w.r.t* one of these independent variables p_α, while holding the other $M-1$ parameters constant,

$$\frac{d\digamma}{dp_\alpha} \;=\; \frac{\partial \digamma}{\partial \tau} \cdot \frac{\partial \tau}{\partial p_\alpha} + \frac{\partial \digamma}{\partial p_\alpha}$$

The above equation makes sense only if τ is at some value such as τ^*, which is when it is dependent on p_α. Now, since $\digamma(\tau, \mathbf{p})\Big|_{\tau=\tau^*} = \frac{\partial \,\mathrm{Cov}(\tau,\mathbf{p})}{\partial \tau}\Big|_{\tau=\tau^*} = 0$, we have

$$\begin{aligned} \frac{d\digamma}{dp_\alpha}\bigg|_{\tau^*} &= 0 \\ \Longrightarrow \qquad \frac{\partial \tau}{\partial p_\alpha}\bigg|_{\tau^*} &= -\frac{\partial \digamma / \partial p_\alpha}{\partial \digamma / \partial \tau}\bigg|_{\tau^*} \end{aligned} \tag{4.28}$$

where,

$$\begin{aligned} \frac{\partial \digamma}{\partial p_\alpha}\bigg|_{\tau^*} &= \frac{\partial^2 \,\mathrm{Cov}(\tau, \mathbf{p})}{\partial \tau \partial p_\alpha}\bigg|_{\tau^*} \\ &= \sum_l \frac{\partial \lambda_l}{\partial p_\alpha} e^{\lambda_l \tau^*} V_{il} \sum_k U_{kl} C_{kj} + \sum_l \lambda_l \tau^* e^{\lambda_l \tau^*} \frac{\partial \lambda_l}{\partial p_\alpha} V_{il} \sum_k U_{kl} C_{kj} \\ &\quad + \sum_l \lambda_l e^{\lambda_l \tau^*} \frac{\partial V_{il}}{\partial p_\alpha} \sum_k U_{kl} C_{kj} + \sum_l \lambda_l e^{\lambda_l \tau^*} V_{il} \sum_k \frac{\partial U_{kl}}{\partial p_\alpha} C_{kj} \\ &\quad + \sum_l \lambda_l e^{\lambda_l \tau^*} V_{il} \sum_k U_{kl} \frac{\partial C_{kj}}{\partial p_\alpha} \end{aligned}$$

and

$$\left.\frac{\partial F}{\partial \tau}\right|_{\tau^*} = \left.\frac{\partial^2 \operatorname{Cov}(\tau, \mathbf{p})}{\partial \tau^2}\right|_{\tau^*} = \sum_l \lambda_l^2 e^{\lambda_l \tau^*} V_{il} \Big(\sum_k U_{kl} C_{kj} \Big)$$

The expression for $\frac{\partial \tau^*}{\partial p_\alpha}$ holds true only at specific values of p_α denoted as $(\tilde{p}_\alpha)$ and at time τ^*, which is equivalent to saying that the terms on the R.H.S of equation (4.28) are applicable only in a small region around $(\tilde{p}_\alpha, \tau^*)$, due to the effect of linearization of the highly nonlinear covariance function. Hence, equation (4.28) is rewritten more specifically as:

$$\left.\frac{\partial \tau}{\partial p_\alpha}\right|_{\substack{at\ a\ specific \\ value\ of\ (p_\alpha,\tau)=(\tilde{p}_\alpha,\tau^*)}} = -\left.\frac{\partial F/\partial p_\alpha}{\partial F/\partial \tau}\right|_{(\tilde{p}_\alpha,\tau^*)} \tag{4.29}$$

NOTE:

- τ^* is the τ at which the covariance function $\operatorname{Cov}(\tau, p_\alpha)_{ij}$ attains maximum value (at a specific fixed value of $p_\alpha = \tilde{p}_\alpha$). Hence, each $\tilde{p}_\alpha$ corresponds to a specific, or in other words, results in a specific τ^*
- $\left.\frac{\partial \tau}{\partial p_\alpha}\right|_{(\tilde{p}_\alpha,\tau^*)}$ is the sensitivity of τ^* to a very small variation in $\tilde{p}_\alpha$. Hence, equation (4.29) is applicable only in a very small region around $\tilde{p}_\alpha$. Moving further away from this value $\tilde{p}_\alpha$, *i.e.,* for a different value of p_α, one would obtain a different value for τ^* which requires recalculation of the sensitivity. This is true even for $\frac{\partial \operatorname{Corr}^*}{\partial p_\alpha}$.
- This is due to the highly *non-linear* nature of $\operatorname{Cov}(\tau, \mathbf{p})_{ij}$. The *non-linearity* is due to both $\mathbf{p}$ and τ. In the next chapter where we tabulate the numerical values of the sensitivities in cases of some regulatory systems, we mention whether these sensitivities are of the type *linear* or *non-linear* depending on whether for larger changes in the parameter values, the sensitivities that are evaluated, *do* or *do not* hold true. In simple words, if $\frac{\Delta \operatorname{Corr}^*}{\Delta p_\alpha} \equiv \frac{\partial \operatorname{Corr}^*}{\partial p_\alpha}$ for large changes in the values of p_α, then the sensitivity *w.r.t* this parameter is said to be of type *linear* else of type *non-linear*.

4.3 Summary

The reaction rate constants, termed as parameters p_α in the present context, influence the fluctuation properties of the system and this influence is what we have quantified as sensitivities in the present chapter. This quantification also serves the purpose of estimating the values of rate constants given the experimental values for the correlations. In the previous chapter, τ^* was shown to be a function in the decay rates of the mRNA and proteins and therefore sensitive to changes in their values. Likewise, in the next chapter (and in Appendix B) where we derive the expressions for τ^* in the case of a two-gene *elementary activator* system, the techniques demonstrated in this chapter would prove useful in deriving numerical values for the sensitivities of τ^* *w.r.t* each of the parameters of the system. This would give useful insight into the effect that these parameters have on the dynamic correlation functions or simply on the *features* of these correlations that are $[\text{Corr}^*, \tau^*]$.

Chapter 5

Regulatory Mechanisms

As was demonstrated in chapter 3, for the case of a single gene, the dynamic correlation between two molecular species gives insight into the process behind their interactions. The defining *features* of the correlation function are the stationary correlations (at time $\tau = 0$), the peak value of dynamic correlations labelled as Corr* and the time τ^* at which this peak occurs. From the analytical expression for τ^* it was clear that the decay rates of mRNAs and proteins (k_M^- and k_P^-) were in effect controlling the temporal aspect of the dynamic correlations, which meant that the sensitivities $\frac{\partial \tau^*}{\partial k_M^-}$ and $\frac{\partial \tau^*}{\partial k_P^-}$ would be high. In Tables 5.3 and 5.6 where numerical values for these sensivities are calculated for the two-gene activator and repressor systems, we find that indeed τ^* is sensitive only to the decay rates of mRNAs and proteins. On similar lines, we notice that the correlation magnitudes are also sensitive in various parameters of the GRN. With this background in mind, our intention here is therefore to evaluate the dynamic correlation functions between two species of a GRN and monitor their behaviour as observed in their *features*, for significant *variations* in certain *factors* of the GRN. These *factors* are *(i)* the values that the rate constants take, for which the correlations are sensitive, *(ii)* the regulatory mechanisms that are involved in the transcription process, such as activation or repression and finally *(iii)* the network mechanism describing the connectivity between genes. In this chapter we shall deal with the first two *factors*, while the last *factor* of network mechanism is dealt with in the next chapter. Because the dynamic correlations between any two species are effectively due to the reaction rate constants, they are expected to behave uniquely for *variations* in each of the above *factors*, *i.e.,* for different values of the rate constants, different regulatory mechanisms present between two genes and for different network connectivities between them. Therefore the dynamic correlation serves as a *signature* that could be used to differentiate or characterize various regulatory networks. The choice of species for most of our work are the two proteins, one that acts as regulator protein or the transcription factor and the other that is the regulated protein.

Firstly, we consider the case of a two-gene activator system, which we refer to as the *elementary activator*, where the element X activates the element Y or realistically speaking, protein P_x acts as the activating transcription factor for the production of protein P_y. Variations are then introduced in the values of rate constants which result in changes in the correlation between P_x and P_y. Next, the regulatory mechanism which in this case is that of activation by P_x, is replaced by activation that is effected via the protein dimerization process, *i.e.,* the proteins P_x now form dimers P_{x_2} which then bind to the upstream sequences of the gene of Y. This regulatory mechanism throws up distinct protein correlations. We further study the regulatory mechanism wherein P_x acts as a repressor, and for which there is a qualitative change in the protein correlations as compared to the case of activation.

To effectively analyze the dynamic correlations of a gene network we need a basic model of the gene regulation process. Simple yet effective models are those that not only incorporate the essence of the regulatory process but are easy to understand and replicate. Therefore we choose a set of elementary reactions that captures the significant steps of the regulatory process. We take into account the translation process of the mRNAs, decay of mRNAs and proteins, and transcription of the downstream gene. If X and Y denote the regulator and regulated elements, the *constituent* molecular species of such a model would then be the regulated gene G_y, mRNAs M_x, M_y and the proteins P_x, P_y. P_x acts as a transcription factor and regulates the transcription of the mRNA M_y. The *non-constituent* elements are the RNA polymerases and ribosomes that increase the complexity of the model. By employing fewer elements we intend to reduce the complexity of the model, while retaining the essence of transcriptional regulation. Since the focus of study is the regulatory link between genes, the production of mRNA transcripts of upstream gene X is assumed to be through a constant source and at a constant rate of flux.

5.1 Elementary Activator

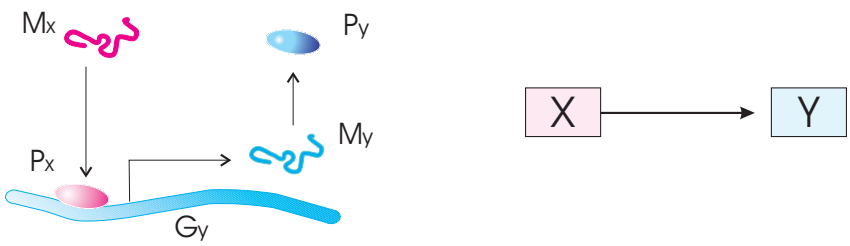

FIGURE 5.1: The figure on the left is a schematic representation of the regulatory process where the mRNA M_y is transcribed from the coding sequence of its gene G_y. P_x is a positive regulator or an activator of this transcription. The equivalent network representation is shown on the right where the activation of P_y by P_x is represented as element Y being activated by the element X, arrow indicating activation.

A schematic representation of an *elementary activator* is given in Figure 5.1, where the gene-node X directly activates Y without any other intervening influences on the promoter region of G_y, and the resulting biochemical reactions of which are given in Table 5.1. The regulatory sequence upstream of G_y is bound by the transcription factor P_x, which is an activator in the present case, resulting in the complex $\mathcal{C}_y$ which is in effect the active or ON state of the gene G_y. The transcription of the mRNA of the downstream gene M_y is then initiated from this complex. The remaining reactions are similar to the case of the single-gene system. The values for the rate constants that are biologically relevant are chosen such that the mean steady state values of the mRNAs and proteins and their decay rates correspond to a regulatory network in the yeast organism. This is the $\{CHA4 \rightarrow CHA1\}$ regulatory link, on which we shall elaborate later.

$\left[\phi \xrightarrow{k^+_{M_x}} M_x\right]$	$\left[P_x + G_y \underset{k_{off}}{\overset{k_{on}}{\rightleftharpoons}} \mathcal{C}_y\right]$	$\left[M_y \xrightarrow{k^+_{P_y}} M_y + P_y\right]$	$\left[M_x \xrightarrow{k^-_{M_x}} \phi\right]$	$\left[P_x \xrightarrow{k^-_{P_x}} \phi\right]$
$\left[M_x \xrightarrow{k^+_{P_x}} M_x + P_x\right]$	$\left[\mathcal{C}_y \xrightarrow{k^+_{M_y}} \mathcal{C}_y + M_y\right]$		$\left[M_y \xrightarrow{k^-_{M_y}} \phi\right]$	$\left[P_y \xrightarrow{k^-_{P_y}} \phi\right]$

TABLE 5.1: Reaction set describing the process of activation between elements X and Y. Production of M_x is assumed to be of Poissonian birth-death process. $\phi \rightarrow M_x$ therefore denotes spontaneous creation of M_x from a constant source ϕ.

The deterministic dynamics of this system are governed by a set of coupled ODEs that describe the time-evolution of the mean concentration levels of the species. Referring back, this is nothing but the evolution of the mean of the Gaussian distributions of the species, where the distributions are the solution to the Fokker Planck equation (3.7). Alternatively, these rate equations could also be obtained by simply ignoring the fluctuations around them through linearization of the propensity functions, as was done for obtaining equation (3.5). The significance of the fluctuation dissipation relations comes to the fore, where we show that these deterministic rate equations are sufficient for deriving the covariances between the species that are stochastic variables in reality.

Hence, the internal fluctuations and the deterministic dynamics of the system are interrelated. The rate equations are as follows:

$$\begin{aligned}
\frac{dG_y}{dt} &= -k_{on}P_xG_y + k_{off}\mathcal{C}_y \\
\frac{d\mathcal{C}_y}{dt} &= -k_{off}\mathcal{C}_y + k_{on}P_xG_y \\
\frac{dM_x}{dt} &= k^+_{M_x} - k^-_{M_x}M_x \\
\frac{dM_y}{dt} &= k^+_{M_y}\mathcal{C}_y - k^-_{M_y}M_y \\
\frac{dP_x}{dt} &= k^+_{M_x}M_x - k^-_{P_x}P_x - k_{on}P_xG_y + k_{off}\mathcal{C}_y \\
\frac{dP_y}{dt} &= k^+_{P_y}M_y - k^-_{P_y}P_y
\end{aligned}$$

where the angled brackets, in $\langle M_x \rangle$ etc that represent the mean concentration levels, are omitted for convenience. The total amount of G_y present in the cell at time t is $\mathcal{C}_y(t) + G_y(t) = G_y(t_0)$. Hence, $\mathcal{C}_y(t) = \alpha - G_y(t)$ where $\alpha = G_y(t_0)$ is the initial concentration. Assuming the presence of a single copy of the gene in the yeast nucleus of volume 1 μm^3 the approximate initial concentration of G_y is 1 nM. Using the above substitution for $\mathcal{C}_y$ and solving for the second of the above rate equations we have at steady state, $\mathcal{C}_y = \frac{\alpha P_x}{K_D+P_x}$ and $G_y = \frac{\alpha K_D}{K_D+P_x}$, where $K_D = k_{off}/k_{on}$ is the dissociation constant associated with the binding and unbinding of P_x to the regulatory region upstream of G_y. The mean steady state value of the product M_y then follows the kinetics of Michaelis-Menten (hyperbolic) type where $\langle M_y \rangle = \frac{k^+_{M_y}}{k^-_{M_y}} \frac{\alpha P_x}{K_D+P_x}$ where $k^+_{M_y}$ is the transcription rate. The Hill-coefficient in this case is equal to 1 corresponding to activation through protein monomers. Note that the basal transcription rate is not included in this model, as it does not represent the regulatory link between the two genes. Now, retaining G_y as the non-redundant variable and substituting for $\mathcal{C}_y$ in the other equations, we eliminate the redundancy in the above set of ODEs. The molecular species or simply the variables that completely define the dynamics of the system now are $[G_y, M_x, M_y, P_x, P_y]$ whose rate equations are,

$$\frac{dG_y}{dt} = -k_{on}(P_xG_y) + k_{off}(\alpha - G_y) \tag{5.1}$$

$$\frac{dM_x}{dt} = k^+_{M_x} - k^-_{M_x}M_x \tag{5.2}$$

$$\frac{dM_y}{dt} = k^+_{M_y}(\alpha - G_y) - k^-_{M_y}M_y \tag{5.3}$$

$$\frac{dP_x}{dt} = k^+_{M_x}M_x - k^-_{P_x}P_x - k_{on}(P_xG_y) + k_{off}(\alpha - G_y) \tag{5.4}$$

$$\frac{dP_y}{dt} = k^+_{P_y}M_y - k^-_{P_y}P_y. \tag{5.5}$$

These rate equations not only describe the system's deterministic behaviour but, by the fluctuation-dissipation relations, also describe the internal (linear) fluctuations of

the system. This is clear from the very expression of the dynamic covariances of these fluctuations. The basic quantities required to solve for the dynamic covariances are the deterministic rates and the stoichiometry which are obtained from the above rate equations. The deterministic rates of the system are nothing but the individual fluxes responsible for the increase or decrease in the production rate of the five variables. They are the terms on the R.H.S of the ODEs. Therefore the vector of deterministic rates is,

$$\begin{aligned}\mathbf{R} \;=\; \Big(&k_{on}P_xG_y, k_{off}(\alpha - G_y), k^+_{M_y}(\alpha - G_y), k^+_{M_x}, k^+_{P_x}M_x, k^+_{P_y}M_y,\\ &k^-_{M_x}M_x, k^-_{M_y}M_y, k^-_{P_x}P_x, k^-_{P_y}P_y\Big)^T\end{aligned}$$

The elements of the stoichiometric matrix $\boldsymbol{\nu}$ simply indicate whether the above deterministic rates either increase or decrease each of the variables and if so, by what amount.

$$\boldsymbol{\nu} = \begin{pmatrix} -1 & +1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & +1 & 0 & 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & +1 & 0 & 0 & 0 & 0 & -1 & 0 & 0 \\ -1 & +1 & 0 & 0 & +1 & 0 & 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & +1 & 0 & 0 & 0 & -1 \end{pmatrix}$$

The relation between the Jacobian $\mathbf{A}$, and the above quantities is therefore,

$$\mathbf{A}_{ik} = \frac{\partial(\sum_{j=1}^{M} \nu_{ij} R_j(\langle \mathbf{x} \rangle))}{\partial \langle x_k \rangle}$$

where $\mathbf{x}$ is the vector of the variables $[G_y, M_x, M_y, P_x, P_y]$. The Jacobian matrix for the system of these variables is,

$$\mathbf{A} \;=\; \begin{pmatrix} -(k_{off} + k_{on}P_x) & 0 & 0 & -k_{on}G_y & 0 \\ 0 & -k^-_{M_x} & 0 & 0 & 0 \\ -k^+_{M_y} & 0 & -k^-_{M_y} & 0 & 0 \\ -(k_{off} + k_{on}P_x) & k^+_{P_x} & 0 & -(k^-_{P_x} + k_{on}G_y) & 0 \\ 0 & 0 & k^+_{P_y} & 0 & -k^-_{P_y} \end{pmatrix}. \qquad (5.6)$$

It now becomes essential to work within the range of the parameter values that is biologically plausible. For this purpose we search through the well-documented database of the organism *Saccharomyces cerevisiae* more commonly known as baker's yeast. Transcriptional regulation in yeast is mediated through cis-acting sequence elements that are located upstream to the gene sequence. The transcription factors that bind to these Upstream Activating Sequences (UASs) work as activators when they assist the transcription initiation complex in binding strongly to the promoter region of the gene and thus transcribing the mRNAs. Since it would be extreme to expect two genes working

together, but in isolation to other genes, we *assume* that the downstream gene is regulated *entirely* by the product of the upstream gene. In reality, however, this regulation would involve tens of different TFs acting in a coordinated combinatorial fashion on the UASs. As a first step towards modelling such a complex process, we assume the presence of a single type of TF. This assumption is later expanded to account for the presence of more than one type of TF.

The information regarding the TFs that regulate each gene in yeast, is made available to the general public by projects such as the *Saccharomyces cerevisiae systematic sequencing* project well known as the *SGD* project[1] and more recently by the *YEASTRACT*[2] group Teixeira and et al. (2006); Monteiro and et al. (2008). From this database we adopt a regulatory system where there is direct evidence of an activating link between two genes. Such an activating link exists between the genes *CHA4* and *CHA1* of yeast Bornaes and et al. (1993); Holmberg and Schjerling (1996). The *CHA1* gene encodes the catabolic L-serine deaminase responsible for the utilization of serine as nitrogen sources. The protein *Cha4p* acts as a DNA-binding transcription factor and activates the regulation of the *CHA1* gene. In reality, there may be numerous other TFs other than *Cha4p* that activate/repress the production of the mRNA of *CHA1* gene. In fact, a *YEASTRACT* search for the TFs of *CHA1* results in 22 of which one is *Cha4p*. However, our assumption combines the effect of all the other TFs onto the one TF *Cha4p*. In the next chapter, we study the effect of including other TFs, such as the activator binding to the upstream region of *CHA1* in the presence of a repressor protein. In the present case, we consider *CHA4* to be the upstream X gene and *CHA1* to be the regulated downstream gene, Y.

Activator System					
Std Name	Sys Name	mRNA (nM)	protein (nM)	mRNA $t_{1/2}$ (min)	protein $t_{1/2}$ (min)
CHA4	YLR098C	0.2	395.6	21	65
CHA1	YCL064C	4.5	36602	10	70

TABLE 5.2: The values are of the CHA4 → CHA1 activator link in yeast. Nuclear volume is taken to be around 1 μm^3, resulting in nano-Molar concentrations for the mRNA and protein species. The mRNA and protein half-lives $t_{1/2}$ are in minutes.

The mean steady state protein and mRNA levels of the *CHA4* and *CHA1*, and their respective decay rates or half-lives are obtained from the experimental datasets of Arava et al. (2003); Ghaemmaghami and et al. (2003); Wang et al. (2002); Belle et al. (2006) and are reproduced in Table 5.2. It is important to note that the experimental conditions under which each of the above values were obtained could be quite different. However, since closely linked genes work together under varying conditions, we assume that the

[1] http://www.yeastgenome.org
[2] http://www.yeastract.com

network connectivity is static in nature. Hence, we utilize the above values in our models. The time taken for the yeast cells to divide, denoted as t_{double} is assumed to be 90 minutes. The mRNA and protein decay rates are therefore:

$$k^-_{(M,P)} = \frac{ln(2)}{(t_{1/2})_{(M,P)}} + \frac{ln(2)}{t_{double}} \tag{5.7}$$

Therefore, the decay rates in the case of the $CHA4 \rightarrow CHA1$ system are:

$$\begin{aligned}
k^-_{M_x} &= \frac{ln(2)}{21} + \frac{ln(2)}{90} = 0.0407\ min^{-1} \\
k^-_{M_y} &= \frac{ln(2)}{10} + \frac{ln(2)}{90} = 0.0770\ min^{-1} \\
k^-_{P_x} &= \frac{ln(2)}{65} + \frac{ln(2)}{90} = 0.0184\ min^{-1} \\
k^-_{P_y} &= \frac{ln(2)}{70} + \frac{ln(2)}{90} = 0.0176\ min^{-1}
\end{aligned}$$

To derive the values for the other rate constants, we do the following. The expressions for the levels of mRNAs and proteins at steady state are derived by setting the rate equations to zero. They are,

$$\begin{aligned}
k^+_{M_x} &= k^-_{M_x} M_x = 8.14 \times 10^{-3}\ nM\ min^{-1} \\
k^+_{M_y} &= \frac{k^-_{M_y} M_y}{P_x} = 0.52\ min^{-1} \\
k^+_{P_x} &= \frac{k^-_{P_x} P_x}{M_x} = 36.3\ min^{-1} \\
k^+_{P_y} &= \frac{k^-_{P_y} P_y}{M_y} = 143.2\ min^{-1}
\end{aligned}$$

The only remaining rate constants which are the ON and OFF rates of the DNA are chosen such that their ratio which is the dissociation constant $K_D = k_{off}/k_{on}$ is equal to 200 nM. Though this value is biologically respectable, the logical reasoning is as follows. If the amount of protein P_x is very large compared to the dissociation constant K_D, the gene is present mostly in the active state and hence $M_y \approx \frac{k^+_{M_y}\alpha}{k^-_{M_y}}$ involves a constant flux $k^+_{M_y}\alpha$ similar to the case of M_x. On the other hand, if protein concentration is less than the dissociation constant, $M_y = \frac{k^+_{M_y}}{k^-_{M_y}} \frac{\alpha P_x}{K_D}$ as the gene is now mostly present in the inactive state. Hence, in our model of the elementary activator, we take a value of 200 nM which is of the order of the mean concentration of $P_x = 395.6\ nM$ and hence allows the gene to be in the active state and also includes the effect of P_x and G_y in the rate of production of M_y. Individually, the rates are $k_{on} = 1\ nM^{-1}\ min^{-1}$ and $k_{off} = 200\ min^{-1}$.

5.1.1 Dynamic correlation between regulated and regulator proteins

Correlations between molecular species of a regulatory network are interesting to observe. We focus on the correlations between the regulator and the regulated proteins. The $\mathrm{Corr}[P_x(t), P_y(t+\tau)]$ function as defined by equation (3.15) is the covariance normalized by the stationary auto-covariances. For example, at time $\tau = 0$, we have

$$\begin{aligned}\mathrm{Corr}[P_x(t), P_y(t)] &= \frac{\mathrm{Cov}[P_x(t), P_y(t)]}{\sqrt{\mathrm{Cov}[P_x(t), P_x(t)]\,\mathrm{Cov}[P_y(t), P_y(t)]}} \\ &= \frac{4.219 \times 10^6}{\sqrt{(2.436 \times 10^5)(2.029 \times 10^8)}} = 0.6\end{aligned}$$

Likewise, the correlation at all other times $\tau > 0$ is calculated and is plotted in Figure 5.3. Note that the dynamic character of the covariances is preserved in the correlations. For convenient reference in further sections, we shall term the correlation in Figure 5.3 as that corresponding to the *base case*, where the parameter values that are used to derive this correlation are from the $CHA4 \rightarrow CHA1$ system and these values are equivalently referred to as *base values*.

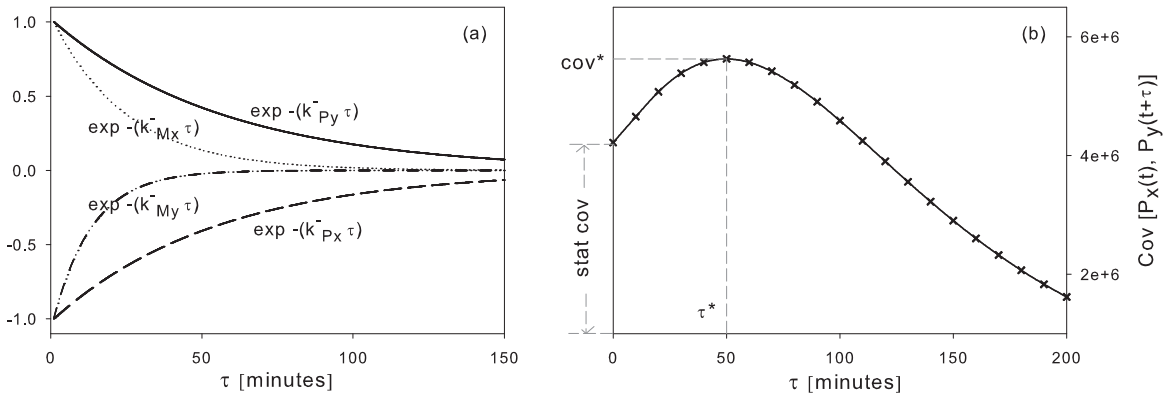

FIGURE 5.2: Exponentials raised to the negative power of the eigenvalues of the $X \rightarrow Y$ system are shown in figure (a). The coefficients of these exponentials are obatined from equation (3.14), but are scaled in the figure above. The resulting sum of these exponentials gives a characteristic shape to the time-covariance function, shown in figure (b). Also, the magnitude of the stationary covariance between the two proteins is marked for reference.

Positive regulation between any two genes of a network not only induces positive correlations between their respective proteins but also gives a characteristic shape to them, as seen in Figure 5.3, where the characteristic peak Corr^* is due to the type of the regulatory mechanism which in this case is that of activation. Corr^* and τ^* are the defining *features* of the correlation functions, which are functions in the rate constants and hence are sensitive in them and also sensitive in the type of regulatory mechanism. In the present case, $\mathrm{Corr}^* = 0.8$ and $\tau^* = 49$ minutes. Physically, this means that at steady-state, the deviations δP_y from the average protein concentration $\langle P_y \rangle$ are influenced

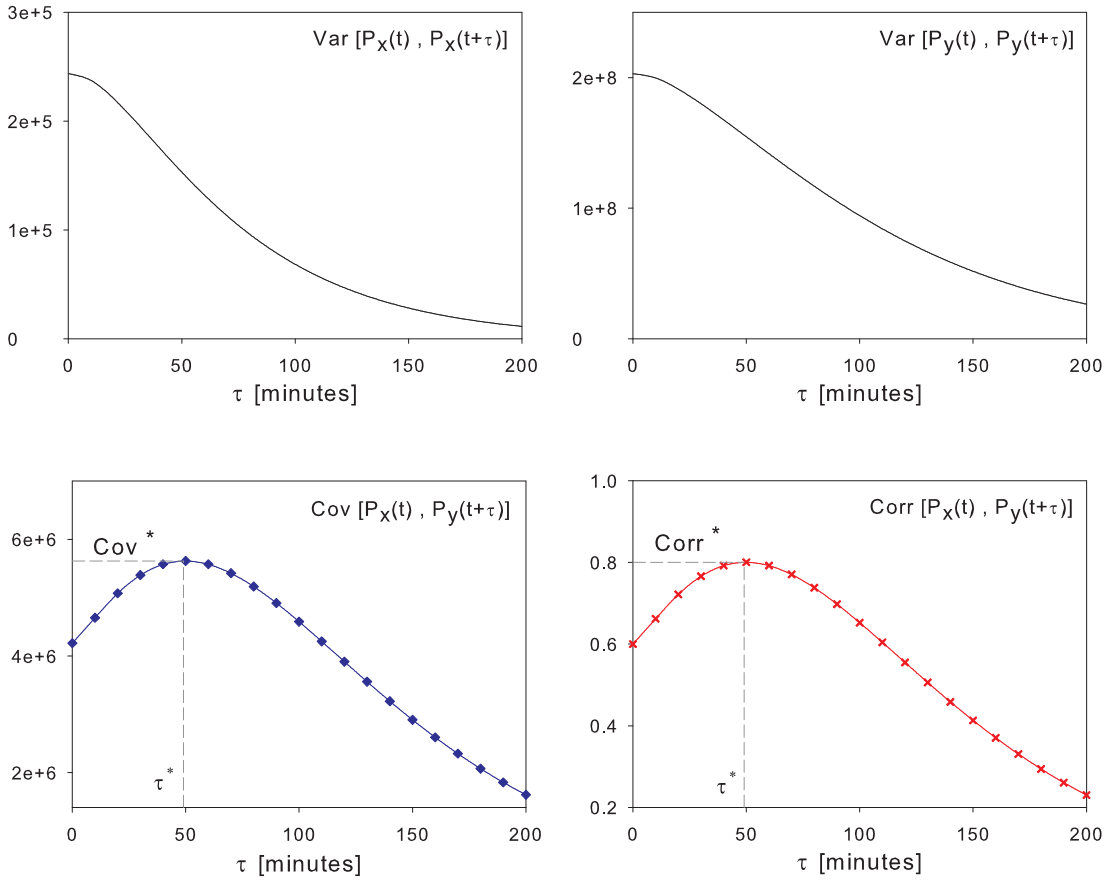

FIGURE 5.3: The dynamic autocovariances $\mathrm{Cov}[P_x(t), P_x(t+\tau)]$ and $\mathrm{Cov}[P_y(t), P_y(t+\tau)]$ are exponentially decreasing functions. Due to the stationary nature of the fluctuations these functions are time-independent and hence $\mathrm{Cov}[P_x(t), P_x(t+\tau)] \equiv \mathrm{Cov}[P_x(t_1), P_x(t_1+\tau)]$. The characteristic shape of the time-covariance function is preserved in the time-correlation function due to the normalization of the time-covariances by stationary auto-covariance values as done in equation (3.15).

mostly by δP_x occuring 49 minutes earlier to it. This time delay is the obvious result of the *causal* link between the two genes. On the other hand, the stationary correlation (at $\tau = 0$), which is of non-zero value (0.6) would ideally not indicate of any such *causality* and hence the significance of the dynamic correlations. This is supported by the simple fact that the $\mathrm{Corr}[P_y(t), P_x(t+\tau)]$ function has the same non-zero value at time $\tau = 0$, while for $\tau > 0$, is a monotonically decreasing function. Hence, stationary correlation alone cannot in effect predict the direction of the *causal* link between the genes. Ideally it would not make much sense for the stationary correlation to have a non-zero value, since that would mean that δP_x would have an instantaneous effect on δP_y. However the actual reason behind the non-zero value of the stationary correlations is the slow decaying auto-correlation function of P_x. This dynamic auto-correlation is essentially due to the decay process of the protein. Therefore, if δP_y at say time (t) is influenced by δP_x at say time $(t - t_1)$, the stationary correlation between the two proteins at time (t) could still be non-zero if the auto-correlation function of δP_x is non-zero for a time-period $(> t_1)$, which is what is happening here. Therefore, the stationary correlation

between the two proteins at some time t is non-zero and positive and is in all likelihood the result of the causal link between them, but this causality cannot be established by this stationary correlation alone.

5.1.2 Eliminating fast reactions

Reacting systems such as gene regulatory networks are usually composed of elementary reactions that can be broadly classified based on their operating timescales. For example, elementary reactions such as the binding and unbinding of the TFs to the upstream sequences of the regulated gene can be categorized as fast reactions whose timescales are in seconds. These reactions are fast in comparison to the other elementary reactions such as the transcription, translation and degradation that are in the range of minutes. For simulation purposes it is obvious that the elimination of these fast reactions would not only speed up the numerical simulations but also simplify the representation of the reaction set. The speed up in the simulations is obviously due to the elimination of reactions that occur thousands of times more frequently than the slower reactions during the same period of simulation time. One way of elimination is to simply neglect the fast reactions occuring at equilibrium and substitute equivalent expressions for the rate constants in the slower reactions. These equivalent expressions would then be of the Michaelis-Menten form with appropriate Hill co-efficients. Bundschuh et al. (2003) employ such simple procedures in the case of three sample regulatory networks and demonstrate that the stationary variances of a molecular species such as proteins would remain very much the same as for the case with the original set of reactions comprising the fast reactions. On similar lines we would like to show here that on eliminating the fast reactions of binding and unbinding $\left[P_x + G_y \underset{k_{off}}{\overset{k_{on}}{\rightleftharpoons}} \mathcal{C}_y\right]$, the dynamic correlation function between the proteins P_x and P_y remain the same. The reduced system now has the same reaction set as in Table 5.1 but without the above binding/unbinding reactions and the $\left[\mathcal{C}_y \xrightarrow{k^+_{M_y}} \mathcal{C}_y + M_y\right]$ being substituted for by the new elementary reaction of $\left[\phi \xrightarrow{k_{eff}} M_y\right]$, where k_{eff} is the effective rate constant for this slow reaction which is $k_{eff} = k^+_{M_y}\left(\frac{\alpha P_x}{K_D + P_x}\right)$ and follows the Michaelis-Menten kinetics. k_{eff} is obtained by the simple fact that at steady state the mean concentration of the bound complex $\mathcal{C}_y$ of the regulated gene is of the same MM-form. Now, the reduced system of reactions, whose timescales are comparable to each other, result in the following deterministic rate

equations,

$$\begin{aligned}
\frac{dM_x}{dt} &= k^+_{M_x} - k^-_{M_x} M_x \\
\frac{dM_y}{dt} &= k_{eff} P_x - k^-_{M_y} M_y \\
\frac{dP_x}{dt} &= k^+_{M_x} M_x - k^-_{P_x} P_x \\
\frac{dP_y}{dt} &= k^+_{P_y} M_y - k^-_{P_y} P_y
\end{aligned}$$

whose Jacobian is now a 4×4 matrix,

$$\mathbf{A} = \begin{pmatrix} -k^-_{M_x} & 0 & 0 & 0 \\ 0 & -k^-_{M_y} & \left(\frac{\alpha k^+_{M_y} K_D}{(K_D+P_x)^2}\right) & 0 \\ k^+_{P_x} & 0 & -k^-_{P_x} & 0 \\ 0 & k^+_{P_y} & 0 & -k^-_{P_y} \end{pmatrix}.$$

The usual procedure for obtaining the dynamic correlation functions is followed as in the case of the original system and as expected, the correlations are exactly of the same values as shown in Figure 5.3 for the original case. Such effective simplification is helpful not only for simulation purposes but also for ease in representation. Suppose a regulatory network is made up of n number of activators, each activator system comprising of 3 variables which are the proteins, mRNAs and the gene. The total number of variables would then be $3n$ which could easily be reduced to $2n$ by employing the above reduction technique, whilst the dynamic correlations remain effectively the same.

5.1.3 Deterministic response to perturbations

The relation between the internal fluctuations and the system's deterministic response to perturbation commonly referred to as fluctuation-dissipation relations was described earlier in chapter 3 with the aid of a sample system of a single-gene. Here, we shall extend the investigation of fluctuation-dissipation relations in the case of the elementary activator system. To get a clear picture let us represent the network structure of the elementary activator simply as $\phi \rightarrow M_x \rightarrow P_x \rightarrow M_y \rightarrow P_y$ instead of the schematic of Figure 5.1. Firstly, the deterministic response, which is the response in the average or mean values of the molecular species for external perturbation, is determined. The external perturbation in this case is simulated by instantaneously increasing the mean steady state value of the mRNA M_x represented as $\langle M_x(t_{ss})\rangle$ to $\langle M_x(t_{ss})\rangle + \Delta M_x$. Due to the above network structure of the system, the responses in the downstream molecular species of P_x, M_y and P_y are as shown on the right side of Figure 5.4. The time at which these responses attain maximum value are labelled as t_{resp}, which is $36, 53, 104$ minutes for $P_x(t), M_y(t)$ and $P_y(t)$ respectively. In Appendix B we derive the

analytical expressions for t_{resp} with the suitable approximations where decay rates of proteins are much smaller than that of the mRNAs. Though this approximation might not be perfectly true for the decay rates in the case of the $CHA4 \rightarrow CHA1$ system, for many other genes where the half-lives of proteins are of the order of 60-120 minutes and the mRNAs are known to be in the range of a few minutes (around 4-10), these approximations hold true. However, the qualitative relation between the expressions of t_{resp} and τ^* is what we are really interested in, and not the exact expressions themselves.

The fluctuation-dissipation theorem says that the source behind these deterministic responses is also responsible for the internal fluctuations. While the deterministic dynamics are in effect described by the rate equations and the equivalent Jacobian matrix, the covariances that represent the internal fluctuations are also described by the same rate equations and Jacobian. This was shown in chapter 3 and that the FDT holds true in such chemical reacting systems. The comparison between the response curves and the dynamic correlations is only for confirmation. Internal fluctuations as represented by the dynamic correlations between the molecular species have similar characteristics as the deterministic response curves. The corresponding time delay for attaining maximum correlation represented by τ^* are functions in the same decay rates as in the case of t_{resp}. For comparison with the response curves, the dynamic correlations between M_x and the other downstream species are plotted on the left of Figure 5.4. These are correlations between M_x at time t and the downstream species at a time-delay of $t + \tau$ and have the characteristic non-monotonic shape with $\tau^* \neq 0$. In the correlations between the pairs of (M_x, P_x), (M_x, M_y) and (M_x, P_y) respectively, the τ^* corresponding to each of the three correlation functions, are of the value of 21, 36 and 81 minutes respectively. The analytical expressions for these times are also derived in Appendix B.

The progressively higher values of τ^* are due to the progressive inclusion of additional reaction steps that correspond to the production of P_x, M_y and P_y. This progressive shift of τ^* is due to the decay rates of the additional species P_x, M_y and P_y. Similarly in the dynamic correlation $\mathrm{Corr}[P_x(t), M_y(t + \tau)]$ the corresponding $\tau^* = 11$ minutes is the smallest of the τ^*'s of all the possible correlation functions out of the four molecular species. This is due to the proximity of the species P_x and M_y within the network structure and also due to the half-life of M_y being 10 minutes which is the smallest among that of all the other species. It is therefore evident that the reaction *structure* and the values taken on by the parameters have a significant influence on the time-correlation functions. More importantly the regulatory activity, which in this case is that of activation, is clearly reflected in these plots that have a non-monotonic shape. Coming back to the times t_{resp} and τ^*, the difference between the two is due to the fact that $\langle M_x(t_{ss}) \rangle$ after being perturbated was allowed to go back to its original state by decaying freely. This meant that at every infinitesimal time instant after the perturbation M_x was at a perturbed value, thereby sustaining the response in the dowstream species for a longer duration. In the case of the response of P_x, the duration for which the

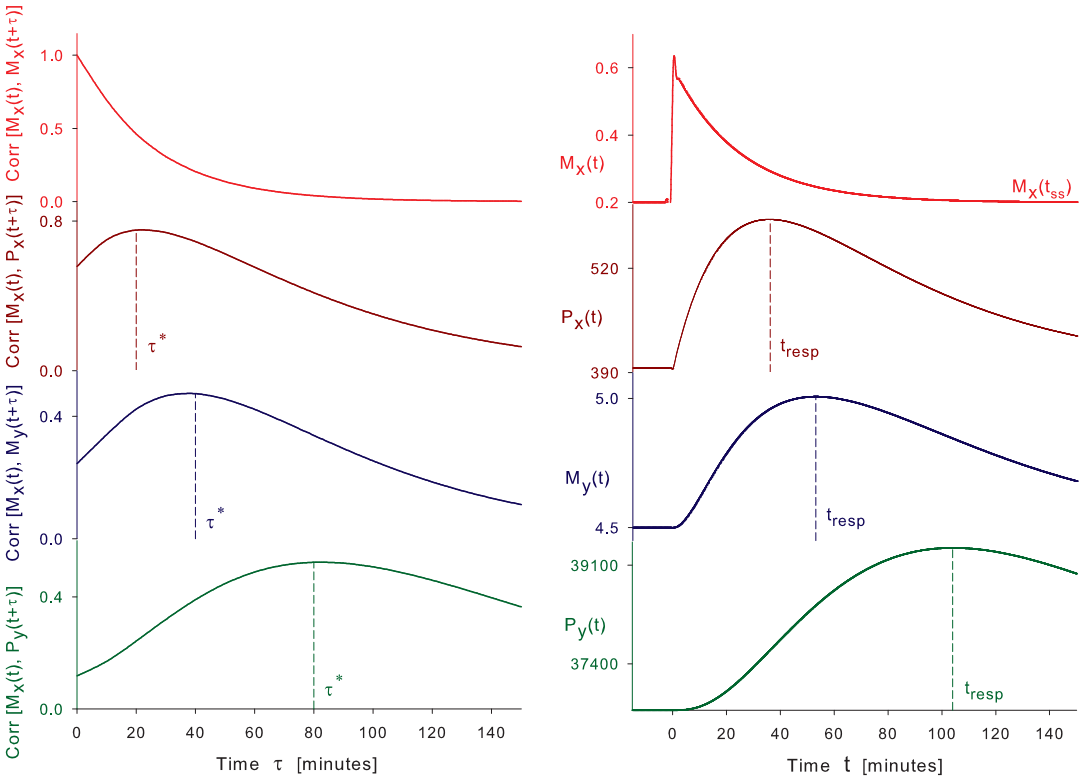

FIGURE 5.4: To the left are the dynamic correlation functions between M_x and the other variables of the $X \rightarrow Y$ network and to the right are the deterministic responses in the other variables for perturbation in the mean steady state value of M_x.

decay in M_x has a perturbation type of effect on P_x should be intuitively around the value of the half-life of M_x which is $\left(21^{-1} + 90^{-1}\right)^{-1} \approx 17$ minutes, while the actual difference $(t_{resp} - \tau^*)$ is around $36 - 21 = 15$ minutes. The reasoning is on similar lines for the difference between t_{resp} and τ^* values in the case of M_y and P_y. The difference in these time-delays would have been close to zero had the perturbation in M_x lasted only for a very short time-interval after which it is instantaneously forced back to its original steady state value. Through the above exercise we have demonstrated that the fluctuation-dissipation relation is a reliable guide in such biological systems.

5.1.4 Effect of parameters on stationary correlations

The stationary covariance between molecular species is a function of various parameters of the system. These parameters being nothing but the reaction rate constants, the stationary covariances could easily be controlled by varying the values of these rate constants. Thattai and van Oudenaarden (2001) showed this in the case of a simple model comprising of only mRNA and protein species. They derived a simple expression for the stationary variance of the proteins in terms of the parameters as $\left(\frac{\langle P \rangle . k_P^+ / k_M^-}{1 + k_P^- / k_M^-}\right)$. As a continuation to their work, Ozbudak et al. (2002) experimentally demonstrated the validity of this expression by altering the transcription and translation rates and

then measuring the protein variance. In section 3.3 we derived the expressions for the stationary variances and covariances in the case of a model comprising of not only mRNA and proteins but also a gene. The additional parameters in this model were the ON/OFF rates of the gene. The expressions of the covariances in this case act as a clear guidance for experimentalists who are interested in altering the fluctuation behaviour of the system. Conversely, the change in the fluctuation behaviour or in other words the change in the covariances, due to variations in parameter values, will provide us with some knowledge with regard to the system at hand. Though inference of the analytical expression of covariances is too much to ask for, a basic understanding of the network structure, regulatory mechanism and the values of the parameters is obtained via such simple procedures. As our focus here is on the protein-protein stationary covariance $\mathrm{Cov}[P_x(t), P_y(t)]$, we derive the same by solving the Lyapunov Equation $\mathbf{AC} + \mathbf{CA}^T + \mathbf{BB}^T = \mathbf{0}$, for the case of the elementary activator system. Since the set of variables for this system are $[G_y, M_x, M_y, P_x, P_y]$, the element of the $\mathbf{C}$ matrix corresponding to $\mathrm{Cov}[P_x(t), P_y(t)]$ is C_{45}. Therefore,

$$\mathrm{Corr}[P_x(t), P_y(t)] = \frac{C_{45}}{\sqrt{C_{44}C_{55}}}$$

The analytical expression for the above correlation can be derived but is way too cumbersome to be reproduced here and also to be of any sensible use. We evaluate the numerical values instead. By varying each of the ten rate constants *individually*, over a vast range, we note the change in value of the above correlation. In Figure 5.5, the change in the correlation for changes in the decay rates is shown. The behaviour of the correlation for variations in the other six parameters is shown in Figure 5.6. Note that the stationary correlation is nothing but the dynamic correlation $\mathrm{Corr}[P_x(t), P_y(t+\tau)]$ at time $\tau = 0$. The parameters are varied from {0.05 times their *base value*} to {30× *base value*}, where the *base values* are those derived from the $CHA4 \rightarrow CHA1$ system.

Firstly, from Figure 5.5, we see that the correlations are smaller for lower values of the decay rates, *i.e.,* for higher half-life times. As the proteins/mRNAs decay slowly, the number of molecules present at any given time remain mostly constant and hence any fluctuations in P_x would not be prominently correlated with the fluctuations in P_y. On the other hand, for smaller half-lives or larger decay rates, due to the lesser number of protein/mRNA molecules present, any fluctuations in them are highly correlated. Also, from the expression for P_x at steady state, which is $\frac{k^+_{P_x} M_x}{k^-_{P_x}}$ and similarly from $P_y = \frac{k^+_{P_y} M_y}{k^-_{P_y}}$, it is obvious that for increase in the decay rates the mean levels of protein concentration decreases whilst from Figure 5.5 it is clear that the stationary correlation increases. The increase in the stationary correlation is greater in the range where the decay rates are smaller in value. A notable difference amongst the four decay rates is the way in which the correlation increases faster and later decreases for increasing $k^-_{M_x}$

and $k^-_{P_x}$. This could possibly point towards the network structure which in this case is $X \rightarrow Y$.

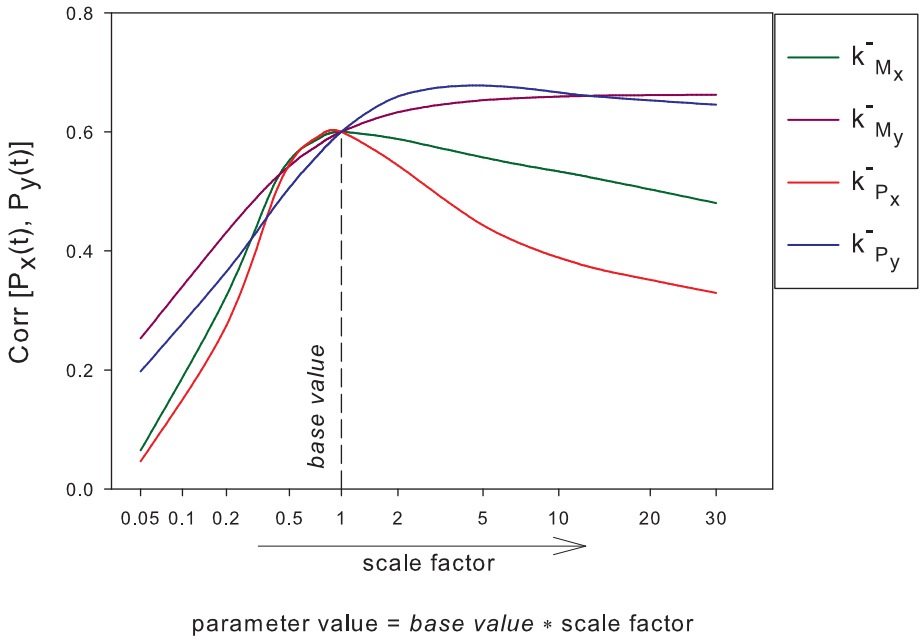

FIGURE 5.5: The above figure shows the variation in the stationary correlation $\text{Corr}[P_x(t), P_y(t)]$ (Figure 5.2(b)) for variations in the decay rates. The decay rates are varied *individually* whilst keeping the other parameters fixed at their base values. The range of the variation in the parameter value is from $\{0.05 \times base\ value\}$ to $\{30 \times base\ value\}$. The parameter values are on the x-axis, whose scale-type is changed to non-linear for ease in representating the plots.

Similar reasoning suffices for the variation in the correlation for changes in the other six parameters. In Figure 5.6, the values of the transcription rate, translation rate and the ON and OFF rates of the DNA are varied from the *base* value in steps. Firstly, for the case of the ON and OFF rates, for lower values of k_{ON} and for higher values of k_{OFF}, the stationary correlation is high. That is, for higher values of the dissociation constant, defined by k_{OFF}/k_{ON}, the transcription factor binds less to the DNA-complex thereby reducing the rate of production of M_y and inturn of P_y. However, the stationary correlation depicts a different picture, mainly due to the fact that the fluctuations in P_x are now less rapid and therefore correlate more with other species of the system. On the other hand, for higher values of the transcription rate, logic dictates that the stationary correlation between the proteins be high, which definitely is the case as seen in the Figure 5.6. Finally, translation rates have no effect on the stationary correlations since they donot directly contribute to the transcription process. Analyzing the effects of these rate constants on the fluctuation properties of the molecular species is important for two reasons. One is that the molecular fluctuaions as observed in the correlations provide significant information regarding the rate constants of the reaction processes. The other complementary reason is that, such an analyses act as analytical tools that

aid in designing new synthetic regulatory networks whose fluctuation properties and performance in general, could be controlled.

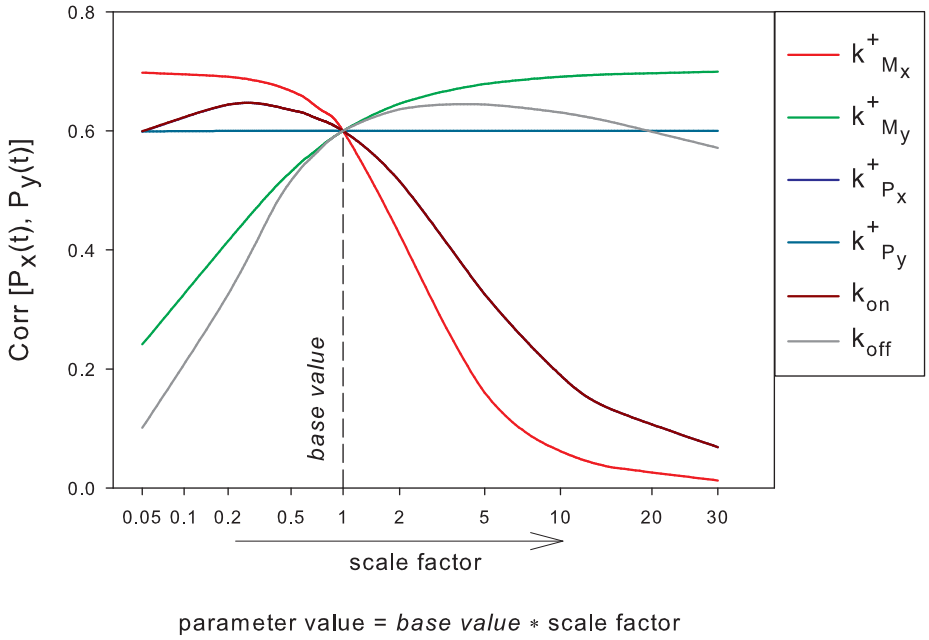

FIGURE 5.6: The above figure shows the variation in the stationary correlation $\text{Corr}[P_x(t), P_y(t)]$ for variations in the transcription rate, translation rates, binding/unbinding rates etc. The parameters are varied *individually* whilst keeping the other parameters fixed at their base values. Note: the correlations corresponding to variations in $k^+_{P_x}$ are nearly the same as the correlations corresponding to variations in k_{on}, hence the overlap in their plots.

5.1.5 Effect of parameters on dynamic correlations

Though the sensitivity of the stationary correlation for variations in the parameters is significant, it is the dynamic correlation that reveals the *causality* between the genes and is therefore a very important statistic. In Figures 5.7 to 5.11 the dynamic correlation between the proteins is plotted over a range of values for the four decay rates and also the dissociation constant K_D. The dynamic covariances given by equation (3.14) have the general form $\sum_{l=1}^{N} h_l e^{\lambda_l \tau}$, where the co-efficient (h's) of the exponentials are some functions in the eigenvectors of the Jacobian and λ's are the eigenvalues. In the case of GRNs involving two or three genes, the eigenvalues are mostly simple functions in the parameters of the system, whereas the elements of the eigenvectors and hence the h's are complicated functions in the parameters $\mathbf{p}$. It is therefore difficult to derive any explicit relation between $\mathbf{p}$ and the correlations. One way of getting around this difficulty is by simply studying the effect $p_\alpha \pm \triangle\, p_\alpha$ has on the magnitude and the temporal character of the correlations or more specifically on the features Corr^* and τ^*. The observed sensitivity of τ^* especially in the decay rates is re-confirmed by performing a sensitivity analysis as described in chapter 4 and whose results are tabulated in Table 5.3.

Elementary Activator						
Genetic Process	Parameter (p_α)	value	$\frac{\partial\,\mathrm{Corr}^*}{\partial p_\alpha} \approx$	Sensitivity	$\frac{\partial \tau^*}{\partial p_\alpha} \approx$	Sensitivity
Constant flux	$k^+_{M_x}$	8.14×10^{-3}	-27	yes / non-Linear	-10	no
Translation	$k^+_{P_x}$	36.3	-3×10^{-3}	no	-0.01	no
	$k^+_{P_y}$	143.2	5×10^{-7}	no	-0.002	no
Transcription	$k^+_{M_y}$	0.52	0.2	yes / Linear	-0.2	no
Decay process	$k^-_{M_x}$	4.07×10^{-2}	2.2	yes / non-Linear	-124	yes / non-Linear
	$k^-_{M_y}$	7.70×10^{-2}	-0.08	yes / non-Linear	-168	yes / non-Linear
	$k^-_{P_x}$	1.84×10^{-2}	6.1	yes / non-Linear	-302	yes / non-Linear
	$k^-_{P_y}$	1.76×10^{-2}	1.2	yes / non-Linear	-1410	yes / non-Linear
TF binding/unbinding	k_{on}	1	-0.1	yes / non-Linear	-0.08	no
	k_{off}	200.0	5×10^{-4}	yes / non-Linear	9×10^{-5}	no

TABLE 5.3: Sensitivities of $\mathrm{Corr}^*[P_x(t), P_y(t+\tau)]$ *w.r.t* each of the parameters of the elementary activator system. Though the expression for the partial derivative $\frac{\partial\,\mathrm{Corr}}{\partial p_\alpha}$ is applicable at any value of τ and not just at τ^*, here we evaluate $\frac{\partial\,\mathrm{Corr}}{\partial p_\alpha}$ only at $\tau^* = 49$ minutes, which is nothing but $\frac{\partial\,\mathrm{Corr}^*}{\partial p_\alpha}$. $\frac{\partial \tau^*}{\partial p_\alpha}$, which is the the sensitivity of τ^* *w.r.t* the parameters is also given. The type of sensitivity *i.e.,* the way in which the Corr^* and τ^* vary for slight variations in p_α, is mentioned. *Non-linearity* implies that the sensitivities are applicable only around small regions of p_α. The terms *yes* or *no* in the *sensitivity* column depend on whether the product of the parameter's value and its corresponding $\frac{\partial\,\mathrm{Corr}^*}{\partial p_\alpha}$ are significant or not.

The mathematical reasoning for the sensitivity of the feature τ^* *w.r.t* only the decay rates is as follows. While the constituent and non-redundant variables of the elementary activator are $[G_y, M_x, M_y, P_x, P_y]$, the *DNA-TF* complex $\mathcal{C}_y$ is the redundant variable. Thus, there are five eigenvalues for the deterministic system, of which three are the decay rates $[-k^-_{M_x}, -k^-_{M_y}, -k^-_{P_y}]$ whilst the remaining two are roots to the quadratic equation $x^2 + bx + c = 0$, where

$$\begin{aligned} b &= k_{off} + P_x k_{on} + G_y k_{on} + k^-_{P_x} \\ c &= k^-_{P_x}(k_{off} + k_{on} P_x) \end{aligned}$$

Considering the initial amount of G_y to be 1 nM, the expression for its mean concentration at steady state is $\langle G_y \rangle = \frac{K_D}{K_D + \langle P_x \rangle} = \frac{k_{off}}{k_{off} + k_{on} P_x}$. We conveniently omit the angled brackets $\langle\rangle$ for the rest of the section. Therefore, $k_{off} + k_{on} P_x = \frac{k_{off}}{G_y}$, which makes $b = \frac{k_{off}}{G_y} + k_{on} G_y + k^-_{P_x} = \left(G_y + \frac{K_D}{G_y}\right) k_{on} + k^-_{P_x}$. Since the dissociation constant K_D is usually of the order of hundreds of nM in concentration and whereas the mean steady state value of G_y is less than 1 nM, the approximation $b \approx \frac{k_{off}}{G_y} + k^-_{P_x}$ is valid. Similarly, c can be re-written as $\frac{k_{off} k^-_{P_x}}{G_y}$. The roots of the quadratic equation $x^2 + bx + c = 0$ with the approximated expressions for b and c are now $-k_{off}/G_y$ and $-k^-_{P_x}$. With this approximation, four out of the five eigenvalues are the decay rates of the mRNAs and proteins whilst the fifth eigenvalue, which is of high magnitude, is $-k_{off}/G_y$. Since the covariance function is a sum of exponentials that are raised to the power of these eigenvalues, the slow eigenvalues corresponding to the decay rates have a predominant effect. For lower (higher) values in these decay rates (half-life times) the exponentials decay slowly to zero. Now, since the covariance at each point in the τ-axis is a summation of these exponentials, the slow decaying exponential induces a delay in the covariance attaining its peak magnitude which appears as a shift in τ^* for the new covariances. This is the reason for the high sensitivity values of $\frac{\partial \tau^*}{\partial p_\alpha}$ for the four decay rates. This behaviour is observed in the correlations of the mRNAs or in fact between any two molecular species of the system.

Though all the four decay rates show up as the eigenvalues, only those connected with the element Y induce more sensitivity in τ^* as shown in Figures 5.9 and 5.10 and further verified from Table 5.3, where the terms $\frac{\partial \tau^*}{\partial p_\alpha}$ for the decay rates of $[M_x, M_y, P_x, P_y]$ are $[-124, -168, -302, -1410]$ respectively. As the decay rates (half-life) of P_y and M_y decrease (increase) the Corr* shifts in time, *i.e.,* τ^* increases with a simultaneous decrease in the amplitude of the correlation. As the half-life of say P_y increases, more of its molecules are present in the environment due to the continuation in its production at a constant rate. With more protein present in the medium at any given time, it becomes difficult to keep *track* of those protein molecules that came into existence due to the action of a *specific* set P_x molecules. Therefore the *relation* between the protein species is now *masked* or, in other words, the correlation between them is no more prominent

than was otherwise. Hence, for any small perturbation in upstream protein values at steady state the effect now seen in the downstream protein levels are lessened and appear after an increased time-delay of τ^*.

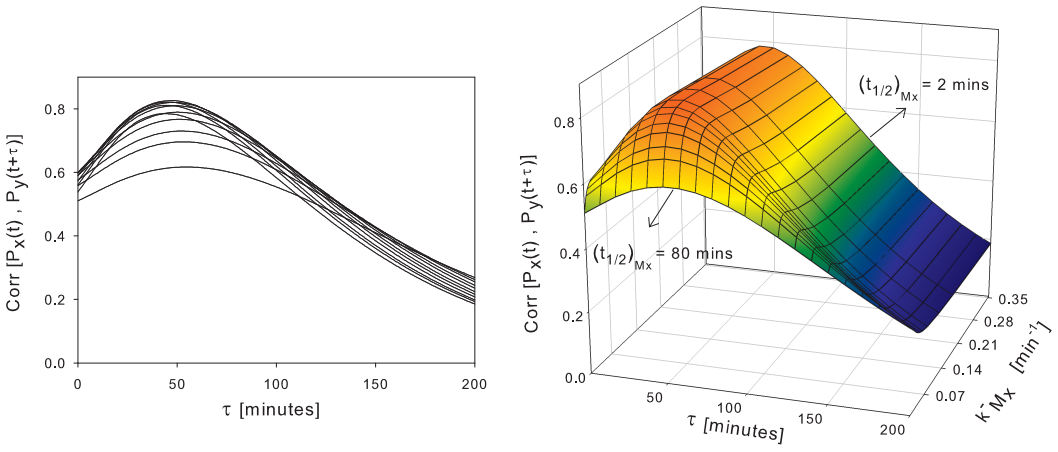

FIGURE 5.7: On the left is shown the time-correlation function between P_x and P_y for various values of the decay rate $k^-_{M_x}$ of the mRNA M_x. To vary $k^-_{M_x}$, the half-life $(t_{1/2})_{M_x}$ is assigned values of $2, 4, 6, 10, 14, 18, 24, 30, 40, 50, 80$ minutes. We notice that τ^* increases marginally from 43 minutes to 54 minutes for increase in $(t_{1/2})_{M_x}$. For higher values of $(t_{1/2})_{M_x}$, τ^* gets capped at around 54 minutes, due to the effect of the cell-doubling time t_{double} of equation (5.7). On the right is shown a 3 dimensional view of the effect of the decay rate on the correlations. The correlation plots corresponding to half-lives of 2 and 80 minutes are marked out explicitly. The relation between the decay rate and the half-life is given in equation (5.7).

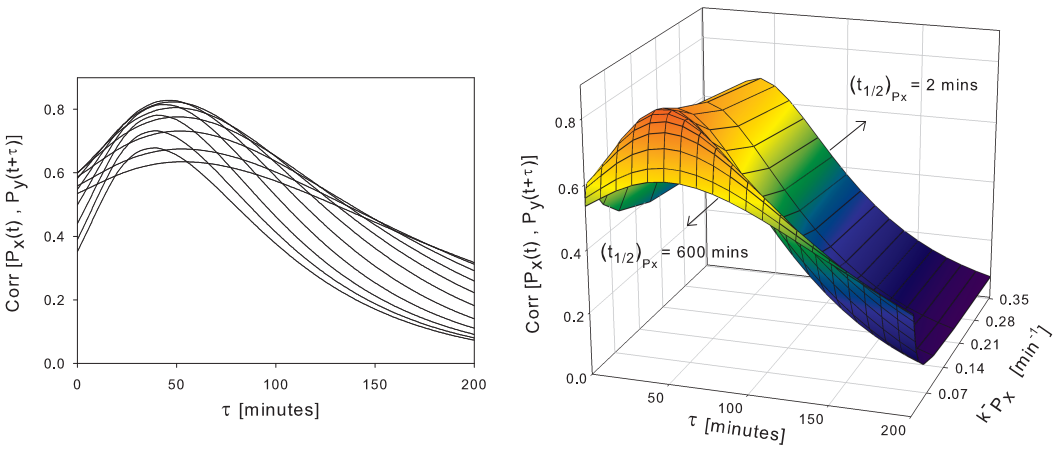

FIGURE 5.8: $k^-_{P_x}$ is varied by assigning the values of $2, 4, 8, 15, 25, 40, 60, 90, 150, 300, 600$ minutes to the half-life $(t_{1/2})_{P_x}$. In the Corr$[P_x(t), P_y(t+\tau)]$ function τ^* marginally increases from 37 minutes to 52 minutes at which point gets capped due to the effect of t_{double}.

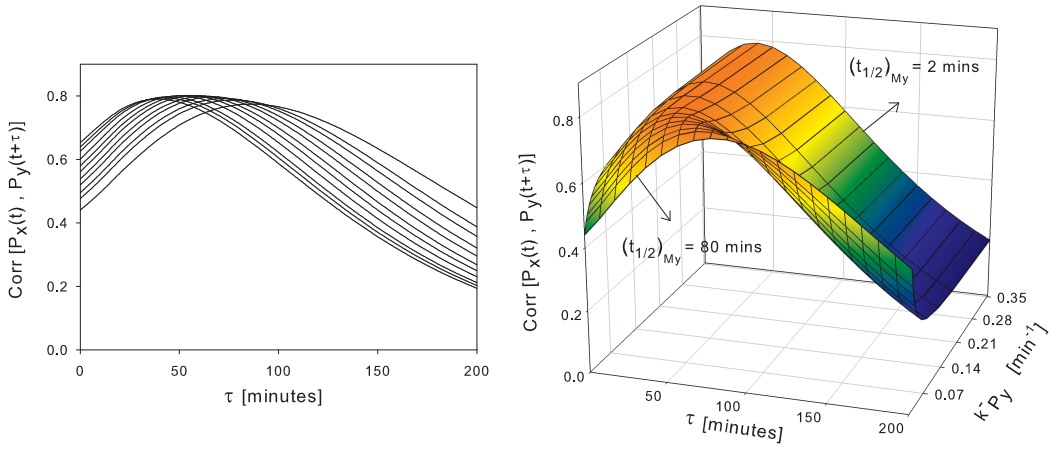

FIGURE 5.9: $k^{-}_{M_y}$ is varied by assigning the values of $2, 4, 6, 10, 14, 18, 24, 30, 40, 50, 80$ minutes to the half-life $(t_{1/2})_{M_y}$. In the Corr$[P_x(t), P_y(t+\tau)]$ function τ^* increases from 38 minutes to 85 minutes for increase in the mRNA half-life.

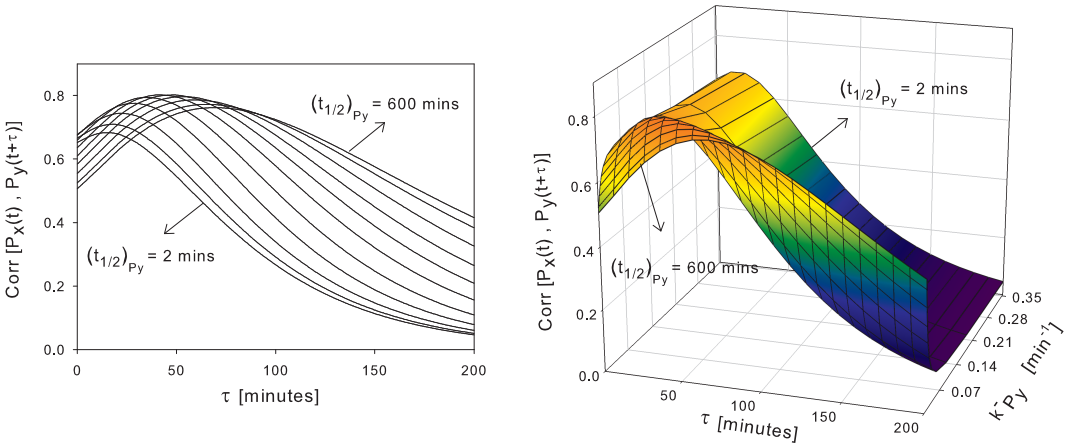

FIGURE 5.10: $k^{-}_{P_y}$ is varied by assigning the values of $2, 4, 8, 15, 25, 40, 60, 90, 150, 300, 600$ minutes to the half-life $(t_{1/2})_{P_y}$. In the Corr$[P_x(t), P_y(t+\tau)]$ function τ^* progressively increases from 14 minutes to 68 minutes for increase in the protein half-life.

Finally with regard to the dissociation constant K_D taking on a wide range of values and corresponding changes in the probability of the TF binding to the operator region, the average amount of mRNA and protein of the regulated element also vary greatly. However, interestingly the dynamic correlation between the TF protein P_x and the regulated protein P_y remains unchanged and nearly the same as that of the *base case*, except for K_D values less than around 200 nM when there is progressive reduction only in Corr* while τ^* still remains stagnant at its base value of 49 minutes. The insensitivity of τ^* to the binding/unbinding rates is due to the nature of the binding/unbinding reactions which are very much faster that the other elementary reactions. This issue was discussed in detail in section 5.1.2. Further, the correlations are totally insensitive

to cases where k_{on} and k_{off} take on different values such that their ratio $K_D = \frac{k_{off}}{k_{on}}$ remains constant.

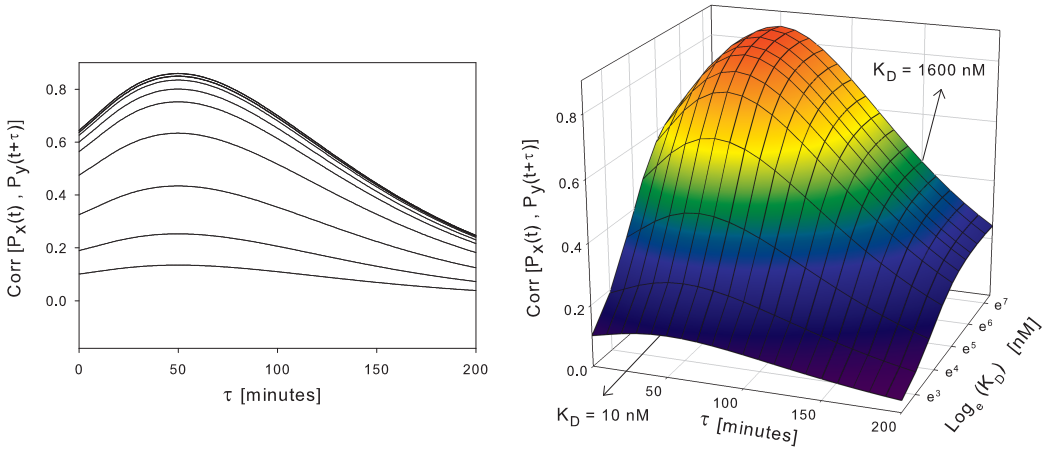

FIGURE 5.11: The dynamic correlation is evaluated for different values of the dissociation constant $K_D = k_{off}/k_{on}$. This dissociation constant is assigned values of $10, 20, 40, 80, 140, 200, 300, 400, 600, 1000, 1600\ nM$, by either increasing k_{off} or by decreasing k_{on}.

5.1.6 Effect of parameters on dynamic correlations - with mean steady state values fixed

On solving for the deterministic rate equations at steady state, the expressions for the mean values of the species are obtained in terms of the rate constants, indicating that these rate constants would impact both the deterministic as well as the fluctuation properties at steady state. For the time-independent case of fluctuations, Thattai and van Oudenaarden (2001) demonstrated that some rate constants exert influence simultaneously on the mean protein levels and also on their variances. In this section we study the effect of parameters on just the fluctuations, with the mean concentration levels held constant at the values given in Table 5.2. For example, supposing that the pair $[k^+_{M_x}, k^-_{M_x}]$ are at their base values, any deviations of equal measures in them does not affect $\langle M_x \rangle = k^+_{M_x}/k^-_{M_x}$. However there is a profound effect on the correlations for equal variations in $[k^+_{M_x}, k^-_{M_x}]$ as seen in Figure 5.12(a). The sensitivity of the dynamic correlations to changes in the parameter values is suggestive of possible ways in which biological networks could alter their fluctuation properties without affecting the mean concentration levels of proteins and mRNAs. In Figure 5.12(a) we show the variation in the correlations between the proteins for equal changes in the values of the parameters $[k^+_{M_x}, k^-_{M_x}]$. On comparison with Figure 5.7, where only the decay rate $k^-_{M_x}$ is varied, we notice that the change in the correlations is larger due to variation in $k^+_{M_x}$. It also suggests that this parameter only influences the magnitude of correlations and not its temporal character.

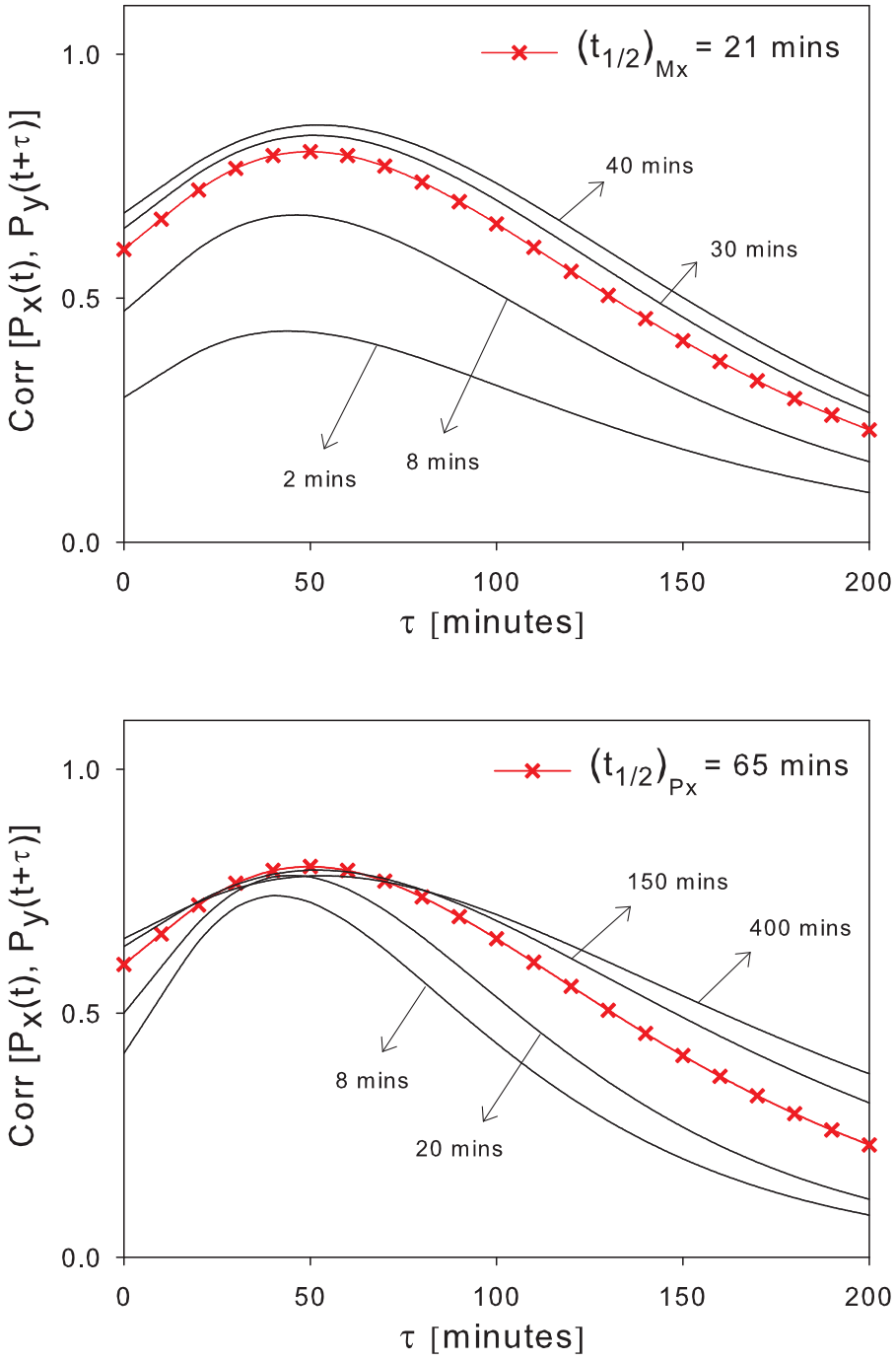

FIGURE 5.12: (a) The dynamic correlation between P_x and P_y is sensitive to changes in the production and decay rates of M_x. The correlation with all the parameters at their base values is in red, whereas the correlations that are due to changes in $[k^+_{M_x}, k^-_{M_x}]$ so that $\langle M_x \rangle = \frac{k^+_{M_x}}{k^-_{M_x}}$ remains constant, is in black. The other eight parameters are fixed at their base values. While Corr* decreases significantly for increase in the two parameters, τ^* registers miminal change from 56 to 46 minutes. (b) The translation and protein decay rates $k^+_{P_x}$ and $k^-_{P_x}$ are varied simultaneously so that $\langle P_x \rangle = \frac{\langle M_x \rangle k^+_{P_x}}{k^-_{P_x}}$ remains constant. While there is not much difference in Corr*, τ^* decreases from 55 to 45 minutes for increase in these parameters.

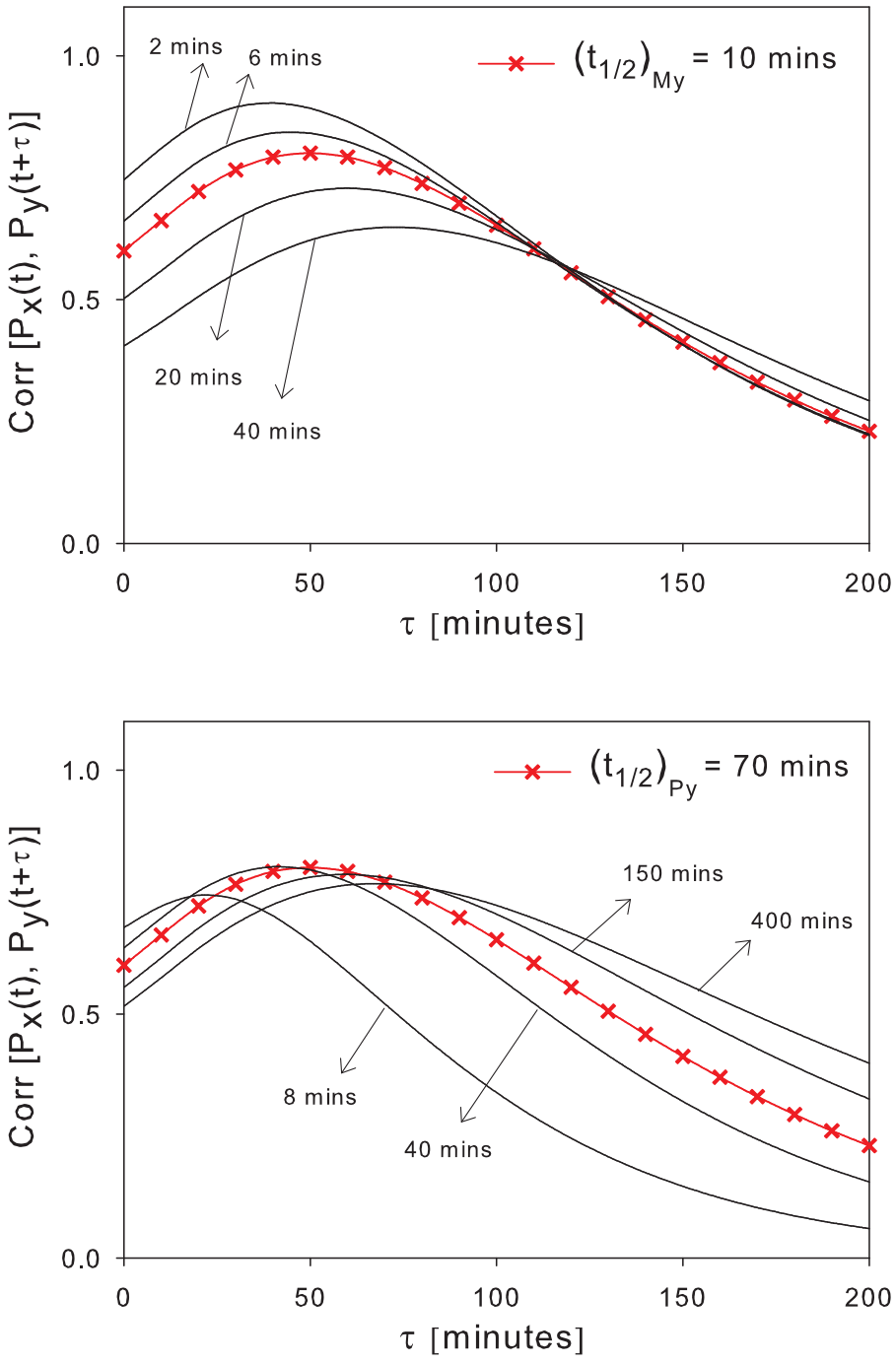

FIGURE 5.13: (a) Transcription rate constant $k^+_{M_y}$, and $k^-_{M_y}$ are varied simultaneously in steps, such that $\langle M_y \rangle$ remains constant. While Corr* does change by an appreciable amount, there is a huge decrease in τ^*, from 87 to 43 minutes. (b) Similar is the case for variations in $k^+_{P_y}$ and $k^-_{P_y}$. There is again a large reduction in τ^*, from 102 to 36 minutes accompanied by noticable change in the *shape* of the plots, as was in the case of $[k^+_{P_x}, k^-_{P_x}]$. The observed variation in τ^* is due to the decay rate rather than the translation rate. This is confirmed from the sensitivities of Table 5.3, where $\frac{\partial \tau^*}{\partial k^-_{P_y}} = -1410$, whereas $\frac{\partial \tau^*}{\partial k^+_{P_y}} = -0.002$.

Similarly, for simultaneous variations in $[k^+_{M_y}, k^-_{M_y}]$ the effect on Corr* of the protein correlations is more (Figure 5.13(a)), whilst in the case where only $k^-_{M_y}$ is varied Corr* just shifts in time (Figure 5.9). Therefore, the rate of production of the mRNAs have an effect on the correlations, under conditions where the mean levels of species are constant. On the other hand, the translation rates do not have such a prominent effect.

5.2 Effect of Dimerization

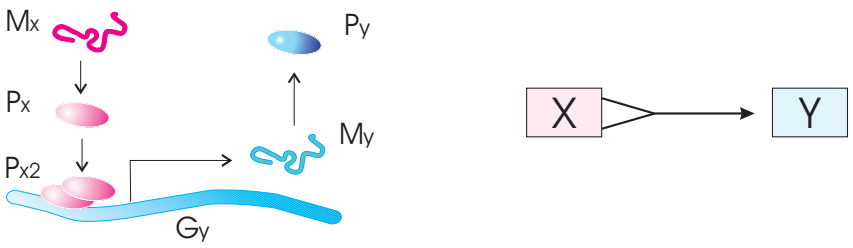

FIGURE 5.14: On the left is a schematic representation of the transcription of M_y from the dimer (P_{x_2}) bound gene complex. The equivalent network representation is shown to the right.

In the previous section on the elementary activator, we studied in detail how changes in parameter values causes different behaviour in the dynamic correlations. These correlations show different behaviours also for changes in the type of regulatory mechanism. In this section the mechanism is of activation via dimerization. Often TFs act in co-operation for activating or repressing the transcription of genes. Dimerization is the basic co-operativite mechanism where the protein monomers associate to form dimers, which then bind to the upstream region of the promoters to influence transcription. Incorporating the dimerization reactions into the model of the elementary activator, the steps that describe the transcription process now are: As in the case of activation through

$\left[\phi \xrightarrow{k^+_{M_x}} M_x\right]$	$\left[2P_x \underset{k_b}{\overset{k_a}{\rightleftharpoons}} P_{x_2}\right]$	$\left[\mathcal{C}_y \xrightarrow{k^+_{M_y}} \mathcal{C}_y + M_y\right]$	$\left[M_x, M_y \xrightarrow{k^-_{M_x}, k^-_{M_y}} \phi\right]$
$\left[M_x \xrightarrow{k^+_{P_x}} M_x + P_x\right]$	$\left[P_{x_2} + G_y \underset{k_{off}}{\overset{k_{on}}{\rightleftharpoons}} \mathcal{C}_y\right]$	$\left[M_y \xrightarrow{k^+_{P_y}} M_y + P_y\right]$	$\left[P_x, P_y \xrightarrow{k^-_{P_x}, k^-_{P_y}} \phi\right]$

TABLE 5.4: Reaction set describing the process of activation via protein dimerization between 2 genes. The dissociation constant of the dimerization process is $K_{Dim} = \frac{k_b}{k_a}$ whose value in this case is set equal to the value of the gene dissociation constant $K_D = 200\ nM$.

non-cooperativity, the redundant variable is chosen as $\mathcal{C}_y$, while the non-redundant variables are now $[G_y, M_x, M_y, P_x, P_{x_2}, P_y]$ and the corresponding rate equations describing

the time evolution of the deterministic or mean concentration levels of the variables are,

$$\begin{aligned}
\frac{dG_y}{dt} &= k_{off}(\alpha - G_y) - k_{on}G_yP_{x_2} \\
\frac{dM_x}{dt} &= k^+_{M_x} - k^-_{M_x}M_x \\
\frac{dM_y}{dt} &= k^+_{M_y}(\alpha - G_y) - k^-_{M_x}M_x \\
\frac{dP_x}{dt} &= k^+_{P_x}M_x - k^-_{P_x}P_x + 2k_bP_{x_2} - 2k_aP_x^2 \\
\frac{dP_{x_2}}{dt} &= k_{off}(\alpha - G_y) - k_{on}G_yP_{x_2} - k_bP_{x_2} + k_aP_x^2 \\
\frac{dP_y}{dt} &= k^+_{P_y}M_y - k^-_{P_y}P_y.
\end{aligned}$$

The vector of the deterministic rates, the stoichiometric matrix and the Jacobian are all derived from the above rate eqautions and are then used to obtain the dynamic correlation between P_x and P_y as described in the case of the elementary activator system.

$$\begin{aligned}
\mathbf{R} = \Big(&k_{on}G_yP_{x_2}, k_{off}(\alpha - G_y), k^+_{M_y}(\alpha - G_y), k_bP_{x_2}, k_aP_x^2, k^+_{M_x}, k^+_{P_x}M_x, k^+_{P_y}M_y, \\
&k^-_{M_x}M_x, k^-_{M_y}M_y, k^-_{P_x}P_x, k^-_{P_y}P_y\Big)^T
\end{aligned}$$

$$\boldsymbol{\nu} = \begin{pmatrix}
-1 & +1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & +1 & 0 & 0 & -1 & 0 & 0 & 0 \\
0 & 0 & +1 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 \\
0 & 0 & 0 & +2 & -2 & 0 & +1 & 0 & 0 & 0 & -1 & 0 \\
-1 & +1 & 0 & -1 & +1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & +1 & 0 & 0 & 0 & -1
\end{pmatrix}$$

$$\mathbf{A} = \begin{pmatrix}
(-k_{off} - k_{on}P_{x_2}) & 0 & 0 & 0 & -k_{on}G_y & 0 \\
0 & -k^-_{M_x} & 0 & 0 & 0 & 0 \\
-k^+_{M_y} & 0 & -k^-_{M_y} & 0 & 0 & 0 \\
0 & k^+_{P_x} & 0 & (-4k_aP_x - k^-_{P_x}) & 2k_b & 0 \\
(-k_{off} - k_{on}P_{x_2}) & 0 & 0 & 2k_aP_x & (-k_b - k_{on}G_y) & 0 \\
0 & 0 & k^+_{P_y} & 0 & 0 & -k^-_{P_y}
\end{pmatrix}.$$

A value of 200 nM is set for the dimer dissociation constant $K_{Dim} = k_b/k_a$ so that the mean concentration of the dimers $P_{x_2} = P_x^2/K_{Dim}$ is 783 nM which is approximately twice that of the monomers $P_x = 395$ nM. The dynamic correlation between P_x and P_y are evaluated for these parameter values and are as shown in Figure 5.15 (blue curves). The dimerization process clearly alters the shape of the correlations from that

of the elementary activator (red curves), such that the correlations are now nearly equal to 1 and decays very slowly. The increase in correlation magnitude is because, the fluctuations in P_x are now amplified by the equilibrium expression of the dimer which is proportional to P_x squared and it is this dimer that then activates the production of P_y. Therefore fluctuations in P_x have an amplifying influence on the fluctuations of P_y and hence the increase in the correlation magnitude. As for the slow decay of these correlations, it is once again the coupled reactions of dimerization that do not *forget* the fluctuations in P_x for a long time and hence the influence on P_y fluctuations is *sustained*. In other words, this is due to the coupled nature of the dimerization reactions, which forces the P_x fluctuations to influence P_{x_2} fluctuations which influence the P_x back again and so on, thereby sustaining the correlations. In Figure 5.15(a) we see that for short half-lives of P_x (about 8 minutes), the effect of dimerization begins to wear off, the possible reason being that P_x molecules are now being produced and degraded faster (due to increase in values of $[k^+_{P_x}, k^-_{P_x}]$) and consequently not able to participate much in the dimerization process. On the other hand, in Figure 5.15(b) we do not see any such effect for decrease in the half-life of P_y due to its non-participation in the dimerization process.

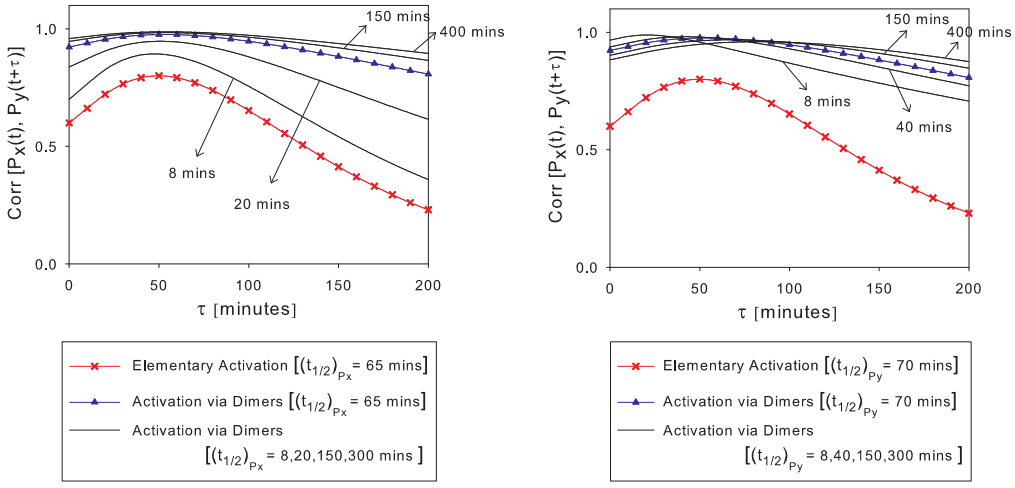

FIGURE 5.15: (a) $k^-_{P_x}$ is varied in steps from its base value of 0.0184 min^{-1} that corresponds to a half-life of $(t_{1/2})_{P_x} = 65$ minutes. The correlation corresponding to this value of $k^-_{P_x}$ is shown in blue. $k^+_{P_x}$ is simultaneously varied such that $\langle P_x \rangle$ remains unchanged. For increase in the values of $[k^+_{P_x}, k^-_{P_x}]$, there is an observable increase in Corr*, whilst τ^* decreases from 53 to 46 minutes. (b) Likewise, for step-wise increase in the parameters $[k^+_{P_y}, k^-_{P_y}]$, τ^* once again reduces from 81 to 20 minutes, whilst Corr* is nearly constant.

The effect of different values of K_{Dim} on the protein correlations is shown in Figure 5.16. For each value that K_{Dim} takes, the transcription rate $k^+_{M_y} = k^-_{M_y} M_y / \mathcal{C}_y$ takes on corresponding values since the amount of gene present in active state has the form $\mathcal{C}_y = \frac{\alpha P_{x_2}}{K_D + P_{x_2}} = \frac{\alpha P_x^2}{K_D K_{Dim} + P_x^2}$. Therefore Figure 5.16 shows the resultant effect of the pair

$[K_{Dim}, k^+_{M_y}]$. The overall effect of dimerization is that there is a higher level of correlation between the proteins and more importantly, the *shape* of the dynamic correlations is different than that in the monomer case. In fact, for a low value of $K_{Dim} = 40\ nM$, corresponding to $P_{x_2} = 3913\ nM$, the correlation function loses its characteristic *bell*-shape and its magnitude remains nearly constant at around 0.98 for a long duration of time. Under such conditions, any perturbation in P_x has an equivalent effect on P_y over a long period of time, *i.e.,* the effect of correlation is sustained over time. Further, *(a)* only the dissociation constants K_D and K_{Dim} affect the correlations and not the corresponding individual forward and reverse rates, *(b)* though $\mathrm{Cov}[P_{x_2}(t), P_y(t+\tau)]$ and $\mathrm{Cov}[P_x(t), P_y(t+\tau)]$ exhibit distinct behaviours, their normalized dynamic correlations are exactly the same, due to normalization by their respective stationary auto-covariances. Therefore either P_x or P_{x_2} could be considered as the output variable.

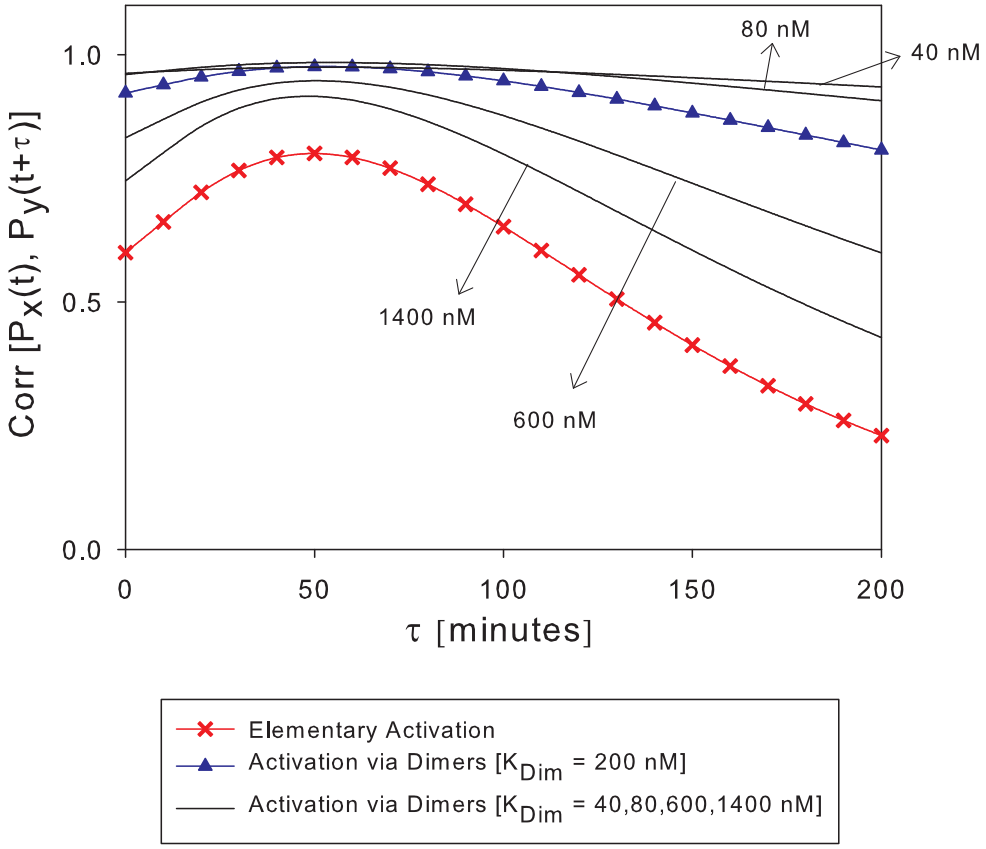

FIGURE 5.16: Activation by dimers: The base value of K_{Dim} is 200 nM and the correlation corresponding to this value is plotted in blue. Black curves represent the correlations that correspond to variations in the value of K_{Dim}, which is assigned values of $40, 80, 600$ and $1400\ nM$ [by adjusting either k_b and/or k_a so that their ratio is K_{Dim} as desired]. An equivalent variation in the value of $k^+_{M_y}$ is done such that the mean value of M_y remains unchanged from its original value. For increase in the values of $[k^+_{M_y}, K_{Dim}]$, τ^* decreases marginally from 53 to 47 minutes. For purposes of comparison, the base case correlation corresponding to the elementary activator system is included and is shown in red.

5.2.1 Eliminating fast reactions

As in the case of the elementary activator system, coupled fast reactions in the dimerization system are the ones corresponding to the binding and unbinding events of the dimers to the regulatory sequences of the gene. However, an additional set of fast reactions are those of protein dimerization. On eliminating the binding/unbinding reactions as in the case of the elementary activator (section 5.1.2), an effective rate constant $k_{eff} = k^+_{M_y}\left(\frac{\alpha P_{x_2}}{K_D+P_{x_2}}\right)$ is introduced in the rate equation for M_y. The fast variable, which is G_y, is eliminated thus and the dynamic correlation between P_x and P_y for such a reduced system is found that gives very nearly the same dynamic behaviour as that of the original unreduced system.

$$\begin{aligned}
\frac{dM_x}{dt} &= k^+_{M_x} - k^-_{M_x}M_x \\
\frac{dM_y}{dt} &= k^+_{M_y}\left(\frac{\alpha P_{x_2}}{K_D + P_{x_2}}\right) - k^-_{M_y}M_y \\
\frac{dP_x}{dt} &= k^+_{M_x}M_x - k^-_{P_x}P_x \\
\frac{dP_{x_2}}{dt} &= -k_b P_{x_2} + k_a P_x^2 \\
\frac{dP_y}{dt} &= k^+_{P_y}M_y - k^-_{P_y}P_y
\end{aligned}$$

$$\mathbf{A} = \begin{pmatrix}
-k^-_{M_x} & 0 & 0 & 0 & 0 \\
0 & -k^-_{M_y} & 0 & \left(\frac{\alpha k^+_{M_y} K_D}{(K_D+P_{x_2})^2}\right) & 0 \\
k^+_{P_x} & 0 & (-4k_aP_x - k^-_{P_x}) & 2k_b & 0 \\
0 & 0 & 2k_aP_x & -k_b & 0 \\
0 & k^+_{P_y} & 0 & 0 & -k^-_{P_y}
\end{pmatrix}.$$

However, further elimination of the other pair of fast reactions that are of the dimerization process, does not result in the same correlation. To eliminate the dimer P_{x_2} and its corresponding dimerization reactions, the equilibrium value of $P_{x_2} = P_x^2/K_{Dim}$ is used in the first reduced system. The rate equations for this second reduced system are now,

$$\begin{aligned}
\frac{dM_x}{dt} &= k^+_{M_x} - k^-_{M_x}M_x \\
\frac{dM_y}{dt} &= k^+_{M_y}\left(\frac{\alpha P_x^2}{K_D K_{Dim} + P_x^2}\right) - k^-_{M_y}M_y \\
\frac{dP_x}{dt} &= k^+_{M_x}M_x - k^-_{P_x}P_x \\
\frac{dP_y}{dt} &= k^+_{P_y}M_y - k^-_{P_y}P_y
\end{aligned}$$

$$\mathbf{A} = \begin{pmatrix} -k^-_{M_x} & 0 & 0 & 0 \\ 0 & -k^-_{M_y} & \left(\frac{2\alpha k^+_{M_y} K_D K_{Dim} P_x}{(K_D K_{Dim} + P_x^2)^2}\right) & 0 \\ k^+_{P_x} & 0 & -k^-_{P_x} & 0 \\ 0 & k^+_{P_y} & 0 & -k^-_{P_y} \end{pmatrix}.$$

On reducing the first set of fast reactions, the stationary covariance between the proteins remains the same as in the original system at 5.04×10^8. However, this stationary covariance in the second reduced system decreases to a value of 5.12×10^6. Consequently, the dynamic correlation between P_x and P_y in this case is found to be of quite less value when compared to the original and the first reduced systems. This is shown in Figure 5.17, where the stationary correlations for the original and second reduced systems are 0.92 and 0.63 respectively. Therefore the dimer reactions hold the key in characterizing this regulatory system, which is reflected in the dynamic correlations between the proteins.

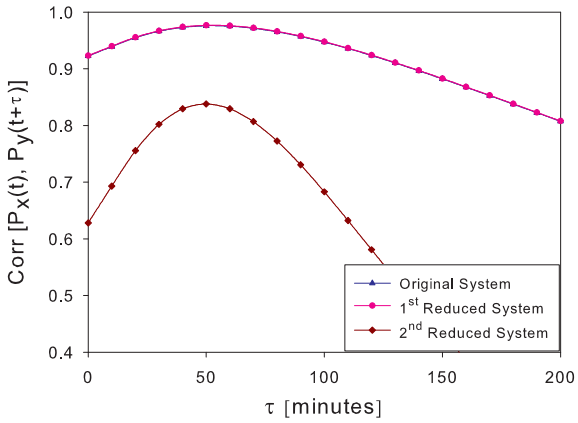

FIGURE 5.17: Correlations between proteins P_x and P_y of the dimer activator system, where two sets of fast reactions are eliminated.

On similar lines, Bundschuh et al. (2003) give an example case of a negative feedback gene where the second reduced system results in large *stationary variances* in the proteins. They reasoned that the *buffering* of fluctuations in monomer proteins was not possible in the second reduced system due to the absence of the dimerization process, since any large fluctuations in the total number of proteins $P_x + 2P_{x_2}$ results in smaller fluctuations in the population of the monomers in the presence of the dimerization reactions. Overestimation of the stationary fluctuations was a problem in their system due to the presence of a feedback mechanism, but in our system that has no such feedback, underestimation is the issue. Therefore, the dynamic correlations are underestimated in our second reduced system. Also, unlike the other works that studied the effect of dimerization, our focus here is on the dynamic correlations between proteins and not just on the stationary variances.

5.3 Elementary Repressor

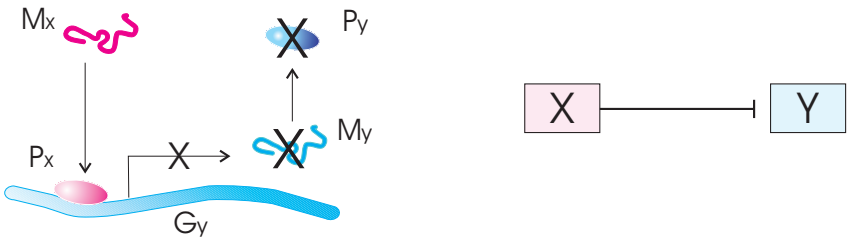

FIGURE 5.18: The figure on the left is a schematic representation of the regulatory process where the transcription of mRNA M_y from the coding sequence of its gene G_y is blocked by the action of the repressor P_x on the promoter region of the DNA. The equivalent network representation is shown to the right where the dashhead indicates repression of element Y by X.

Repression by a transcription factor is as common an event as activation. Protein P_x acting as a repressor of P_y binds upstream to the promoter region of G_y and blocks its transcription by the RNA polymerase molecule. However transcription continues at a basal rate of $k^+_{M_y}$ when there is no binding of P_x. On comparing the reaction set to that of the activator system the only difference here is in the step corresponding to the production of M_y, whose mean steady state value now is $(\textit{basal transcription rate}) \times G_y/k^-_{M_y} = k^+_{M_y} G_y/k^-_{M_y}$. This change in the reaction *structure* results in negative correlations that is an exact mirror image of the correlations of the activator network. The rate constants and the mean concentrations of mRNAs and proteins are of the same value as in the case of the activator network. This is deliberately done so that the effect of the regulatory mechanism, which in this case is repression, gets illuminated in the dynamic correlations.

$\left[\phi \xrightarrow{k^+_{M_x}} M_x\right]$	$\left[P_x + G_y \underset{k_{off}}{\overset{k_{on}}{\rightleftharpoons}} \mathcal{C}_y\right]$	$\left[M_y \xrightarrow{k^+_{P_y}} M_y + P_y\right]$	$\left[M_x \xrightarrow{k^-_{M_x}} \phi\right]$	$\left[P_x \xrightarrow{k^-_{P_x}} \phi\right]$
$\left[M_x \xrightarrow{k^+_{P_x}} M_x + P_x\right]$	$\left[G_y \xrightarrow{k^+_{M_y}} G_y + M_y\right]$		$\left[M_y \xrightarrow{k^-_{M_y}} \phi\right]$	$\left[P_y \xrightarrow{k^-_{P_y}} \phi\right]$

TABLE 5.5: Reaction set describing the process of repression in $X \dashv Y$. Gene G_y is switched to inactive state $\mathcal{C}_y$ by the repressor P_x. M_y is transcribed when the promoter region is free of this repressor.

The rate equations are similar to the case of the elementary activator except for the production rate of M_y which is now transcribed from the unbound gene molecule G_y.

$$\frac{dG_y}{dt} = -k_{on}P_xG_y + k_{off}(\alpha - G_y) \tag{5.8}$$
$$\frac{dM_x}{dt} = k^+_{M_x} - k^-_{M_x}M_x \tag{5.9}$$
$$\frac{dM_y}{dt} = k^+_{M_y}G_y - k^-_{M_y}M_y \tag{5.10}$$
$$\frac{dP_x}{dt} = k^+_{M_x}M_x - k^-_{P_x}P_x - k_{on}P_xG_y + k_{off}(\alpha - G_y) \tag{5.11}$$
$$\frac{dP_y}{dt} = k^+_{P_y}M_y - k^-_{P_y}P_y \tag{5.12}$$

The vector $\mathbf{R}$ of the deterministic rates, and the stoichiometry matrix are obtained through the above set of rate equations, and are:

$$\mathbf{R} = \Big(k_{on}P_xG_y, k_{off}(\alpha - G_y), \underbrace{k^+_{M_y}G_y}, k^+_{M_x}, k^+_{P_x}M_x, k^+_{P_y}M_y, k^-_{M_x}M_x, k^-_{M_y}M_y, k^-_{P_x}P_x, k^-_{P_y}P_y\Big)^T$$

$$\boldsymbol{\nu} = \begin{pmatrix} -1 & +1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & +1 & 0 & 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & +1 & 0 & 0 & 0 & 0 & -1 & 0 & 0 \\ -1 & +1 & 0 & 0 & +1 & 0 & 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & +1 & 0 & 0 & 0 & -1 \end{pmatrix}$$

Further, the Jacobian matrix corresponding to the above set of rate equations is:

$$\mathbf{A} = \begin{pmatrix} -(k_{off} + k_{on}P_x) & 0 & 0 & -k_{on}G_y & 0 \\ 0 & -k^-_{M_x} & 0 & 0 & 0 \\ \underbrace{k^+_{M_y}} & 0 & -k^-_{M_y} & 0 & 0 \\ -(k_{off} + k_{on}P_x) & k^+_{P_x} & 0 & -(k^-_{P_x} + k_{on}G_y) & 0 \\ 0 & 0 & k^+_{P_y} & 0 & -k^-_{P_y} \end{pmatrix}.$$

The negative correlations (Figure 5.19) are due to the change in the rate equation for M_y that is reflected as a corresponding change in the Jacobian and the reaction rates. In the Jacobian $\mathbf{A}$ and the vector $\mathbf{R}$ we have marked this change in the element by underbracing it. While the vector of the deterministic rates $\mathbf{R}$ is exactly the same as in the case of the activator, the only change is in the rate $k^+_{M_y}(\alpha - G_y)$ which now is $k^+_{M_y}G_y$. Similarly, the Jacobian matrix A is exactly the same as that of the activator system (5.6) with the only exception being that element A_{31} is now $k^+_{M_y}$ in place of $-k^+_{M_y}$.

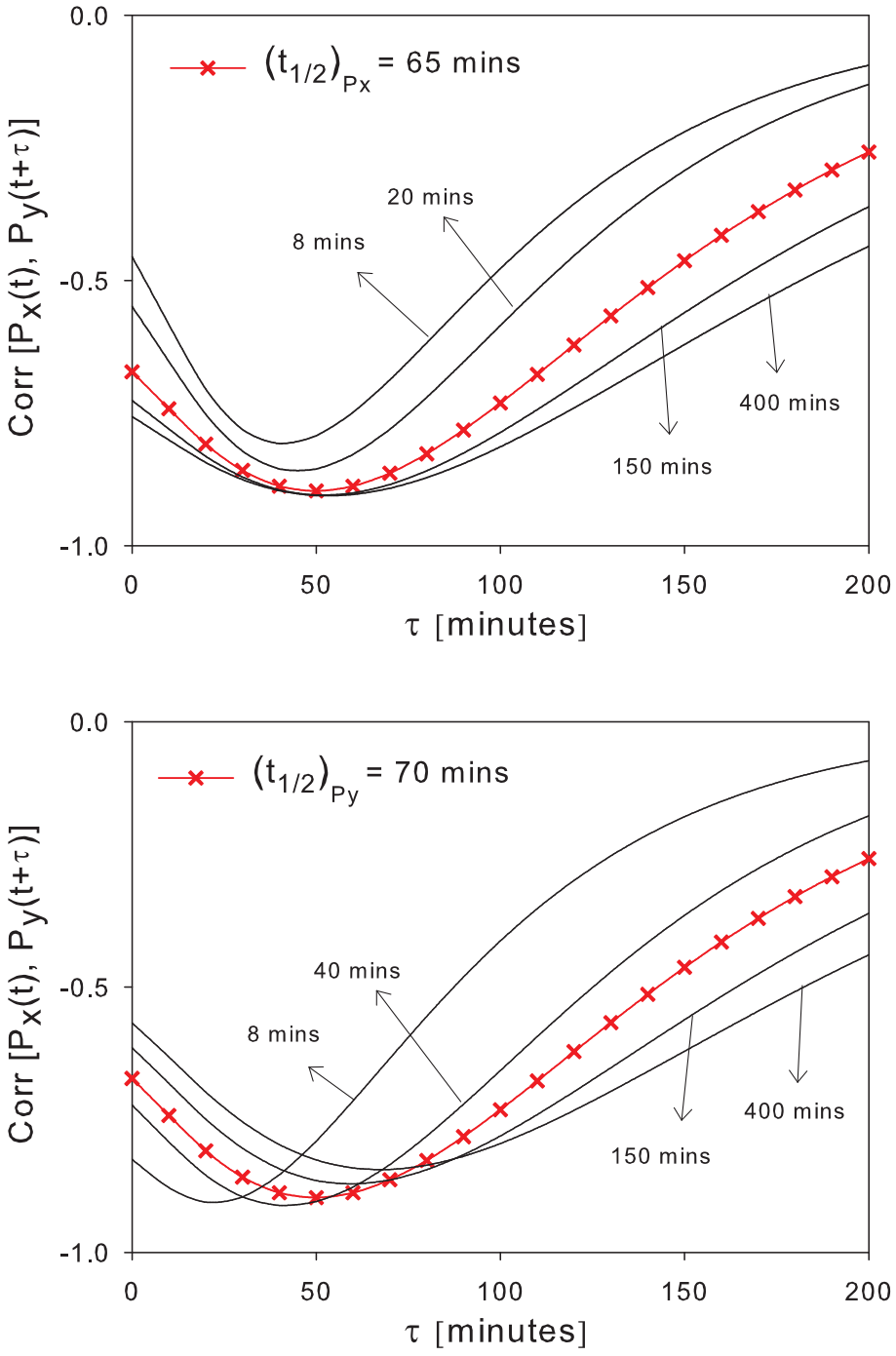

FIGURE 5.19: (a) $[k^+_{P_x}, k^-_{P_x}]$ are varied simultaneously so that the mean concentration level of $P_x = M_x \frac{k^+_{P_x}}{k^-_{P_x}}$ remains constant. While there is an increase in Corr*, τ^* decreases from 55 to 45 minutes for step-wise increase in these parameters. (b) Similarly for step-wise increase in $[k^+_{P_y}, k^-_{P_y}]$, τ^* once again reduces from 102 to 36 minutes.

Therefore, though the eigenvalues λ's remain the same the reason for the negative correlations is that the co-efficients of exponentials in equation (3.14) are now functions of a new set of eigenvectors whose elements involve $k^{+}_{M_y}$. This is the reason that, though the concentrations of the regulated and regulator proteins are not typical of a repressor *i.e.,* $P_y > P_x$, the correlations faithfully reveal the type of regulatory activity that could not be guessed otherwise. For the case where $P_x > P_y$, the observed fluctuation properties remain the same except for reduction in the correlation magnitude. As an example, for $\langle M_y \rangle = 0.1\ nM$ and $\langle P_y \rangle = 100\ nM$ the $\text{Corr}^* = -0.47$.

The correlations between the repressor P_x and the repressed protein P_y exhibits changes in them for corresponding changes in the parameter values. In Figure 5.19 the protein correlations for different half-lives of the proteins are shown. Note the close similarity between these correlations and those in Figures 5.12(b) and 5.13(b) respectively. A change in the regulatory mechanism, which in this case is repression, has influenced the dynamic correlations such that they are now negative in value but otherwise behave very much similar to the case of the elementary activator. The close similarity between the correlations of the two systems with different regulatory mechanisms is due to the fact that the reaction *structure* of the systems are essentially the same, with the only change being the way M_y is transcribed. Hence, the sensitivities of the correlations *w.r.t* the parameters is very similar to those of the activator case. These sensitivities are given in Table 5.6.

5.4 Summary

In this chapter we have demonstrated how the stationary as well as the dynamic correlations between molecular species of a regulatory system vary *(a)* for changes in the values of the parameters, and *(b)* for different types of regulatory mechanisms. The parameters that induce temporal sensitivity in the dynamic correlations are the eigenvalues of the deterministic system, which are in most cases the decay rates of proteins and mRNAs. The decay rates of proteins are especially important in the case of protein correlations. Therefore, the dynamic correlations between the proteins, or also between the mRNAs, illuminate or characterize the type and form of the regulatory mechanism present between two genes.

Elementary Repressor						
Genetic Process	Parameter (p_α)	value	$\frac{\partial \text{Corr}^*}{\partial p_\alpha} \approx$	Sensitivity	$\frac{\partial \tau^*}{\partial p_\alpha} \approx$	Sensitivity
Constant flux	$k^+_{M_x}$	8.14×10^{-3}	4.8	yes / non-Linear	-8.5	no
Translation	$k^+_{P_x}$	36.3	-1×10^{-5}	no	-0.004	no
	$k^+_{P_y}$	143.2	-2×10^{-7}	no	-0.002	no
Transcription	$k^+_{M_y}$	1.03	-0.04	yes / Linear	-0.2	no
Decay process	$k^-_{M_x}$	4.07×10^{-2}	0.8	yes / non-Linear	-128	yes / non-Linear
	$k^-_{M_y}$	7.70×10^{-2}	-0.07	yes / non-Linear	-168	yes / non-Linear
	$k^-_{P_x}$	1.84×10^{-2}	1.6	yes / non-Linear	-302	yes / non-Linear
	$k^-_{P_y}$	1.76×10^{-2}	-3.2	yes / non-Linear	-1410	yes / non-Linear
TF binding/unbinding	k_{on}	1	-3×10^{-4}	yes / non-Linear	-0.08	no
	k_{off}	200.0	1×10^{-6}	yes / non-Linear	-4×10^{-4}	no

TABLE 5.6: Table giving Corr^* and τ^* sensitivities for the case of elementary repressor system. Note the similarity between the values of sensitivities here and in the case of the elementary activator given in Table 5.3.

Chapter 6

Network Mechanisms

Complex regulatory networks, found in unicellular organisms such as bacteria, have been observed to be made up of over-represented small networks comprising of 2-3 genes Milo and et al. (2002). These smaller networks are found to be common not only in microorganisms but also in animals and plants with the difference being in the precise manner in which they interconnect to form the bigger complex network, which in turn performs a specific regulatory function. Whilst studying the bigger network formed via interconnecting the elementary networks holds the key to unraveling many aspects of regulatory functions, the pre-requisite to this, however, is to develop a thorough understanding of these elementary networks. It is towards this objective, that the classification of certain small networks as elementary ones, is crucial. The recurrence of these networks over a wide range of organisms is considered to be an indication of their special properties. These basic networks that are the building blocks of the bigger and more complex networks, have widely been investigated in theory Shen-Orr et al. (2002), and also experimentally with the aid of synthetic networks Alon (2007), and have been found to have well-defined characteristics that are useful for proper functioning of the complex regulatory network. In this chapter, we follow a similar objective of characterizing these networks, but on the basis of their dynamic fluctuation properties. This approach is important since it would adhere to our statements made in the introduction regarding the benefits of single-cell measurements over multi-cell measurements, and would assist in recognizing the regulatory structures of these networks with the aid of such measurements.

Here we consider different types of network mechanisms that, by virtue of differences in their network structures and consequently in their Jacobian matrices, display different dynamic correlation behaviours. Each of these networks consists of an *input* gene-node X, an *output* node Y and an additional gene-node denoted as Z. For example, in the case of *cascaded activation*, the network would be $X \rightarrow Z \rightarrow Y$, where the intermediary transcription factor P_z is now responsible for the activation of P_y. On the other hand,

the *combinatorial activation network* is $X \rightarrow Y \leftarrow Z$ involving activation of Y by the combinatorial action of P_x and P_z. On combining the above two types of network mechanisms we arrive at a third type of network mechanism known as the *feedforward loop*. The coherent and incoherent feedforward loops are discussed towards the end of the chapter. As we shall see in this chapter, the effect that the node Z has on the dynamic fluctuations is of significance in determining the network mechanism of the GRN.

6.1 Cascading Regulatory Networks

Among such basic networks, of particular interest has been the *cascade*, comprising of anywhere around 2 to 7 genes, where each gene regulates the transcription of the gene dowstream to it. In order to study the role of cascades in the functioning of the large network, it is important to analyze their behaviour in isolation. Towards this, Rosenfeld and Alon (2003) sifted through the databases of transcriptional interactions in organisms such as *Escherichia coli* and *Saccharomyces cerevisiae*, and found that cascades of length 2-3 were most common in sensory transcription networks of such organisms, where these networks respond to external fluctuating conditions. On the other hand, larger cascades of length of around 6-7 were common in developmental transcription networks of multicellular organisms such as sea urchin and Drosophila Davidson et al. (2002).

The relation between the mean steady state levels of the *output* and *input* variables, known as the transfer function, and its hyperbolic shape is due to the Michaelis-Menten kinetics of the reactions. For longer cascades, this function results in switching-like behaviour, inducing a sigmoidal shape that generates a sharp threshold for switching the *output* variable to higher concentration levels. This was confirmed experimentally by Pedraza and van Oudenaarden (2005) who designed a synthetic cascade regulatory networks and monitored the expression levels of the proteins through fluorescence microscopy. The fluorescence levels were monitored for changes in the concentration levels of the inducer molecules that were the variables at the front end of the cascade and which induced the production of the protein variables that were dowstream in the cascade. Similarly, the synthetic circuit designed by Hooshangi et al. (2005) consisted of the *input* variable anhydrotetracycline (aTc) regulating, via a cascade of repressors, the transcription of the *output* variable, which was the enhanced yellow fluorescent protein (*eyfp*). Both studies concluded that for increased number of steps in the cascade, the steady state transfer curve became steeper. This enhanced sensitivity of a gene compared to its upstream gene, when the inducer concentration is varied, demonstrated the utility of cascades for generating steep responses. Apart from the above, the significance of regulatory cascades also lies in the fact that the time taken for the downstream

components to respond to signals upstream, which is important in timing regulatory decisions, can be varied as a function of reaction rates and also the number of stages of the cascade. A direct relation to these *response* times is the time taken for the dynamic correlations to reach maximum value which is τ^*. In section 3.3 we proved that such a relation does indeed exist and hence the need for single-cell measurements through which dynamic correlations could be evaluated. Therefore, the quantities of interest are the dynamic covariances or correlations between the proteins, more so between the *input* and *output* proteins P_x and P_y.

6.1.1 Cascading Activation - I

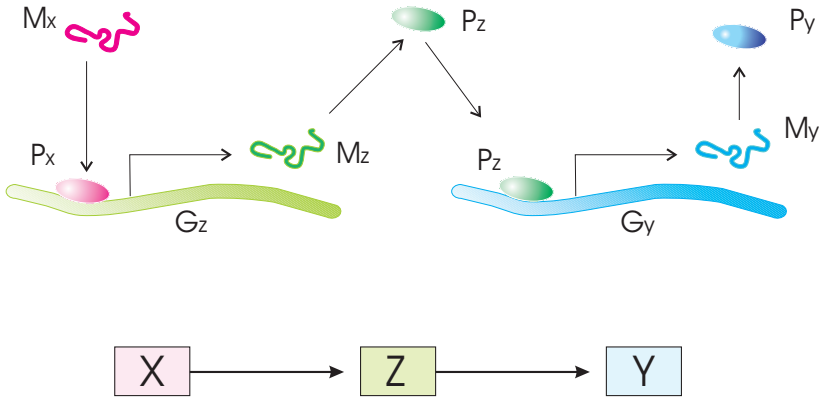

FIGURE 6.1: Schematic and network representations of a cascade involving activators.

Figure 6.1 shows a schematic representation of a two-stage cascade with the proteins acting as activators. P_x activates the production of P_z, which in turn activates the production of P_y. The mean concentration levels and the decay rates of the *input* and *output* proteins and mRNAs are of the same value as that of the *elementary activator* system of section 5.1. However, the mean concentration levels of the intermediary species M_z and P_z are assumed to be 2.5 nM and 2000 nM respectively. These concentration levels are chosen such that they fall in between the corresponding concentration levels of the upstream node X and the downstream node Y. The half-lives of M_z and P_z are assumed to be of 8 and 40 minutes respectively. Dilution by cell-division is also a factor in determining the decay rates of these species. The transcription and translation rates of M_z are correspondingly evaluated as $k^+_{M_z} = k^-_{M_z}\langle M_z\rangle/\langle P_x\rangle$ and $k^+_{P_z} = k^-_{P_z}\langle P_z\rangle/\langle M_z\rangle$. The reaction set for this regulatory scheme is given in Table 6.1 and where the dissociation constants $K_{D_1} = k_{off_1}/k_{on_1}$ and $K_{D_2} = k_{off_2}/k_{on_2}$ are assumed to be of same value of 200 nM.

$\left[\phi \xrightarrow{k^+_{M_x}} M_x\right]$	$\left[P_x + G_z \underset{k_{off_1}}{\overset{k_{on_1}}{\rightleftharpoons}} \mathcal{C}_z\right]$	$\left[P_z + G_y \underset{k_{off_2}}{\overset{k_{on_2}}{\rightleftharpoons}} \mathcal{C}_y\right]$	$\left[M_z \xrightarrow{k^+_{P_z}} M_z + P_z\right]$	$\left[M_x, M_z, M_y \xrightarrow{k^-_{M_x}, k^-_{M_z}, k^-_{M_y}} \phi\right]$
$\left[M_x \xrightarrow{k^+_{P_x}} M_x + P_x\right]$	$\left[\mathcal{C}_z \xrightarrow{k^+_{M_z}} \mathcal{C}_z + M_z\right]$	$\left[\mathcal{C}_y \xrightarrow{k^+_{M_y}} \mathcal{C}_y + M_y\right]$	$\left[M_y \xrightarrow{k^+_{P_y}} M_y + P_y\right]$	$\left[P_x, P_z, P_y \xrightarrow{k^-_{P_x}, k^-_{P_z}, k^-_{P_y}} \phi\right]$

TABLE 6.1: Reaction set of the *cascading activation-I* regulatory system.

$\left[\frac{dG_z}{dt} = -k_{on_1} G_z P_x + k_{off_1}(\alpha - G_z)\right]$	$\left[\frac{dG_y}{dt} = -k_{on_2} G_y P_z + k_{off_2}(\beta - G_y)\right]$	
$\left[\frac{dM_x}{dt} = k^+_{M_x} - k^-_{M_x} M_x\right]$	$\left[\frac{dM_z}{dt} = k^+_{M_z}(\alpha - G_z) - k^-_{M_z} M_z\right]$	$\left[\frac{dM_y}{dt} = k^+_{M_y}(\beta - G_y) - k^-_{M_y} M_y\right]$
$\left[\frac{dP_x}{dt} = k^+_{P_x} M_x - k^-_{P_x} P_x - k_{on_1} G_z P_x + k_{off_1}(\alpha - G_z)\right]$	$\left[\frac{dP_z}{dt} = k^+_{P_z} M_z - k^-_{P_z} P_z - k_{on_2} G_y P_z + k_{off_2}(\beta - G_y)\right]$	$\left[\frac{dP_y}{dt} = k^+_{P_y} M_y - k^-_{P_y} P_y\right]$

TABLE 6.2: A reduced set of differential equations describing the time-evolution of the deterministic variables of the *cascading activation-I* regulatory system. Here, the variables $\mathcal{C}_z$ and $\mathcal{C}_y$ are eliminated by the rule of conservation.

$\left[\frac{dM_x}{dt} = k^+_{M_x} - k^-_{M_x} M_x\right]$	$\left[\frac{dM_z}{dt} = k^+_{M_z}\left(\frac{\alpha P_x}{P_x + K_{D_1}}\right) - k^-_{M_z} M_z\right]$	$\left[\frac{dM_y}{dt} = k^+_{M_y}\left(\frac{\alpha P_z}{P_z + K_{D_2}}\right) - k^-_{M_y} M_y\right]$
$\left[\frac{dP_x}{dt} = k^+_{P_x} M_x - k^-_{P_x} P_x\right]$	$\left[\frac{dP_z}{dt} = k^+_{P_z} M_z - k^-_{P_z} P_z\right]$	$\left[\frac{dP_y}{dt} = k^+_{P_y} M_y - k^-_{P_y} P_y\right]$

TABLE 6.3: A further reduced set of differential equations obtained by eliminating G_z and G_y.

The deterministic time-evolution of the variables is defined by the set of ODEs given in Table 6.2 where the DNA-TF complexes are eliminated by the law of conservation:

$$\begin{aligned} \mathcal{C}_z(t) &= G_z(t_0) - G_z(t) = \alpha - G_z(t) \\ \mathcal{C}_y(t) &= G_y(t_0) - G_y(t) = \beta - G_y(t) \end{aligned}$$

Further, the fast reactions that involve the binding and unbinding of the TF to the DNA can also be eliminated by solving for the variables G_z and G_y at steady state and then substituting for them in the other ODEs. As demonstrated in section 5.1.2 such elimination of these fast TF-DNA reactions does not affect the dynamic correlation between the remaining variables of the system. We shall therefore use the set of ODEs given in Table 6.3 for evaluating the dynamic correlations. The variables of such a reduced system are now $[M_x, M_z, M_y, P_x, P_z, P_y]$ for which the corresponding Jacobian matrix is:

$$\mathbf{A} = \begin{pmatrix} -k^-_{M_x} & 0 & 0 & 0 & 0 & 0 \\ 0 & -k^-_{M_z} & 0 & A_{24} & 0 & 0 \\ 0 & 0 & -k^-_{M_y} & 0 & A_{35} & 0 \\ k^+_{P_x} & 0 & 0 & -k^-_{P_x} & 0 & 0 \\ 0 & k^+_{P_z} & 0 & 0 & -k^-_{P_z} & 0 \\ 0 & 0 & k^+_{P_y} & 0 & 0 & -k^-_{P_y} \end{pmatrix} \tag{6.1}$$

where the off-diagonal elements A_{24} and A_{35} are,

$$A_{24} = \frac{\partial(dM_z/dt)}{\partial P_x} = k^+_{M_z}\alpha\left[\frac{K_{D_1}}{(P_x + K_{D_1})^2}\right] \tag{6.2}$$

$$A_{35} = \frac{\partial(dM_y/dt)}{\partial P_z} = k^+_{M_y}\beta\left[\frac{K_{D_2}}{(P_z + K_{D_2})^2}\right] \tag{6.3}$$

The vector of deterministic rates and the stoichiometric matrix are:

$$\begin{aligned} \mathbf{R} = \Bigg(& k^+_{M_x}, k^+_{M_z}\frac{\alpha\frac{P_x}{K_{D_1}}}{(1+\frac{P_x}{K_{D_1}})}, k^+_{M_y}\frac{\beta\frac{P_z}{K_{D_2}}}{(1+\frac{P_z}{K_{D_2}})}, k^+_{P_x}M_x, k^+_{P_z}M_z, k^+_{P_y}M_y, \\ & k^-_{M_x}M_x, k^-_{M_z}M_z, k^-_{M_y}M_y, k^-_{P_x}P_x, k^-_{P_z}P_z, k^-_{P_y}P_y\Bigg)^T \end{aligned}$$

$$\boldsymbol{\nu} = \begin{pmatrix} +1 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & +1 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & +1 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & +1 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & +1 & 0 & 0 & 0 & 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & +1 & 0 & 0 & 0 & 0 & 0 & -1 \end{pmatrix}$$

The quantity of interest is the dynamic correlation between P_x and P_y, which is evaluated by making use of the above matrices via equations (3.14) and (3.15). This function is plotted in Figure 6.2 where the other cross-correlations are also shown. For purposes of comparison, the correlation between these proteins in the case of the two-gene activator system (section 5.1) are also shown (in red). The effect of introducing an intermediary regulatory node Z to the original two-gene activator network is that there is a decrease in the magnitude of the correlation and there is also a shift in this correlation curve along the time-axis τ. The time τ^* at which the correlations achieve maximum value, which was 49 minutes for the two-gene case, is now doubled to 97 minutes.

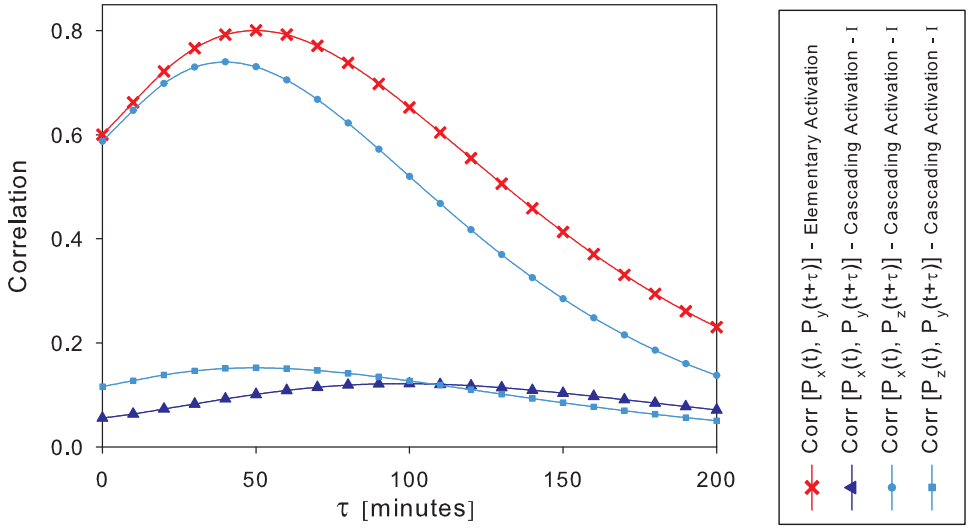

FIGURE 6.2: The dynamic correlations between the proteins of the *cascading activation-I* system are shown in blue, whilst the correlation between P_x and P_y of the two-gene *elementary activation* system is shown in red for comparison. τ^* for the $\text{Corr}[P_x(t), P_z(t+\tau)]$ function is 39 minutes, for the $\text{Corr}[P_z(t), P_y(t+\tau)]$ function is 48 minutes, and in the case of $\text{Corr}[P_x(t), P_y(t+\tau)]$ is 97 minutes. The mean values are: $\langle M_x \rangle = 0.2\ nM, \langle M_z \rangle = 2.5\ nM, \langle M_y \rangle = 4.5\ nM, \langle P_x \rangle = 395\ nM, \langle P_z \rangle = 2000\ nM, \langle P_y \rangle = 36001\ nM$. The corresponding parameter values are: $k^+_{M_x} = 0.008\ nM\ min^{-1}, k^+_{M_z} = 0.35\ min^{-1}, k^+_{M_y} = 0.38\ min^{-1}, k^+_{P_x} = 36\ min^{-1}, k^+_{P_z} = 20\ min^{-1}, k^+_{P_y} = 143\ min^{-1}, k^-_{M_x} = 0.0407\ min^{-1}, k^-_{M_z} = 0.0927\ min^{-1}, k^-_{M_y} = 0.077\ min^{-1}, k^-_{P_x} = 0.0184\ min^{-1}, k^-_{P_z} = 0.025\ min^{-1}, k^-_{P_y} = 0.0176\ min^{-1}$. The TF-DNA dissociation constants are $K_{D_1} = K_{D_2} = 200\ nM$.

The decrease in the level of correlation is due to decrease in covariance levels. For example, if we take the stationary variance $\langle \delta P_y^2 \rangle$ for the two-gene case, which is $\text{Cov}[P_x(t), P_y(t)]$, it is of value 2.03×10^8 whereas for the three-gene cascade network it is reduced to 5.69×10^7. This could seem contrary to the idea that any fluctuations would increase on addition of a node in a cascade. However, Thattai and van Oudenaarden (2002) show that the condition for noise in the output variable to be bounded or in fact to decrease is that the magnitude of the derivative of transfer functions, which they call the differential amplification factor, should be less than one. Their model is based on

the Langevin method, where a time-dependent noise term having Gaussian white-noise properties, is tagged on to the deterministic dynamical ODEs. The parameter values used in our case adheres to the above condition and therefore the decrease in the variance of P_y. For example, let us consider the rate of production of P_z given by the ODE:

$$\frac{d\langle P_z\rangle}{dt} = k^+_{P_z}\langle M_z\rangle - k^-_{P_z}\langle P_z\rangle = k^+_{P_z}\frac{k^+_{M_z}}{k^-_{M_z}}\Big(\frac{\alpha\langle P_x\rangle}{\langle P_x\rangle + K_{D_1}}\Big) - k^-_{P_z}\langle P_z\rangle \tag{6.4}$$

The differential amplification factor is nothing but the partial derivative of the RHS of the above ODE *w.r.t* P_x, which is,

$$k^+_{P_z}\frac{k^+_{M_z}}{k^-_{M_z}}\Big(\frac{\alpha K_{D_1}}{(\langle P_x\rangle + K_{D_1})^2}\Big) = 20.0 \times \frac{0.348}{0.0926} \times \Big(\frac{1.0 \times 200}{(395.6 + 200)^2}\Big) = 0.042$$

Similarly for the variable P_y, the factor is 0.03. Since these factors are less than one, the stationary covariance in the protein P_y reduces in value as compared to that of the two-gene network. Similarly, the stationary covariance between the proteins P_x and P_y also reduces from 4.22×10^6 to 2.08×10^5. Therefore the stationary correlation or the dynamic correlation at time $\tau = 0$ is,

$$\text{Corr}[P_x(t), P_y(t)] = \frac{2.08 \times 10^5}{\sqrt{2.43 \times 10^5}\ \sqrt{5.69 \times 10^7}} = 0.056$$

whilst on the other hand for the two-gene case, this is,

$$\text{Corr}[P_x(t), P_y(t)] = \frac{4.22 \times 10^6}{\sqrt{2.43 \times 10^5}\ \sqrt{2.03 \times 10^8}} = 0.6$$

Coming back to the increase in τ^*, due to the relation between the time τ^* and the response time t_{resp} that we derived in section 3.3, there is an equivalent doubling in the response time as well. This is in accordance with the results obtained by Rosenfeld and Alon (2003), who noticed doubling of the response time, which they define as the time to reach half of the change between the pre-induced steady state and the post-induced steady state. They noticed that for proteins that decay via dilution, this response time was approximately equal to one cell-cycle time for each stage of the cascade, and therefore the time taken for the downstream components of the cascade to respond to signals at the top of the cascade increases with increase in the length of the cascade. However, it was observed by Hooshangi et al. (2005) in their experiments that as the cascade grew in length, there was not only a loss of synchronization in these response times, but also the variation in the protein levels increased. This was a potential problem in a population of cells in their ability to respond uniformly to an external signal. Therefore the limitation of the response measurements are that they are the mean response of a population of cells to an external pertubation, where the issue of cell synchronization creeps in. In our proposed scheme this issue could be addressed by monitoring protein

numbers in single cells over time and then evaluating the dynamic correlation functions between different protein variables of the cascade. Whilst in the case of the deterministic responses, the response times increase for increasing cascade length, the τ^* which is the time taken for correlations to reach maximum value and is related to the response time (section 3.3) also increases equivalently for addition of each stage of the cascade. Consequently, we retain the ability to identify the network structure through their well-known characteristics, which in this case is of increased τ^* of the dynamic correlations between the input variable P_x and the output variable P_y. As our approach avoids issues of cell syncronization, the use of single-cell measurements in characterizing these networks becomes all the more relevant.

6.1.1.1 Sensitivity of protein correlations to decay rates

In the previous chapter we focussed on the correlation between the protein variables and their sensitivity for changes in the parameter values. This gave us a clear picture of the fluctuation properties of the two-gene networks. In the present case of three-genes, let us study the effect of the paramaters on the correlations $\mathrm{Corr}[P_x(t), P_y(t+\tau)]$ between the same protein variables. Firstly, by changing the order of variables to $[M_x, P_x, M_z, P_z, M_y, P_y]$, and consequently re-arranging the Jacobian matrix of (6.1), we get a lower triangular matrix whose diagonal elements are its eigenvalues.

$$\mathbf{A} = \begin{pmatrix} -k^-_{M_x} & 0 & 0 & 0 & 0 & 0 \\ k^+_{P_x} & -k^-_{P_x} & 0 & 0 & 0 & 0 \\ 0 & A_{24} & -k^-_{M_z} & 0 & 0 & 0 \\ 0 & 0 & k^+_{P_z} & -k^-_{P_z} & 0 & 0 \\ 0 & 0 & 0 & A_{35} & -k^-_{M_y} & 0 \\ 0 & 0 & 0 & 0 & k^+_{P_y} & -k^-_{P_y} \end{pmatrix}$$

These are the decay rates of all the mRNA and protein variables and therefore have a significant effect on the defining *features* of the protein correlation curve. For example, we notice in Figure 6.3(a) that not only does the stationary correlation at time $\tau = 0$ increase for increase in the mRNA decay rate $k^-_{M_y}$, but also that there is a larger increase in Corr^* with not much variation in the value of τ^*. Note that the mean steady state value of M_y is held constant at 4.5 nM by inducing equivalent variations in the transcription rate $k^+_{M_y}$. Overall, variations are induced in the elements A_{35} and $-k^-_{M_y}$ of the above Jacobian matrix, which correspond to the variable M_y. Therefore, a fast decaying and fast transcribing mRNA M_y increases the correlation between the proteins. By comparing the above analysis with that of the two-gene case, shown in Figure 5.13(a), we notice that though the correlations at $\tau = 0$ and at τ^* are similarly sensitive to the mRNA decay rate, there is a large variation in τ^*. This is due to the proximity of the gene-node Y to X in the two-gene case as compared to the three-gene case. Hence,

the fluctuation characteristics as revealed in the dynamic correlations can be used as an input into a *network identification algorithm*.

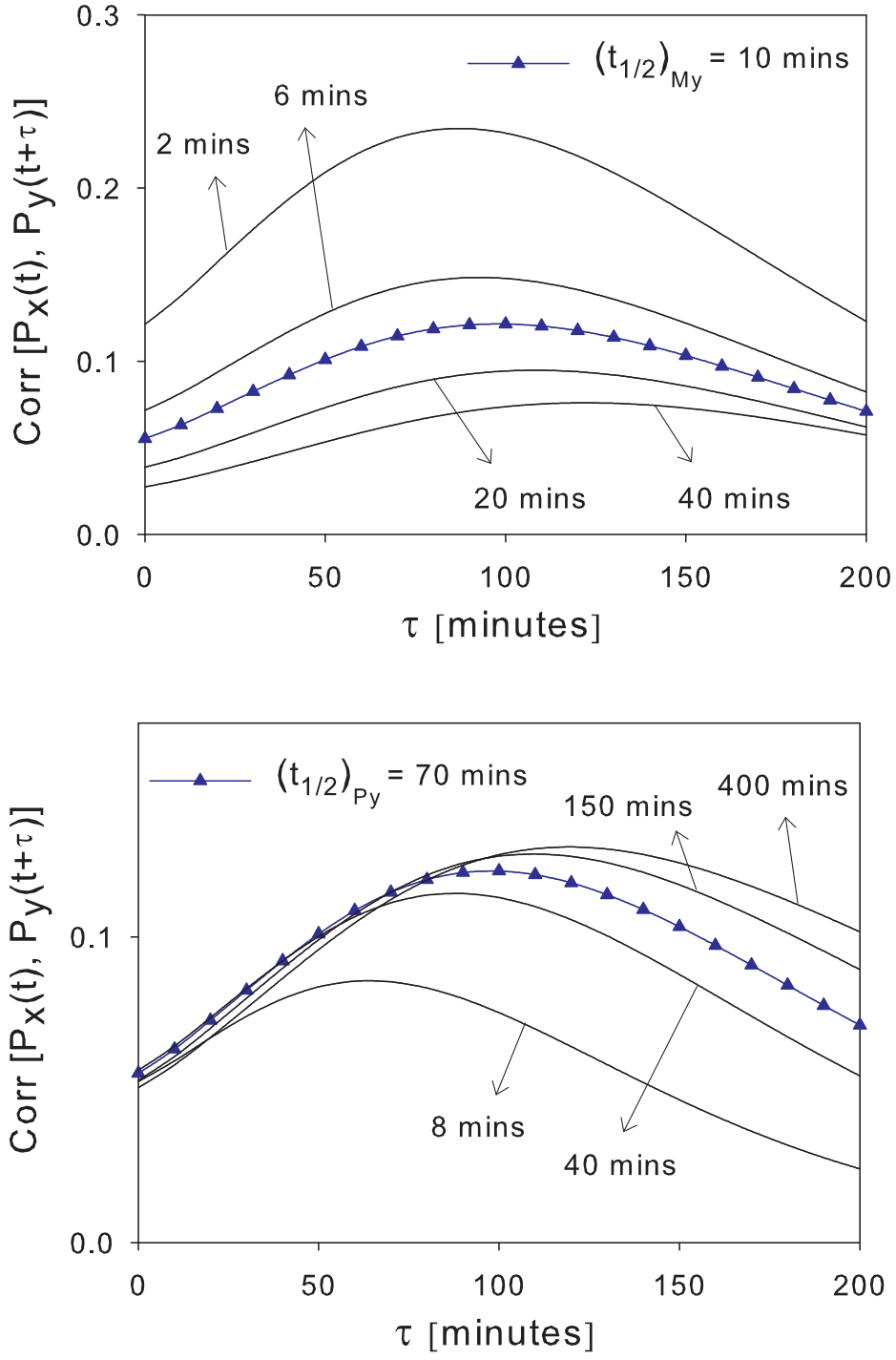

FIGURE 6.3: (a) The half-life of mRNA M_y is varied in steps from 2 to 40 minutes, for which there is a corresponding increase in the value of $\tau^* = [86, 92, 97, 107, 121]$ minutes, accompanied by a decrease in value of Corr*. (b) Similarly, for increase in the half-life of protein P_y from 8 to 400 minutes, τ^* increases as $[63, 87, 97, 109, 118]$ minutes respectively. There is also an increase in Corr* for increased half-life of P_y.

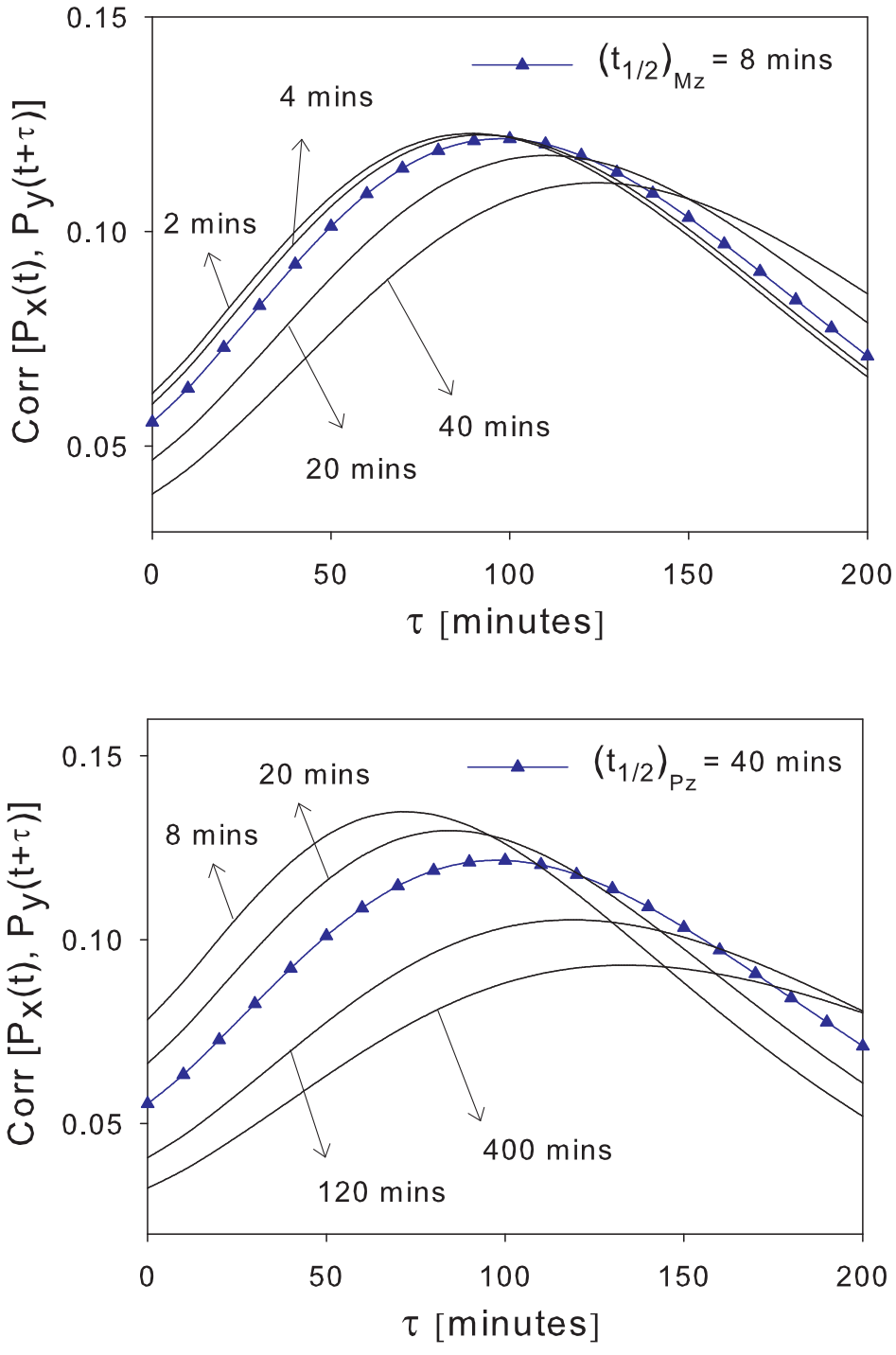

FIGURE 6.4: (a) The half-life of the intermediary mRNA M_z is varied in steps from 2 to 40 minutes, for which there is a corresponding increase in the value of τ^* $[88, 91, 97, 109, 123]$ minutes respectively. (b) Similarly, $k^-_{P_z}$ or rather, the half-life of protein P_z is increased from 8 to 400 minutes for which τ^* increases as $[71, 84, 97, 118, 133]$ minutes respectively. There is also a decrease in both Corr* and the stationary correlation (at $\tau = 0$).

On the other hand, for variations in the decay rate $k^-_{P_y}$ (Figure 6.3(b)), there is a noticable difference in the sensitivity of the stationary correlations as compared to the two-gene case (Figure 5.13(b)). In the three-gene case, the stationary correlation is almost insensitive to this parameter, whilst features such as Corr* and τ^* are sensitive. Similarly, from Figures 6.4(a) and 6.4(b), we conclude that decay rates of mRNAs and proteins of the intermediate node Z have an effect on all the features of protein correlations. In this regard it is interesting to note that, as half-life of P_z is reduced, τ^* moves closer to the value corresponding to the two-gene case whilst the magnitude of correlation remains low, suggesting ways of controlling different features of the correlations.

6.1.2 Cascading Activation - II

A regulatory cascade can be built using repressors where the regulatory function between the input and output proteins is of activation. In fact experiments such as those by Pedraza and van Oudenaarden (2005) and Hooshangi et al. (2005) involve the use of only repressor elements in the cascade. This is because bacterial systems are easier to tinker with, and repressors have a simpler mechanism. In the 2-stage cascade shown in Figure 6.5, the resultant regulatory function between the proteins P_x and P_y, is that of activation.

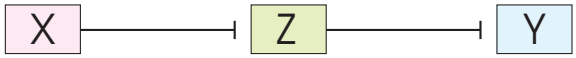

FIGURE 6.5: Network representation of the *cascading activation-II* regulatory system.

This is reflected in the protein correlations shown in Figure 6.6. The difference in these correlations and those of the *cascading activation-I* network is that the correlations between P_x and P_y is increased greatly in magnitude. This is because of the reaction scheme of this network, given in Table 6.4, where there is a presence of basal transcription rate of high value so that the mean value of mRNA and proteins are same as that of the *cascading activation-I* network. The positive correlations are due to the simultaneous sign changes in both the off-diagonal entries A_{24} and A_{35} of the Jacobian matrix,

$$A_{24} = \frac{\partial(dM_z/dt)}{\partial P_x} = -k^+_{M_z}\alpha\left[\frac{K_{D_1}}{\left(P_x + K_{D_1}\right)^2}\right]$$
$$A_{35} = \frac{\partial(dM_y/dt)}{\partial P_z} = -k^+_{M_y}\beta\left[\frac{K_{D_2}}{\left(P_z + K_{D_2}\right)^2}\right]$$

This results in the eigenvectors changing signs twice, and therefore the covariances which are sum of exponentials whose co-efficients are in turn functions of these eigenvectors do not change sign and remain positive. However, the similarity between the two networks is that the τ^*'s of the correlation curves between the three proteins remain exactly the same.

$\left[\phi \xrightarrow{k^+_{M_x}} M_x\right]$	$\left[P_x + G_z \underset{k_{off_1}}{\overset{k_{on_1}}{\rightleftharpoons}} \mathcal{C}_z\right]$	$\left[P_z + G_y \underset{k_{off_2}}{\overset{k_{on_2}}{\rightleftharpoons}} \mathcal{C}_y\right]$	$\left[M_z \xrightarrow{k^+_{P_z}} M_z + P_z\right]$	$\left[M_x, M_z, M_y \xrightarrow{k^-_{M_x}, k^-_{M_z}, k^-_{M_y}} \phi\right]$
$\left[M_x \xrightarrow{k^+_{P_x}} M_x + P_x\right]$	$\left[G_z \xrightarrow{k^+_{M_z}} G_z + M_z\right]$	$\left[G_y \xrightarrow{k^+_{M_y}} G_y + M_y\right]$	$\left[M_y \xrightarrow{k^+_{P_y}} M_y + P_y\right]$	$\left[P_x, P_z, P_y \xrightarrow{k^-_{P_x}, k^-_{P_z}, k^-_{P_y}} \phi\right]$

TABLE 6.4: Reaction set of the *cascading activation-II* regulatory system.

$\left[\frac{dM_x}{dt} = k^+_{M_x} - k^-_{M_x} M_x\right]$	$\left[\frac{dM_z}{dt} = k^+_{M_z}\left(\frac{\alpha K_{D_1}}{P_x + K_{D_1}}\right) - k^-_{M_z} M_z\right]$	$\left[\frac{dM_y}{dt} = k^+_{M_y}\left(\frac{\alpha K_{D_2}}{P_z + K_{D_2}}\right) - k^-_{M_y} M_y\right]$
$\left[\frac{dP_x}{dt} = k^+_{P_x} M_x - k^-_{P_x} P_x\right]$	$\left[\frac{dP_z}{dt} = k^+_{P_z} M_z - k^-_{P_z} P_z\right]$	$\left[\frac{dP_y}{dt} = k^+_{P_y} M_y - k^-_{P_y} P_y\right]$

TABLE 6.5: A reduced set of differential equations describing the time-evolution of the deterministic variables of the *cascading activation-II* regulatory system. The variables $\mathcal{C}_z, \mathcal{C}_y$ and G_z, G_y are eliminated by making use of the conservation rule and by eliminating the fast reactions respectively.

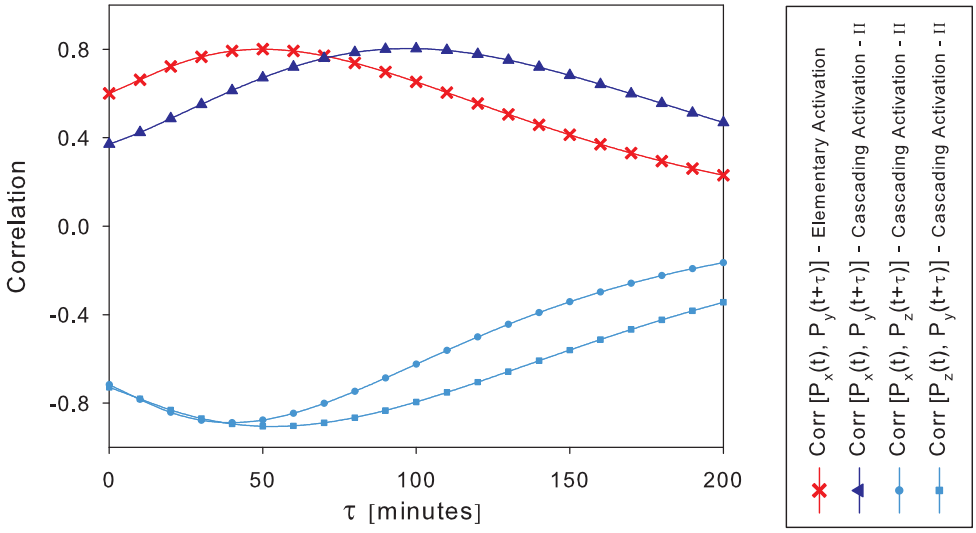

FIGURE 6.6: The dynamic correlations between the proteins of the *cascading activation-II* system are shown in blue, whilst the correlation between P_x and P_y of the two-gene *elementary activation* system is shown in red for comparison. τ^* for the $\text{Corr}[P_x(t), P_y(t+\tau)]$ function is 96 minutes.

6.1.3 Cascading Repression - I

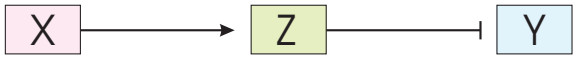

FIGURE 6.7: Network representation of the *cascading repression-I* regulatory system. The resultant regulatory function between the proteins P_x and P_y, is that of repression.

Similar to the above two cases of cascaded activation, we could place an activator behind or front of a repressor element to obtain an overall regulatory function of repression. Under such a condition, the only changes to the system are the simple sign changes in the off-diagonal elements of the Jacobian, which are now,

$$A_{24} = \frac{\partial(dM_z/dt)}{\partial P_x} = +k^+_{M_z}\alpha\left[\frac{K_{D_1}}{\left(P_x + K_{D_1}\right)^2}\right]$$

$$A_{35} = \frac{\partial(dM_y/dt)}{\partial P_z} = -k^+_{M_y}\beta\left[\frac{K_{D_2}}{\left(P_z + K_{D_2}\right)^2}\right]$$

From the correlation plots of Figure 6.8 it is clear that the correlations between the proteins reflect the regulatory function between them *independently* without any other influence.

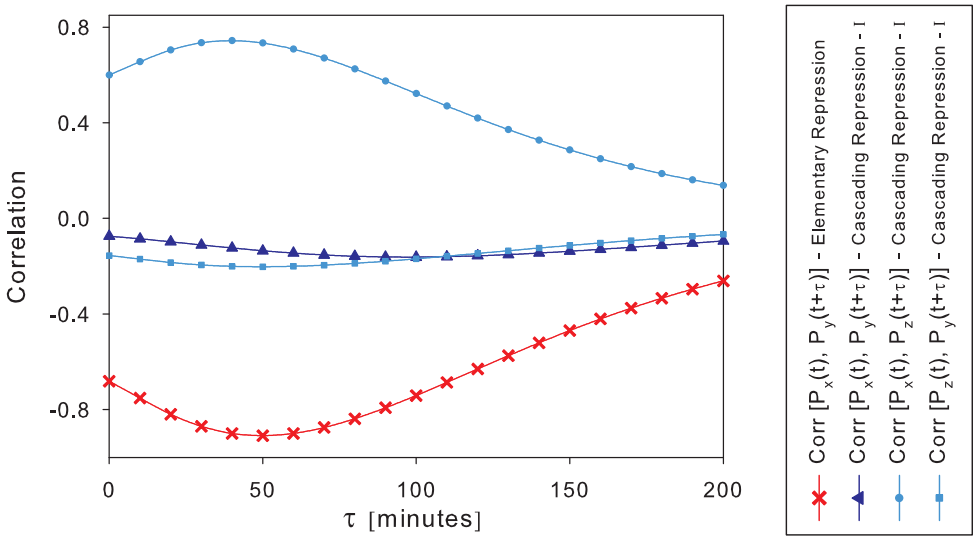

FIGURE 6.8: The dynamic correlations between the proteins of the *cascading repression-I* or $X \rightarrow Z \dashv Y$ network.

6.1.4 Cascading Repression - II

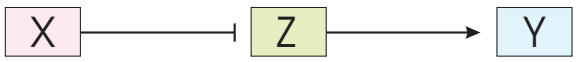

FIGURE 6.9: Network representation of the *cascading repression-II* regulatory system. The resultant regulatory function between the proteins P_x and P_y, is that of repression.

Repression between the input and output nodes of the cascade can also be brought about by the above network mechanism, resulting in the dynamic correlations in Figure 6.10. Once again, the only changes in the Jacobian elements is in the change of their signs,

$$A_{24} = \frac{\partial(dM_z/dt)}{\partial P_x} = -k^+_{M_z}\alpha\left[\frac{K_{D_1}}{(P_x + K_{D_1})^2}\right]$$

$$A_{35} = \frac{\partial(dM_y/dt)}{\partial P_z} = +k^+_{M_y}\beta\left[\frac{K_{D_2}}{(P_z + K_{D_2})^2}\right]$$

The similarity between all the above four types of regulatory cascades is that the time-characteristics of the correlations as reflected by τ^* are exactly the same in each case, whilst the difference being in the type of regulation between nodes of the cascade. In the next section we shall investigate the different types of network mechanisms where there is competitive binding of TFs on the same regulatory sequence of Y.

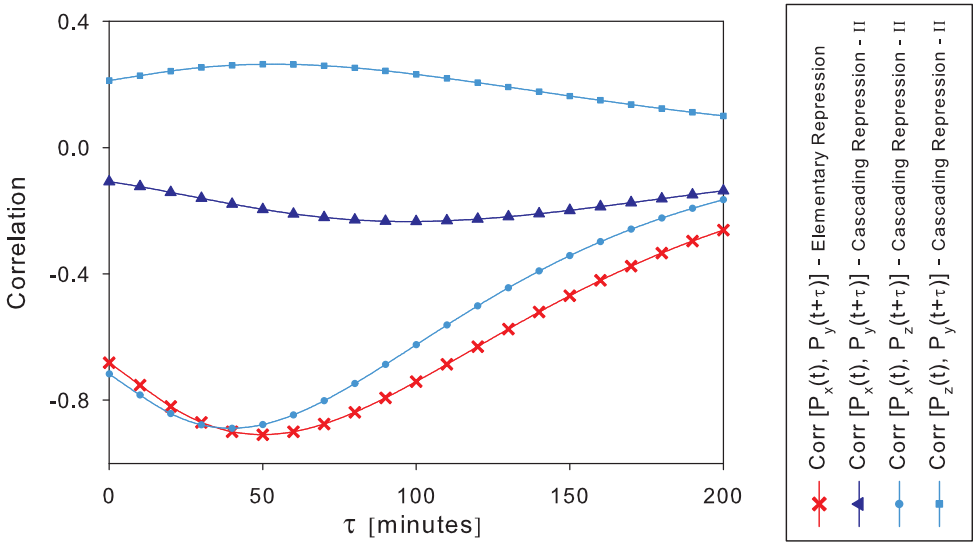

FIGURE 6.10: The dynamic correlations between the proteins of the *cascading repression-II* or $X \dashv Z \rightarrow Y$ network.

6.2 Combinatorial Regulation

In the previous cases of regulation, there was just one type of TF acting on the regulatory region of the DNA. The formation of protein dimers and their recruitment on to the regulatory region was studied in section 5.2, where the magnitude of correlations between the TF proteins and the proteins whose product it regulates, greatly increased. However, it has been known previously that, with at least two types of TFs acting on their respective binding sites on the target DNA, different responses are generated in the regulated protein. This was investigated by Ptashne and Gann (1997) and later by Buchler et al. (2003) who formulated a quantitative model of combinatorial regulation based on the regulation observed in bacteria. They consider the interactions between the TFs and their respective binding-sites on the DNA with certain dissociation constants, and additionally consider interactions of varying strengths between these TFs and RNA polymerase molecules. The result of various combinations for the strengths of the molecular interactions, the dissociation constants and concentrations of the TFs is that a variety of regulatory functions are generated. The significance of their model was that, generating complex regulatory functions was shown to be possible, without resorting to additional classes of regulatory networks. Here, we study the most basic form of combinatorial regulation, where TFs that are labelled beforehand as activators and repressors regulate transcription of the downstream node Y.

6.2.1 Dual Activators

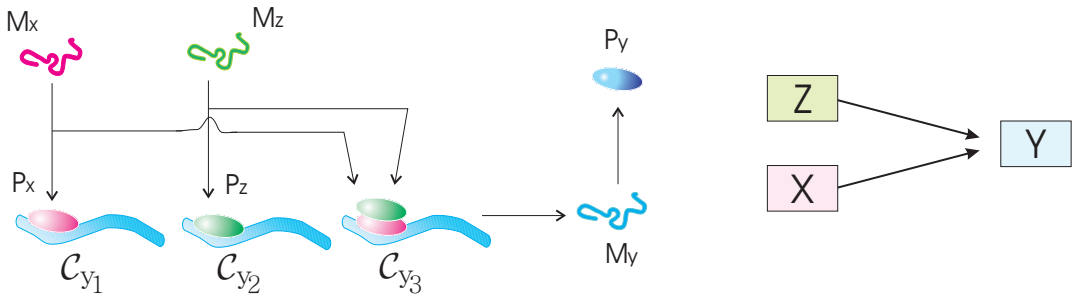

FIGURE 6.11: On the left is a schematic representation of two activators regulating the transcription of G_y. Whilst the three possible ways of activation are shown for a better understanding, the actual model involves only one binding site. To the right is the equivalent network representation.

Here, G_y is activated through the combined action of two types of transcription factors, both of which are activators. For simplicity we make two assumptions, which do not alter the characteristics of this regulatory scheme. Firstly, we neglect the presence of RNAP molecules but instead assume that the gene is self-transcribed once the TF molecule binds to its *cis*-regulatory sequence. This would not only include the effect of transcription during activation by TFs but also the basal transcription process. The second assumption is that there is only one binding site upstream of the target DNA to which the TFs bind. We shall represent this binding site by the variable G_y itself. Therefore, activation is through competitive binding of the two TFs P_x and P_z on to the regulatory region of the gene that is being transcribed. The binding and unbinding strengths may be different for each of the TFs. Transcription in such a case is represented explicitly by two reactions each with its own rate constant and each representing the activation by P_x and P_z. Further activation occurs when both TFs interact and bind together to the regulatory region. This is represented by a third set of binding/unbinding reactions that result in the TF-DNA complex $\mathcal{C}_{y_3}$. The set of binding/unbinding reactions are:

$$P_x + G_y \underset{k_{off_1}}{\overset{k_{on_1}}{\rightleftharpoons}} \mathcal{C}_{y_1}$$

$$P_z + G_y \underset{k_{off_2}}{\overset{k_{on_2}}{\rightleftharpoons}} \mathcal{C}_{y_2}$$

$$P_z + \mathcal{C}_{y_1} \underset{k_{off_3}}{\overset{k_{on_3}}{\rightleftharpoons}} \mathcal{C}_{y_3}$$

By our first assumption, the above complexes self-transcribe to form the mRNA transcripts M_y. To distinguish clearly the influence of each TF on the rate of transcript production, we use three distinct steps of trancription each with its own rate constant:

$$\mathcal{C}_{y_i} \xrightarrow{k^+_{M_{y_i}}} \mathcal{C}_{y_i} + M_y \qquad (i = 1, 2, 3)$$

To obtain the dynamic correlations between the species, we follow the procedure outlined in previous chapters starting with the description of the deterministic dynamics of the system. From the binding/unbinding reactions of the TFs and the DNA, we get the following set of rate equations:

$$\begin{aligned}
\frac{dG_y}{dt} &= k_{off_1}\mathcal{C}_{y_1} - k_{on_1}G_yP_x + k_{off_2}\mathcal{C}_{y_2} - k_{on_2}G_yP_z \\
\frac{d\mathcal{C}_{y_1}}{dt} &= k_{on_1}G_yP_x - k_{off_1}\mathcal{C}_{y_1} + k_{off_3}\mathcal{C}_{y_3} - k_{on_3}\mathcal{C}_{y_1}P_z \\
\frac{d\mathcal{C}_{y_2}}{dt} &= k_{on_2}G_yP_z - k_{off_2}\mathcal{C}_{y_2} \\
\frac{d\mathcal{C}_{y_3}}{dt} &= k_{on_3}\mathcal{C}_{y_1}P_z - k_{off_3}\mathcal{C}_{y_3}
\end{aligned}$$

Since the regulatory sequence of Y to which proteins bind is conserved and in fact equal to one in the present case, from the above set of equations we get,

$$\begin{aligned}
\frac{dG_y}{dt} &= -\frac{d\mathcal{C}_{y_1}}{dt} - \frac{d\mathcal{C}_{y_2}}{dt} - \frac{d\mathcal{C}_{y_3}}{dt} \\
G_y(t) &= \alpha - \mathcal{C}_{y_1}(t) - \mathcal{C}_{y_2}(t) - \mathcal{C}_{y_3}(t)
\end{aligned}$$

where $\alpha = G_y(t_0)$. The variable G_y can now be substituted for in the deterministic dynamical equations of the other variables, which are,

$$\begin{aligned}
\frac{d\mathcal{C}_{y_1}}{dt} &= k_{on_1}P_x(\alpha - \mathcal{C}_{y_1} - \mathcal{C}_{y_2} - \mathcal{C}_{y_3}) - k_{off_1}\mathcal{C}_{y_1} + k_{off_3}\mathcal{C}_{y_3} - k_{on_3}\mathcal{C}_{y_1}P_z \\
\frac{d\mathcal{C}_{y_2}}{dt} &= k_{on_2}P_z(\alpha - \mathcal{C}_{y_1} - \mathcal{C}_{y_2} - \mathcal{C}_{y_3}) - k_{off_2}\mathcal{C}_{y_2} \\
\frac{d\mathcal{C}_{y_3}}{dt} &= k_{on_3}\mathcal{C}_{y_1}P_z - k_{off_3}\mathcal{C}_{y_3} \\
\frac{dM_x}{dt} &= k^+_{M_x} - k^-_{M_x}M_x \\
\frac{dM_z}{dt} &= k^+_{M_z} - k^-_{M_z}M_z \\
\frac{dM_y}{dt} &= k^+_{M_{y_1}}\mathcal{C}_{y_1} + k^+_{M_{y_2}}\mathcal{C}_{y_2} + k^+_{M_{y_3}}\mathcal{C}_{y_3} - k^-_{M_y}M_y \\
\frac{dP_x}{dt} &= k_{off_1}\mathcal{C}_{y_1} - k_{on_1}P_x(\alpha - \mathcal{C}_{y_1} - \mathcal{C}_{y_2} - \mathcal{C}_{y_3}) + k^+_{P_x}M_x - k^-_{P_x}P_x \\
\frac{dP_z}{dt} &= k_{off_2}\mathcal{C}_{y_2} - k_{on_2}P_z(\alpha - \mathcal{C}_{y_1} - \mathcal{C}_{y_2} - \mathcal{C}_{y_3}) + k_{off_3}\mathcal{C}_{y_3} - k_{on_3}\mathcal{C}_{y_1}P_z + k^+_{P_z}M_z - k^-_{P_z}P_z \\
\frac{dP_y}{dt} &= k^+_{P_y}M_y - k^-_{P_y}P_y
\end{aligned}$$

In the above set of variables, the TF-DNA complexes that are involved in the fast binding/unbinding events, can be eliminated at steady state without affecting the correlation properties of the system. This was shown in section 5.1.2 and also in the case of the cascade network. If $K_{D_i} = k_{off_i}/k_{on_i}$ is the TF-DNA dissociation constant, on solving

for the complexes $\mathcal{C}_{y_i}$ at steady state we get the equalities,

$$\begin{aligned} P_x(\alpha - \mathcal{C}_{y_1} - \mathcal{C}_{y_2} - \mathcal{C}_{y_3}) &= K_{D_1}\mathcal{C}_{y_1} \\ P_z(\alpha - \mathcal{C}_{y_1} - \mathcal{C}_{y_2} - \mathcal{C}_{y_3}) &= K_{D_2}\mathcal{C}_{y_2}, \end{aligned} \tag{6.5}$$

from which the expressions for $\mathcal{C}_{y_2}$ is obtained as,

$$\mathcal{C}_{y_2} = \frac{P_z}{K_{D_2}}\frac{K_{D_1}}{P_x}\mathcal{C}_{y_1}.$$

The complex $\mathcal{C}_{y_3}$ at steady-state is,

$$\mathcal{C}_{y_3} = \frac{P_z}{K_{D_3}}\mathcal{C}_{y_1}$$

Substituting for the above expressions for $\mathcal{C}_{y_2}$ and $\mathcal{C}_{y_3}$ in Equation (6.5), we get,

$$P_x(\alpha - \mathcal{C}_{y_1} - \frac{P_z}{K_{D_2}}\frac{K_{D_1}}{P_x}\mathcal{C}_{y_1} - \frac{P_z}{K_{D_3}}\mathcal{C}_{y_1}) = K_{D_1}\mathcal{C}_{y_1}$$

which on solving gives the expressions for the three complexes at steady-state as,

$$\begin{aligned} \mathcal{C}_{y_1} &= \frac{\alpha\frac{P_x}{K_{D_1}}}{\left[1 + \frac{P_x}{K_{D_1}} + \frac{P_z}{K_{D_2}} + \frac{P_xP_z}{K_{D_1}K_{D_3}}\right]} \\ \mathcal{C}_{y_2} &= \frac{\alpha\frac{P_z}{K_{D_2}}}{\left[1 + \frac{P_x}{K_{D_1}} + \frac{P_z}{K_{D_2}} + \frac{P_xP_z}{K_{D_1}K_{D_3}}\right]} \\ \mathcal{C}_{y_3} &= \frac{\alpha\frac{P_xP_z}{K_{D_1}K_{D_3}}}{\left[1 + \frac{P_x}{K_{D_1}} + \frac{P_z}{K_{D_2}} + \frac{P_xP_z}{K_{D_1}K_{D_3}}\right]} \end{aligned}$$

On removing the dynamical equations corresponding to these complexes and on substituting the above expressions in the equations for the mRNAs and proteins, we get a completely reduced system of equations where all the molecular species operate in the same range of time-scales, and more importantly, where the correlations between these species are the same as in the case of the original unreduced system. The new set of

dynamical equations are therefore,

$$\begin{aligned}
\frac{dM_x}{dt} &= k^+_{M_x} - k^-_{M_x} M_x \\
\frac{dM_z}{dt} &= k^+_{M_z} - k^-_{M_z} M_z \\
\frac{dM_y}{dt} &= k^+_{M_{y_1}} \frac{\alpha \frac{P_x}{K_{D_1}}}{\left[1 + \frac{P_x}{K_{D_1}} + \frac{P_z}{K_{D_2}} + \frac{P_x P_z}{K_{D_1} K_{D_3}}\right]} + k^+_{M_{y_2}} \frac{\alpha \frac{P_z}{K_{D_2}}}{\left[1 + \frac{P_x}{K_{D_1}} + \frac{P_z}{K_{D_2}} + \frac{P_x P_z}{K_{D_1} K_{D_3}}\right]} \\
&\quad + k^+_{M_{y_3}} \frac{\alpha \frac{P_x P_z}{K_{D_1} K_{D_3}}}{\left[1 + \frac{P_x}{K_{D_1}} + \frac{P_z}{K_{D_2}} + \frac{P_x P_z}{K_{D_1} K_{D_3}}\right]} - k^-_{M_y} M_y \\
\frac{dP_x}{dt} &= k^+_{P_x} M_x - k^-_{P_x} P_x \\
\frac{dP_z}{dt} &= k^+_{P_z} M_z - k^-_{P_z} P_z \\
\frac{dP_y}{dt} &= k^+_{P_y} M_y - k^-_{P_y} P_y
\end{aligned}$$

From the above rate equations we derive the Jacobian matrix, which on re-arranging in the order corresponding to $[M_x, P_x, M_z, P_z, M_y, P_y]$, is the following lower triangular matrix with the decay rates on the diagonal and hence are the eigenvalues of the system,

$$\mathbf{A} = \begin{pmatrix}
-k^-_{M_x} & 0 & 0 & 0 & 0 & 0 \\
k^+_{P_x} & -k^-_{M_z} & 0 & 0 & 0 & 0 \\
0 & 0 & -k^-_{M_y} & 0 & 0 & 0 \\
0 & 0 & k^+_{P_z} & -k^-_{P_x} & 0 & 0 \\
0 & A_{52} & 0 & A_{54} & -k^-_{P_z} & 0 \\
0 & 0 & 0 & 0 & k^+_{P_y} & -k^-_{P_y}
\end{pmatrix} \tag{6.6}$$

where the off-diagonal elements corresponding to M_y are,

$$\begin{aligned}
A_{52} &= \frac{\partial(dM_y/dt)}{\partial P_x} = \Big(k^+_{M_{y_1}} \alpha \Big[\frac{1}{K_{D_1}} + \frac{P_z}{K_{D_1} K_{D_2}}\Big] - k^+_{M_{y_2}} \alpha \Big[\frac{P_z}{K_{D_1} K_{D_2}} + \frac{P_z^2}{K_{D_1} K_{D_2} K_{D_3}}\Big] \\
&\quad + k^+_{M_{y_3}} \alpha \Big[\frac{P_z}{K_{D_1} K_{D_3}} + \frac{P_z^2}{K_{D_1} K_{D_2} K_{D_3}}\Big]\Big) \Big/ \Big[1 + \frac{P_x}{K_{D_1}} + \frac{P_z}{K_{D_2}} + \frac{P_x P_z}{K_{D_1} K_{D_3}}\Big]^2 \\
A_{54} &= \frac{\partial(dM_y/dt)}{\partial P_z} = \Big(- k^+_{M_{y_1}} \alpha \Big[\frac{P_x}{K_{D_1} K_{D_2}} + \frac{P_x^2}{K_{D_1}^2 K_{D_3}}\Big] + k^+_{M_{y_2}} \alpha \Big[\frac{1}{K_{D_2}} + \frac{P_x}{K_{D_1} K_{D_2}}\Big] \\
&\quad + k^+_{M_{y_3}} \alpha \Big[\frac{P_x}{K_{D_1} K_{D_3}} + \frac{P_x^2}{K_{D_1}^2 K_{D_3}}\Big]\Big) \Big/ \Big[1 + \frac{P_x}{K_{D_1}} + \frac{P_z}{K_{D_2}} + \frac{P_x P_z}{K_{D_1} K_{D_3}}\Big]^2
\end{aligned}$$

The vector of deterministic rates and the stoichiometry of the reduced system are,

$$\mathbf{R} = \Bigg(k^+_{M_x}, k^+_{M_z}, k^+_{M_{y_1}} \frac{\alpha \frac{P_x}{K_{D_1}}}{\left[1 + \frac{P_x}{K_{D_1}} + \frac{P_z}{K_{D_2}} + \frac{P_x P_z}{K_{D_1} K_{D_3}}\right]}, k^+_{M_{y_2}} \frac{\alpha \frac{P_z}{K_{D_2}}}{\left[1 + \frac{P_x}{K_{D_1}} + \frac{P_z}{K_{D_2}} + \frac{P_x P_z}{K_{D_1} K_{D_3}}\right]}, k^+_{M_{y_3}} \frac{\alpha \frac{P_x P_z}{K_{D_1} K_{D_3}}}{\left[1 + \frac{P_x}{K_{D_1}} + \frac{P_z}{K_{D_2}} + \frac{P_x P_z}{K_{D_1} K_{D_3}}\right]},$$
$$k^+_{P_x} M_x, k^+_{P_z} M_z, k^+_{P_y} M_y, k^-_{M_x} M_x, k^-_{M_z} M_z, k^-_{M_y} M_y, k^-_{P_x} P_x, k^-_{P_z} P_z, k^-_{P_y} P_y \Bigg)^T$$

$$\boldsymbol{\nu} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & -1 \end{pmatrix}$$

The above terms are used for calculating the stationary and then the dynamic correlations between the proteins, and are shown in Figure 6.12. The values for the transcription rates $k^+_{M_{y_i}}$ are 0.3 min^{-1}, 0.1 min^{-1} and 0.6 min^{-1} respectively, and are chosen such that the mean steady state value of M_y is 4.5 nM. The mean values of M_z and P_z are assumed to be 0.3 nM and 240 nM respectively, which are in the range of the values for M_x and P_x such that they compete fairly for binding to the regulatory region of G_y. From the Figure 6.12, we notice that there is a slight decrease in the magnitude of the correlations between P_x and P_y whilst its temporal character remains the same as in the case of the single activator network $X \rightarrow Y$. We call this as temporal independency between the regulatory activity of the two TFs. Therefore, if a synthetic regulatory network needs to be designed, where control is desired only over the magnitude of fluctuations with the time-characteristics remaining unaltered, then the above network mechanism serves the desired purpose.

The effect that the decay rates of M_y and P_y have on $\mathrm{Corr}[P_x(t), P_y(t+\tau)]$ is the same as in the $X \rightarrow Y$ network. However, the decay rates of M_z and P_z do affect this correlation function. While the magnitude of the stationary and dynamic covariances is insensitive to variations in these parameters, the correlations do vary in magnitude. This is due to the sensitivity of the normalizing auto-covariance functions. On the other hand, though these decay rates $k^-_{M_z}$ and $k^-_{P_z}$ are the eigenvalues of the system, they do not influence the temporal feature τ^* of the correlations. This is verified by evaluating $\frac{\partial \tau^*}{\partial k^-_{M_z}}$ which turns out to be insignificant ($= 1.83$). Similar is the case for $\frac{\partial \tau^*}{\partial k^-_{P_z}} = 6.67$. This is due to the reduced values of the terms $\frac{\partial F}{\partial k^-_{M_z}}$ and $\frac{\partial F}{\partial k^-_{P_z}}$ that are derived in section 4.2 of the

chapter on sensitivities. Therefore, the above scheme of activation de-sensitizes τ^* and brings about variation only in the correlation magnitudes.

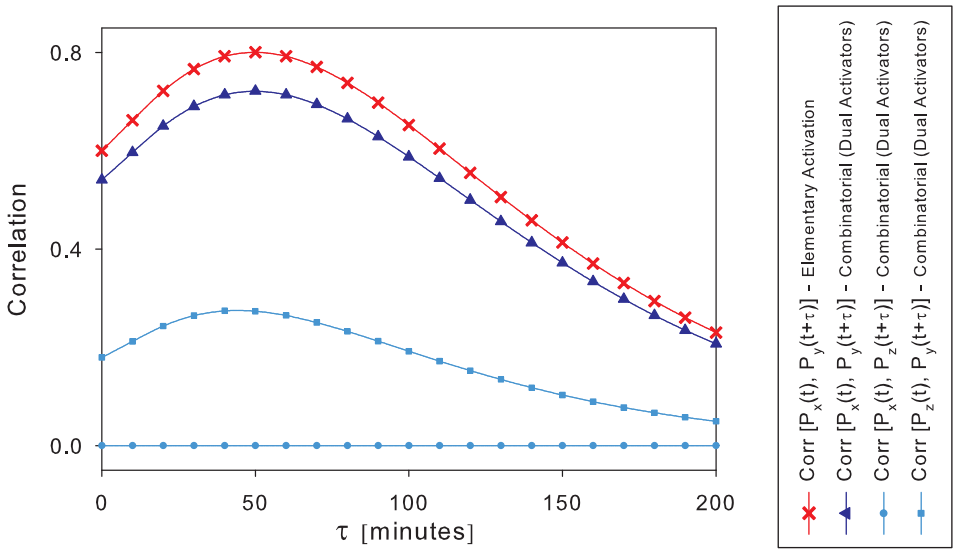

FIGURE 6.12: The dynamic correlations between the proteins of the combinatorial (dual activator) network are shown in blue. The network scheme is $X \rightarrow Y \leftarrow Z$. For sake of comparison, protein correlations in the case of $X \rightarrow Y$ are shown in red. The mean values are: $\langle M_x \rangle = 0.2\ nM, \langle M_z \rangle = 0.3\ nM, \langle M_y \rangle = 4.5\ nM, \langle P_x \rangle = 395\ nM, \langle P_z \rangle = 240\ nM, \langle P_y \rangle = 36001\ nM$. The corresponding parameter values are: $k^+_{M_x} = 0.008\ nM\ min^{-1}, k^+_{M_z} = 0.028\ nM\ min^{-1}, k^+_{M_{y_1}} = 0.3\ min^{-1}, k^+_{M_{y_2}} = 0.1\ min^{-1}, k^+_{M_{y_3}} = 0.656\ min^{-1}, k^+_{P_x} = 36\ min^{-1}, k^+_{P_z} = 20\ min^{-1}, k^+_{P_y} = 143\ min^{-1}, k^-_{M_x} = 0.0407\ min^{-1}, k^-_{M_z} = 0.0927\ min^{-1}, k^-_{M_y} = 0.077\ min^{-1}, k^-_{P_x} = 0.0184\ min^{-1}, k^-_{P_z} = 0.025\ min^{-1}, k^-_{P_y} = 0.0176\ min^{-1}$. The TF-DNA dissociation constants are $K_{D_1} = K_{D_2} = K_{D_3} = 200\ nM$.

6.2.1.1 Activators turn into Repressors

Though it seems quite logical and straightforward, from the reaction scheme of the $X \rightarrow Y \leftarrow Z$ regulatory network, that the proteins P_x and P_z behave as activators in the transcription of M_y, we show that there is a very interesting behaviour of this network contrary to the above logic. In Figure 6.13, we observe that as the value of the transcription rate $k^+_{M_{y_2}}$ increases, the stationary as well as dynamic correlations between P_x and P_y change sign, which means that P_x starts behaving like a repressor and down-regulates the production of P_y for higher values of the transcription rate corresponding to the other TF P_z. Such a behaviour is perplexing, but from the expressions of A_{52}, which is the Jacobian element in equation (6.6), we notice that there is a change of sign in its value as $k^+_{M_{y_2}}$ crosses a threshold value. Since A_{52} is an off-diagonal element of a triangular matrix, the eigenvectors of this matrix change sign once $k^+_{M_{y_2}}$ crosses a threshold value of 0.9 min^{-1}.

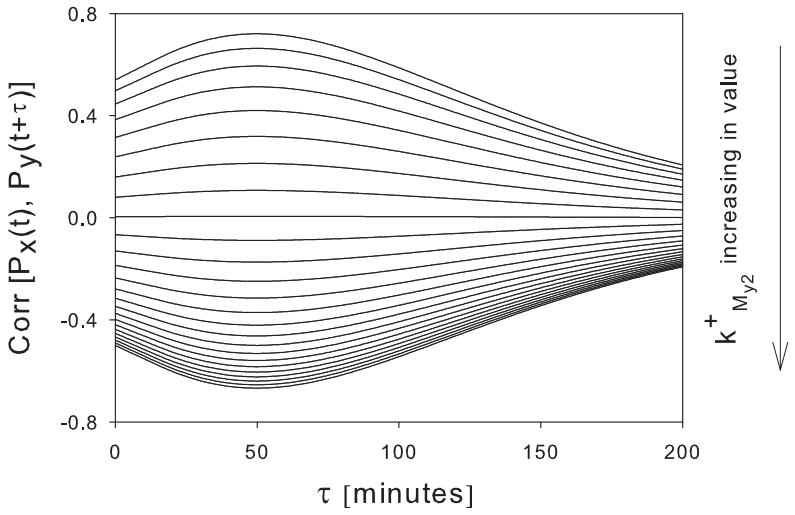

FIGURE 6.13: Correlations between P_x and P_y for changes in the transcription rate $k^+_{M_{y_2}}$.

Since the dynamic covariance function is a sum of exponentials whose co-efficients are functions of these eigenvectors, the covariances and consequently the correlations change signs for sign-changes in the eigenvectors. This is a mathematical explanation for the strange behaviour of this type of regulation. The physical reasoning for this is as follows. At a high value of $k^+_{M_{y_2}}$, P_y is predominatly being produced due to the action of P_z and not much by the action of P_x. Under such a condition, when there is a positive perturbation induced in the mean steady-state value of P_x, the mean concentration levels of the complexes $\mathcal{C}_{y_1}$ and $\mathcal{C}_{y_3}$ increases, whereas the concentration of $\mathcal{C}_{y_2}$ decreases due to the conservation of the DNA molecules. Now, as the transcription rate of the complexes $\mathcal{C}_{y_1}$ and $\mathcal{C}_{y_3}$ remain unaltered while that of $\mathcal{C}_{y_2}$ is high, the amount of M_y and hence of P_y decreases for sudden increase in the mean steady-state value of P_x.

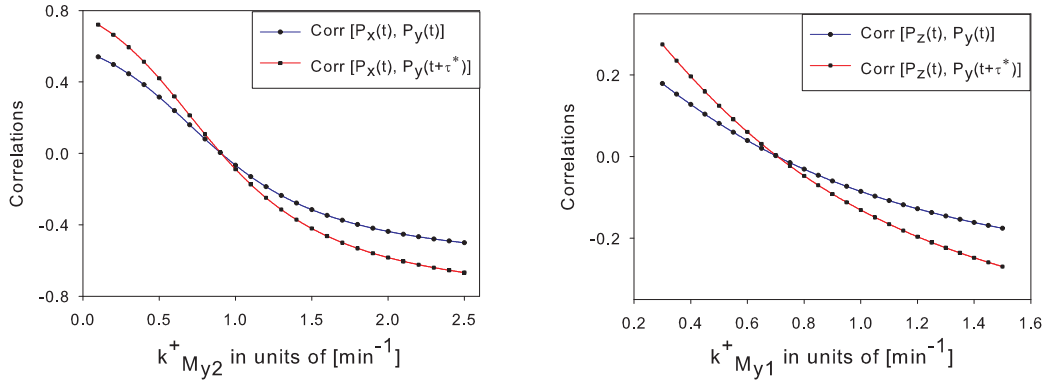

FIGURE 6.14: Variations in the stationary and dynamic correlations between P_x and P_y, for increase in the transcription rate $k^+_{M_{y_2}}$.

In Figure 6.13 is shown the role reversal of P_x from an activator to a that of a repressor for increase in the transcription rate $k^+_{M_{y_2}}$. The variation in the stationary correlations (at $\tau = 0$) and the correlations at τ^*, for increased values of the transcription rates is shown in Figure 6.14.

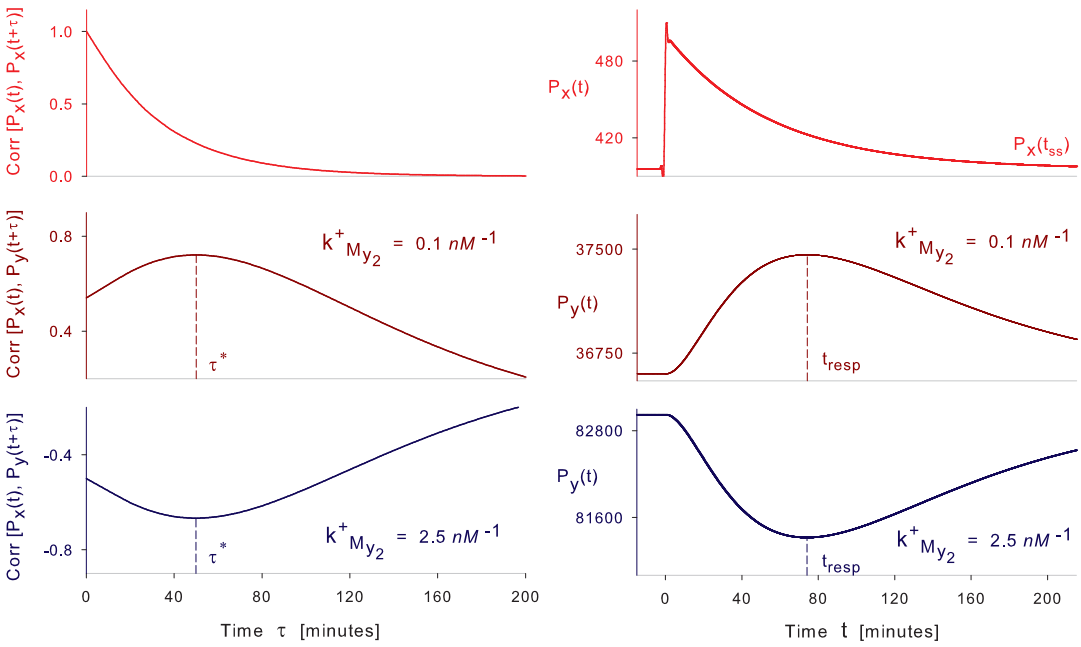

FIGURE 6.15: On the left is shown the correlation between the proteins P_x and P_y for two different values of $k^+_{M_{y_2}}$. On the right is the deterministic response in P_y for perturbation in P_x from its mean steady-state value of 395 nM to 495 nM.

6.2.2 Activator and Repressor

There exist mechanisms of gene regulation which involve the simultaneous action of activators and repressors on the regulatory region of the regulated DNA. Here, we once again assume the presence of only one regulatory region upstream of G_y, for which there is competitive binding of the activator and repressor molecules. The network topology of such a mechanism is $X \rightarrow Y \vdash Z$, where X is the activator, Y being the repressor and Z is the regulated gene-node.

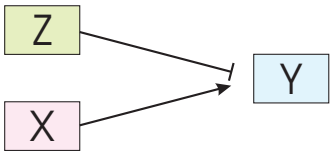

FIGURE 6.16: Network representation of combinatorial (activator and repressor) regulation, where the activator P_x and the repressor P_z both bind to the same *cis*-regulatory sequence of G_y.

The reaction scheme for this regulatory network consists of the following binding/unbinding events:

$$P_x + G_y \underset{k_{off_1}}{\overset{k_{on_1}}{\rightleftharpoons}} \mathcal{C}_{y_1}$$

$$P_z + G_y \underset{k_{off_2}}{\overset{k_{on_2}}{\rightleftharpoons}} \mathcal{C}_{y_2}$$

$$\mathcal{C}_{y_1} \xrightarrow{k^+_{M_y}} \mathcal{C}_{y_1} + M_y$$

For such a mechanism it is interesting to note that the nature of regulatory link between the proteins P_x and P_y which is that of activation ($\rightarrow$) and between P_z and P_y which is that of repression ($\vdash$), are faithfully revealed by their respective correlation functions, as shown in Figure 6.17. This is in spite of their competitive binding on to the same operator region. Further, we analyze the sensitivity of the *features* of the correlations for variations in some of the decay rates. For variations in the decay rates of the mRNA or the protein of the repressor gene-node, the τ^* of the correlations between the activator and the regulated node, which is $\mathrm{Corr}[P_x(t), P_y(t+\tau)]$, remains unaltered at 49 minutes. However the stationary correlation and Corr^* show variations as in the case of the dual activator network. Similar is the case for the correlation between the repressor P_z and P_y, for variations in the decay rates of the mRNA and protein of X. We therefore conclude that the overall *positive* nature of the correlations between an *activator* and its regulated protein is unaffected by the competitive binding of a *repressor* on to the same operator region, and *vice versa*.

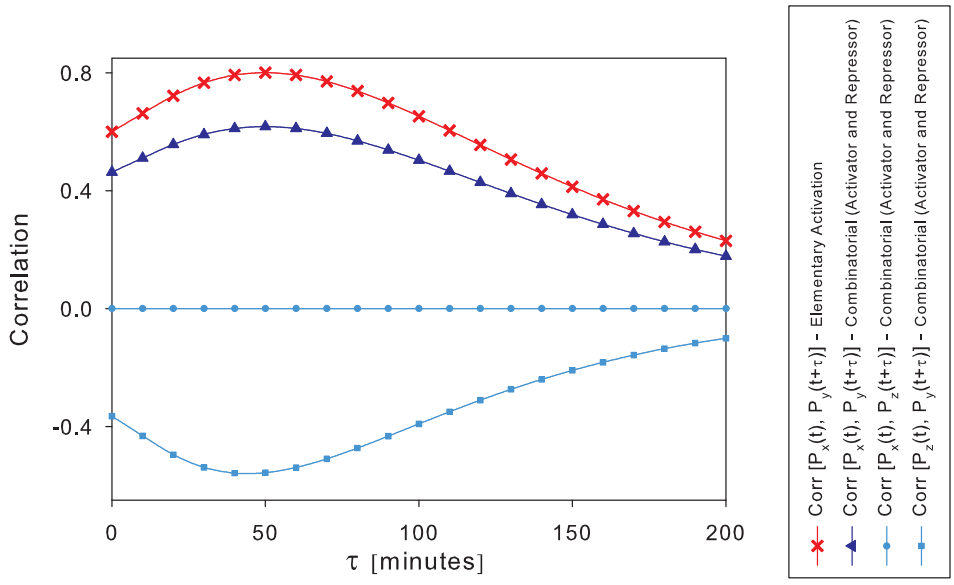

FIGURE 6.17: The dynamic correlations between the proteins of the combinatorial (activator and repressor) network are shown in blue. The network scheme is $X \rightarrow Y \vdash Z$. Protein correlations in the case of the $X \rightarrow Y$ network are shown in red.

An interesting observation in this regulatory mechanism, as well as in the case of the dual activator network, is that for increase in the value of dissociation constant of one TF, say $K_{D_2} = k_{off_2}/k_{on_2}$, the magnitude of the correlation between P_x and P_y increases. Further, this magnitude whether at $\tau = 0$ or τ^* becomes increasingly insensitive to changes in the decay rates of M_z and P_z, as K_{D_2} increases. This is because the relation between P_x and P_y now stands strengthened due to the reduction in the repressor-DNA complex $\mathcal{C}_{y_2}$ as a result of increase in K_{D_2}.

6.3 FeedForward Loops

Another class of well-defined regulatory networks are the feedforward loops consisting of three genes where the downstream gene is regulated by two different TFs, out of which one TF acts as the regulator of the other TF. Thus a loop is formed. The type of regulation can either be that of activation or repression. If for example, both the TFs act as activators of the downstream protein, such a network is called as the coherent feedforward loop. On the other hand, if one is a repressor while the other an activator, the incoherent FFL is formed. It has been reported that the feedforward loop is a recurring network in organisms such as both *Escherichia coli* and the yeast Shen-Orr et al. (2002); Kalir et al. (2005); Mangan et al. (2006). The biological advantages of such loops over the previously described forms of regulation may be difficult to estimate, but however, on modelling these set of networks deterministically or stochastically the possible use of these networks in the functioning of real networks could be elucidated.

6.3.1 Coherent FFL

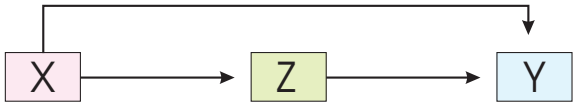

FIGURE 6.18: Network structure of the Coherent Feedforward Loop.

As shown in Figure 6.18, the coherent FFL, where the regulatory function of both the paths from the gene-node X to Y are that of activation, is formed by combining two of our previously studied networks, the cascading activator and the combinatorial dual activator networks. The exact set of transcriptional reactions between the two TFs and Y determines whether the system functions with an AND-logic or an OR-logic. Since the combinatorial network that we investigated was regulated by either of the two activators, in the coherent FFL, we shall follow the same OR-logic. The only changes that need to be done to the Jacobian matrix of the combinatorial network, given in Equation (6.6), is

the addition of an extra element relating to the activation of Z by X. This the element,

$$A_{32} = \frac{\partial(dM_z/dt)}{\partial P_x} = \frac{k^+_{M_z}\frac{\beta}{K_{D_4}}}{\left[1+\frac{P_x}{K_{D_4}}\right]^2}$$

where K_{D_4} is the dissociation constant of P_x binding/unbinding to G_z and its value is assumed to be the same as that of other binding/unbinding processes (200 nM). Similarly, the additional deterministic reaction rate is $k^+_{M_z}\beta\frac{P_x}{K_{D_4}}/\left[1+\frac{P_x}{K_{D_4}}\right]$. We observe that there are some differences in the *features* of the dynamic correlations of this network and of the *combinatorial dual activator* network (Figure 6.12). Firstly, the magnitudes of all the three proteins correlations increase in value with the obvious change being in the Corr$[P_x(t), P_y(t+\tau)]$ function that is now non-zero and positive. Secondly, the time-delay τ^* of peak value of the Corr$[P_x(t), P_y(t+\tau)]$ function increases from 49 to 55 minutes. On the other hand, τ^* of the Corr$[P_z(t), P_y(t+\tau)]$ function decreases from 43 to 34 minutes. Further, due to the network connectivity, there is an advantage over the combinatorial network in the sense that there is now a control over the production of P_z by P_x, which means that variations in the decay rate of P_z or the transcription rate of M_z have an influence on the correlations between P_x and P_y. Additionally from the viewpoint of the *cascading activator* network, where τ^* for the Corr$[P_x(t), P_y(t+\tau)]$ function was around 97 minutes, it now reduces significantly to 55 minutes, nearly equivalent to that of the *elementary activator* network. Therefore the correlations between the *input* and *output* elements of the coherent FFL display a combination of behaviours that are peculiar to the sub-networks, out of which it is made.

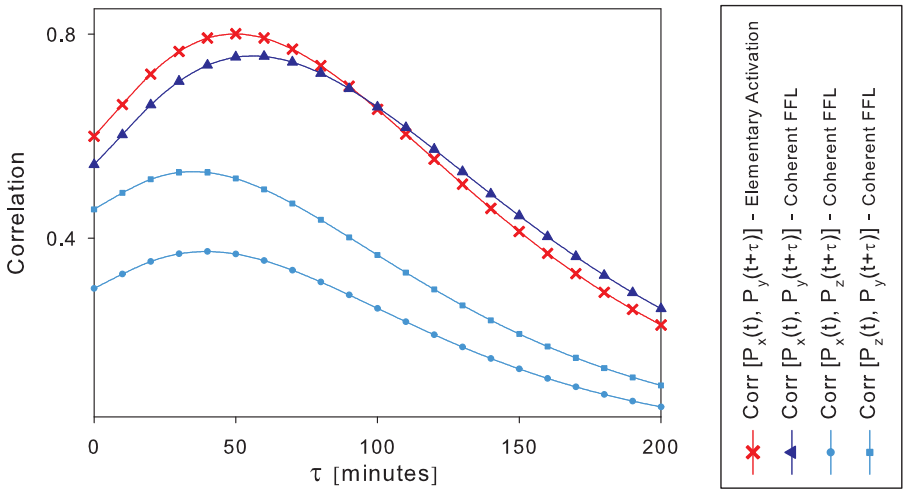

FIGURE 6.19: The dynamic correlations between the proteins of the coherent feedforward loop.

The more interesting behaviour of the coherent FFL is in the way the Corr$[P_x(t), P_y(t+\tau)]$ function behaves for changes in $k^+_{M_{y2}}$, which is the transcription rate of M_y from P_z. However, the threshold value of this parameter, at which the correlation function changes sign, is now 1.4 min^{-1} instead of that in the combinatorial dual activator case where it was 0.9 min^{-1}. The correlation function at this threshold value has a unique property of further changing signs after a time-delay of 95 minutes. This property is one way of characterizing this network. The biological implication of such a behaviour is that the coherent FFL can be used to induce a delayed response in the downstream element for a perturbation in the top most element of the network. Further, in Figure 6.20 we show that for a ten-fold increase in the transcription rate of M_z, this threshold value reduces to 0.75 min^{-1} and is less steep than in the case where $k^+_{M_z}$ was lower. Therefore, dynamic correlations between these proteins reveal a lot about the network structure of such regulatory systems and their potential uses in designing complex networks.

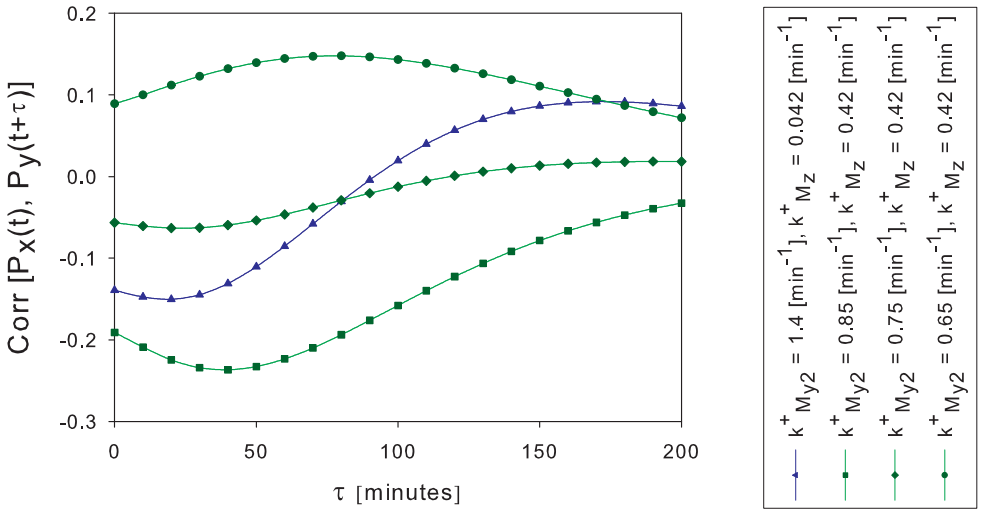

FIGURE 6.20: The dynamic correlations between the proteins of the coherent feedforward loop for changes in the transcription rates.

6.3.2 InCoherent FFL

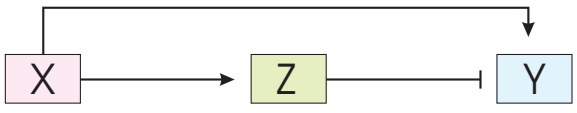

FIGURE 6.21: Network structure of the InCoherent Feedforward Loop.

As shown in the above Figure 6.21, the incoherent FFL is formed by combining the cascading repressor and the combinatorial (activator and regulator) networks. Whilst one path of regulation between the gene-nodes X and Y is that of activation, the other path

is that of repression, hence the incoherency. However, this network displays an interesting behaviour in its protein correlations. The time instant τ^*, at which the dynamic correlations between the proteins of X and Y nodes attains maximum value, is shifted back to 32 minutes from that of the either sub-networks. Therefore such a combination of the smaller sub-networks gives rise to behaviours in the internal fluctuations, by which the larger network can be characterized.

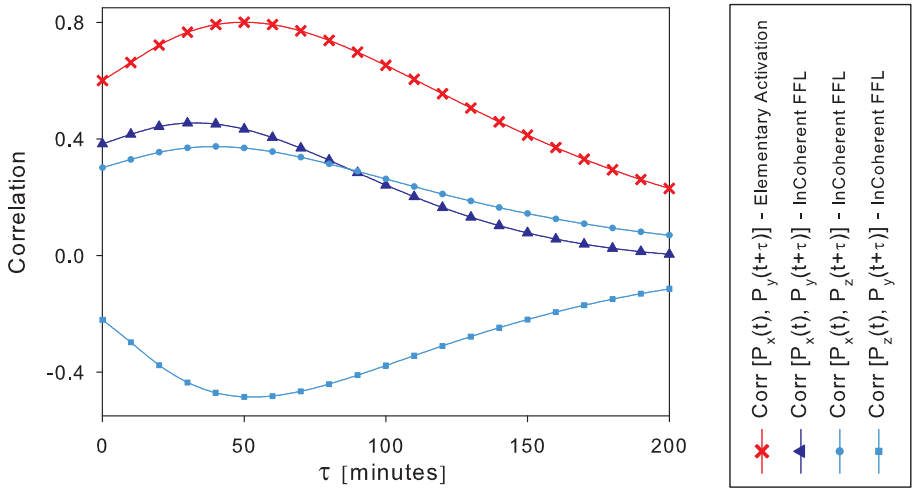

FIGURE 6.22: The dynamic correlations between the proteins of the incoherent feedforward loop.

6.4 Summary

In this chapter, we showed the effectiveness of the dynamic correlations between proteins, in characterizing different network mechanisms based on their internal fluctuations. In the cascade network, the *feature* that stood out was τ^* that was twice the value than that of a simple two-gene network. Further, the magnitudes of correlations also varied for changes in the regulatory function of the cascade. We then discussed in detail the network where two transcription factors regulated the production of a downstream protein. An interesting behaviour was observed in the case of dual activators simultaneously activating a downstream gene. Under varying transcription rates of one activator, the other activator began functioning as a repressor. Finally, we showed that on combining the above network mechanisms, the resulting network such as feedforward loops displayed a variety of behaviours in their internal fluctuations as captured by the dynamic correlations.

In conclusion, we saw how the dynamic correlation functions vary for changes in the essential *factors* of the gene regulatory networks such as the values of the rate constants, the form and type of the regulatory mechanim and the connectivity between the genes

of the network. As a conclusion to the above results of chapters 5 and 6, in Figures 6.23 and 6.24 we provide simple demonstrations of the effectiveness of the dynamic correlations in characterizing small gene regulatory networks. The two most significant *features* of the dynamic correlation functions or plots, which are τ^* and Corr* are plotted in the case of different networks. Clearly, the features enable in identifying the structure of the network. Each cluster of points corresponds to a particualr network. The points within each cluster are obtained by varying the decay rate of P_y. For example, in Figure 6.23 the clusters corresponding to the cascading activator networks have a larger value of τ^*, whilst differing amongst each other in respect of the magnitude of peak correlations Corr*. Similarly, the incoherent and the coherent feedforward loops are clearly identifiable. Finally, the effect of dimerization is clearly noticed with increased Corr* values. Figure 6.24 is a contrasting picture where the correlation features vary in a differing manner, in the case of these networks, for changes in the value of transcription rate. Continuing on similar lines, clusters can be generated in a high dimensional space by varying all the reaction parameters, thereby demonstarting that the dynamic correlations of molecular fluctuations are effective in characterizing small gene regulatory networks. Due to changes in the shape of these correlation plots, the use of the integral over time τ is also a matter that needs to be probed further.

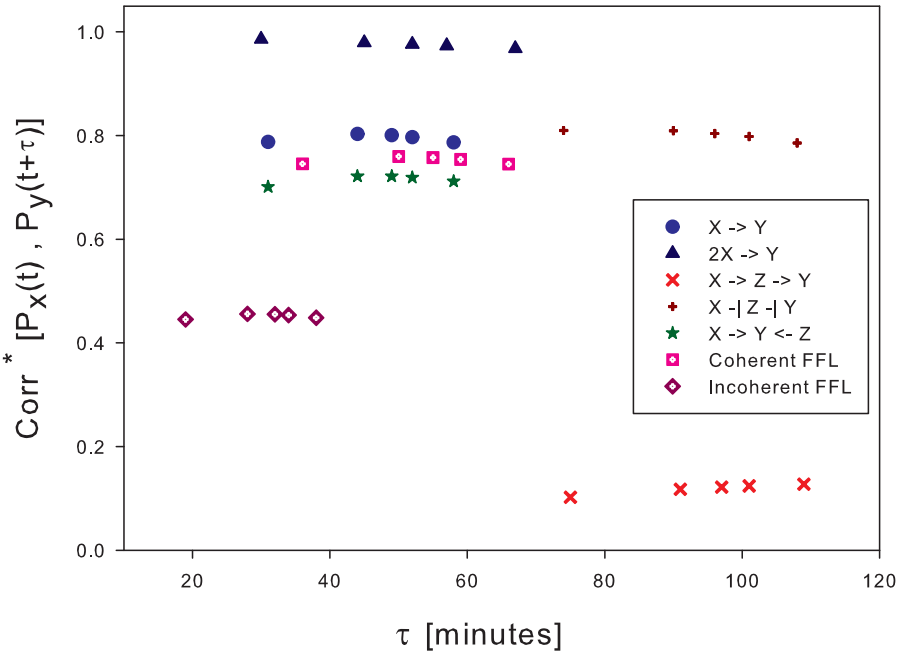

FIGURE 6.23: In the above figure we show how the *features* of the dynamic correlation functions form clusters specific to different regulatory networks. The points in each cluster are obtained by varying the half-life of the protein P_y as $[20, 50, 70, 90, 150]$ minutes respectively. The decay rate of P_y is related to the half-life as shown in equation (5.7).

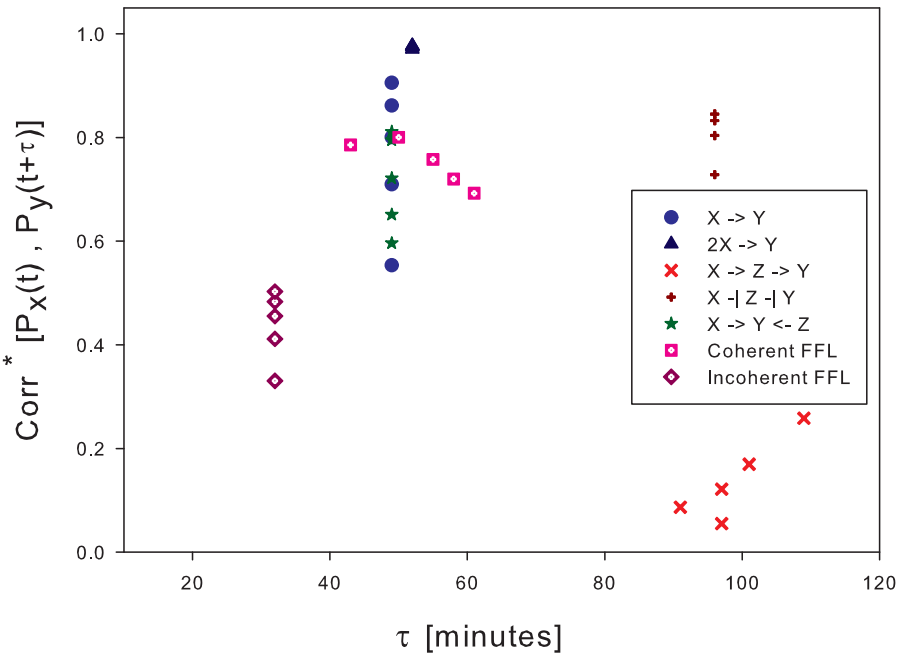

FIGURE 6.24: In the above figure we show how the *features* of the dynamic correlation functions form clusters specific to different regulatory networks. The points in each cluster are obtained by multiplyting the base value of the transcription rate of M_y with $0.2, 0.5, 1, 2, 5]$ respectively. In the case where the regulatory mechanism is that of repression, this transcription rate is the basal rate.

Chapter 7

Conclusions

In order to understand the functioning of a complex gene regulatory network, it is vital to investigate the properties of its building blocks, which are smaller networks having well-defined characteristics. The idea is to develop a better understanding of the *parts* before proceeding to the *system* level Sprinzak and Elowitz (2005); Guido et al. (2006). Such an understanding would improve our ability to recognize the form of the regulatory activity and the network structure in these GRNs by analyzing the protein and mRNA levels. Though GRNs have previously been analyzed in the deterministic or macroscopic realm, here, we follow a stochastic approach where the microscopic behaviour of these networks is revealed. This is due to the effectiveness of such an approach in characterizing the GRNs and also due to the issues concerning the effectiveness of data obtained on averaging from a population of cells. The latter issue was discussed in Chapter 1. Due to the significance of small molecular numbers in generating fluctuations in the levels of the species, stochastic methods are better off at explaining the behaviour of GRNs. The issues regarding stochastic modelling were elaborated in Chapter 2.

The stationary and dynamic covariances between the molecular species are used as indicators for characterizing GRNs. Towards this, our aim in this thesis has been to investigate whether these indicators do indeed reveal different characteristics for GRNs that differ in their regulatory and network mechanisms. In Chapter 3, we derived the dynamic covariances in terms of the eigenvalues and eigenvectors of the system Jacobian, thereby showing that the properties of the fluctuations can be derived by simply knowing the deterministic dynamical behaviour of the system. This formed the basis of the fluctuation-dissipation relationship. Using the analytical framework of Chapter 3 we showed that variations are induced in the internal fluctuations of GRNs, for different values of the reaction rate constants and different forms of gene regulation. Such a framework serves the dual purpose of characterizing different GRNs based on their fluctuation properties, as well as retrieving information about the type and form of regulation between two genes of a network, once such characterization is complete.

In Chapter 5 we demonstrated how the defining *features* such as the stationary correlations $\mathrm{Corr}(\tau = 0)$, the peak correlations Corr^* and the time τ^* at which this peak occurs, depend on the type of regulatory mechanism and on the reaction rate constants. We also showed the ineffectiveness of co-operative activation in altering the dynamic feature of the correlations, whilst increasing the magnitude of correlations between the regulator and regulated proteins. In Chapter 6 we showed that if one had information about the lifetimes of proteins and mRNAs of the two gene-nodes X and Y, the dynamic correlations would infer the presence or absence of an *intermediary* gene. In the case of combinatorial regulation by two TFs on the same regulatory region of a downstream gene, we showed that an activating TF switches its role to that of a repressor under certain varying conditions. This behaviour is faithfully revealed in the dynamic correlations between the proteins. Such analyses proves the use of dynamic correlation functions in analyzing the internal fluctuation properties of GRNs for changes in their regulatory mechanisms and network structures.

Due to the high sensitivity of the correlations to certain reaction rate constants, we could also hope to extract more information about the underlying reactions or processes. In this regard, we demonstrated the effect that various reaction rate constants of the *elementary activator* have on the *features* of the dynamic correlations between the regulator and regulated proteins P_x and P_y. Therefore by adopting a suitable learning methodology, the framework of sensitivity analysis derived in Chapter 4 could be used to estimate the parameters of a regulatory system given the dynamic correlations between its species.

In conclusion, the analytical framework developed in this thesis is demonstrated to be useful in characterizing various small gene regulatory networks and thereby illuminating their fluctuation properties. As a continuation of the above work, regulatory networks formed by combining the networks discussed in this thesis can be studied. For example in section 6.3 on the FeedForward loops, we showed that the *coherent FFL* which is a combination of the *cascading activator* network and the *dual activator* network, reduces the peak correlation time τ^* from that of the cascade network. Similary, the *incoherent FFL* showed different characteristics than those of its *parts*. Continuing this way, one could scale up the network size to incorporate a variety of smaller GRNs. Also, the performance of these smaller networks could alter when they are embedded into a large network. In this respect, it should be noted that the networks considered here are modelled in isolation, the only enternal influence being the spontaneous production and decays of the molecules. Overall, our belief is that knowing the *parts* is crucial to understand the *system* as a whole. Further, networks involving positive and negative feedbacks that are bistable have to be analyzed under each of the stable states, since this is the pre-condition for the applicability of LNA Ziv et al. (2007).

Apart from gaining insights into the functioning of a regulatory network, the framework developed here would be of great help in retrieving the regulatory mechanism and the network structure, given experimental data required for such purposes. Fluorescent reporters have increased our focus towards tracking individual cells, which is a major shift from the microarray realm where bulk averages were the measured quantities. Technologies such as flow cytometry measure the relative fluorescence intensities of individual cells as they flow in a fluid stream thus enabling one to plot histograms of protein fluorescence distributions. Ozbudak et al. (2002); Elowitz et al. (2002); Raser and O'Shea (2004) observed such variations in protein levels for changes to parameters values such as the transcription and translation rates. Their experiments matched neatly with the theoretical analyses. Such analyses however concerned with the stationary noise in proteins and could tell little about the dynamics of the reactions causing the noise, and therefore such measurements that leave out the temporal aspect are of little use in tracking the dynamics of gene regulation.

On the other hand, technologies such as time-lapse microscopy where fluorescently tagged proteins could be tracked over time in individual cells, facilitate our understanding of the relation between the regulatory mechanism between genes and its correlative effect observed in species such as proteins and mRNAs. Variation in protein levels in human cells was observed by Sigal et al. (2006b), who tracked the fluorescently tagged proteins and concluded that the fluctuations varied in time. It was also observed that genes of the same pathway showed correlations between them. Therefore such time-lapse measurements in single cells resulting in time-series of individual molecular species could be of great use in estimating any presence of regulatory activity between the corresponding genes. The analytical framework presented in this thesis would then be an ideal tool in predicting the type of regulatory activity or even the reaction structure between two genes. Due to rapid technological advancements in tracking individual molecules over time in single cells Muzzey and van Oudenaarden (2009), the idea of tracking the causal dynamics between molecular species as revealed by the dynamic correlations, is fast becoming a reality. The work of Yu et al. (2006), where they track single molecules of yellow fluorescent protein (YFP) in living cells by fluorescence microscope, and other single molecule techniques such as Cai et al. (2006) also bring hope for the analytical techniques presented here to come to life.

Appendix A

Stochastic Simulation Algorithm

Let us consider a system with N reacting species $\{s_1, ..., s_N\}$ that react according to M reactions $\{r_1, ..., r_M\}$ within a small volume v at a constant temperature. The dynamical state of this system can be specified as $\mathbf{X(t)} \equiv (X_1(t), ..., X_N(t))$, where $X_i(t) \equiv$ the number of s_i molecules in the system at time t. The algorithm now, has to describe the time evolution of $\mathbf{X(t)}$ from a given initial state $\mathbf{X_0}$. The algorithm centers around the concept of a probability function known as the *propensity function* for each reaction r_μ,

$$\begin{aligned} a_\mu(\mathbf{X})dt \quad &\equiv \quad \text{the probability, given } \mathbf{X(t)} = \mathbf{X}, \text{ that one } r_\mu \\ &\qquad \text{reaction will occur in } v \text{ in the next time interval } dt \end{aligned} \tag{A.1}$$

This *propensity function* is the product of the reaction parameter k_μ and number of reactant combinations h_μ for each μ reaction, where,

$$\begin{aligned} k_\mu dt \quad &\equiv \quad \text{average probability, to first order in } dt, \text{that a particular combination of} \\ &\qquad r_\mu \text{ reactant molecules will react accordingly in the next time interval } dt. \\ h_\mu \quad &\equiv \quad \text{number of distinct molecules reactant combinations for reaction} \\ &\qquad r_\mu \text{ found to be present in } v \text{ at time } t. \end{aligned}$$

Hence, the *propensity function* is now given by,

$$a_\mu(\mathbf{X})dt \quad \equiv \quad h_\mu k_\mu dt \tag{A.2}$$

The *state-change vector* ν_μ is defined by,

$$\begin{aligned} \nu_{\mu i} \quad &\equiv \quad \text{the change in the number of } s_i \text{ molecules produced} \\ &\qquad \text{by one } r_\mu \text{ reaction } (\mu = 1, ..., M \ ; \ i = 1, ..., N) \end{aligned} \tag{A.3}$$

Now, the *propensity function* and the *state-change vector* together completely define each reaction r_μ. Equations (A.2) & (A.3) together imply that this dynamic process is a jump Markov process on an N-dimensional non-negative integer lattice. Now, since the master equation of this system describes how the probability $P(\mathbf{X}, t \mid \mathbf{X}_0, t_0)$ for the state of the system evolves in time, is becomes obvious that the CME is the differential form of the Chapman-Kolmogorov equation of the Markov process. Hence, the Chemical Master Equation is given by,

$$\begin{aligned} \frac{\partial}{\partial t} P(\mathbf{X}, t \mid \mathbf{X}_0, t_0) &= \sum_{\mu=1}^{M} [a_\mu(\mathbf{X} - \boldsymbol{\nu}_\mu) P(\mathbf{X} - \boldsymbol{\nu}_\mu, t \mid \mathbf{X}_0, t_0) \\ &\quad - a_\mu(\mathbf{X}) P(\mathbf{X}, t \mid \mathbf{X}_0, t_0)] \end{aligned} \tag{A.4}$$

With the aid of the equations (A.2) & (A.3) we can derive a term known as the *next-reaction probability density function*, given by,

$$\begin{aligned} P(\tau, \mu \mid \mathbf{X}, t) d\tau &\equiv \text{probability at time } t \text{ that the next reaction in } v \text{ will occur} \\ &\quad \text{in the differential time interval } [t + \tau, t + \tau + d\tau), \\ &\quad \text{and will be an } r_\mu \textit{ reaction}. \\ &= P_0(\tau) \; . \; h_\mu k_\mu dt \end{aligned} \tag{A.5}$$

where $P_0(\tau)$ is the probability that no reaction occurs in the time interval $[t, t + \tau)$. It can then be derived that $P_0(\tau) = exp\big[- \sum_{v=1}^{M} h_v k_v \tau \big]$ Gillespie (1976), which gives rise to,

$$P(\tau, \mu \mid \mathbf{X}, t) = exp\big[- \sum_{v=1}^{M} h_v k_v \tau \big] \; . \; a_\mu(\mathbf{X}) \tag{A.6}$$

The Stochastic Simulation Algorithm Gillespie (1976), is then described by the following steps:

Step 0 (initialization). Set the time $t = 0$. Specify initial values for the N variables $\{X_1, ...X_N\}$. Specify and store the M quantities k_μ and h_μ. Specify the sampling times $t_1 < t_2$....., and a stopping time t_{stop}.

Step 1. Generate one random pair (τ, μ) according to the joint probability density function $P(\tau, \mu \mid \mathbf{X}, t)$ using Monte Carlo techniques.

Step 2. Using the (τ, μ) obtained above, advance t by τ and change the system state $\mathbf{X}$ by $\boldsymbol{\nu}_\mu$. Then recalculte the propensity functions of the occuring reactions as necessary.

Step 3. If t has just advanced through one of the sampling times t_i, output the current molecular numbers of the reacting species, which is the state vector $\mathbf{X}$. If $t > t_{stop}$ or if

all reactants are consumed end the algorithm; or else return to *Step* 1.

Generating random pairs (τ, μ):

To generate a random pair accoring to the probability density function (A.6), one can adopt a simple Monte Carlo technique, commonly known as the *Direct Method*. It is based on the fact that any two-variable probability density function can be written as the product of two one-variable probability density functions. Hence,

$$P(\tau, \mu \mid \mathbf{X}, t) = P_1(\tau \mid \mathbf{X}, t) \; . \; P_2(\mu \mid \tau, \mathbf{X}, t) \tag{A.7}$$

The first term on the *r.h.s* is the probability that the next reaction will occur in $[t+\tau, t+\tau, d\tau)$, irrespective of which reaction it might be, and the second term is the probability that the next reaction will be an r_μ reaction, given that the next reaction occurs at time $t+\tau$. We then obtain P_1 by summing $P(\tau, \mu \mid \mathbf{X}, t)d\tau$ over all μ values as,

$$P_1(\tau \mid \mathbf{X}, t) = \sum_{\mu=1}^{M} P(\tau, \mu \mid \mathbf{X}, t) = \sum_{\upsilon=1}^{M} a_\upsilon \; . \; exp\big[- \sum_{\upsilon=1}^{M} a_\upsilon \tau \big] \tag{A.8}$$

Substituting (A.8) into (A.7),

$$P_2(\mu \mid \tau, \mathbf{X}, t) = P(\tau, \mu \mid \mathbf{X}, t) \; / \; \sum_{\upsilon=1}^{M} P(\tau, \upsilon \mid \mathbf{X}, t) = \frac{a_\mu(\mathbf{X})}{\sum_{\upsilon=1}^{M} a_\upsilon} \tag{A.9}$$

The random values τ and μ are then generated according to the density functions in (A.8) and (A.9) respectively. The random τ and μ may be generated according to (A.8) by simply drawing two random numbers γ_1, γ_2 from the uniform distribution in the unit interval and by taking

$$\tau = \frac{1}{\sum_{\upsilon=1}^{M} a_\upsilon} \; \ln\Big(\frac{1}{\gamma_1}\Big) \tag{A.10}$$

and by taking μ to be that integer for which

$$\sum_{\upsilon=1}^{\mu-1} a_\upsilon \; < \; \gamma_2 \sum_{\upsilon=1}^{M} a_\upsilon \; \leq \; \sum_{\upsilon=1}^{\mu} a_\upsilon$$

In the next section we give a simplified version of the code that was written to generate Monte Carlo simulation runs for a reacting system, and to evaluate the dynamic correlations from the simulation results.

NOTE: The τ used in describing the above algorithm is labelled as 'tau_random' in the code given below. However, the term 'tau' appearing in the code, is the time period over which the dynamic correlations are evaluated (Equation (3.14)).

A.1 A simplified code of the Monte-Carlo simulations and the calculation of dynamic correlations

```
# define TIME_SIZE 40

void main()
{
 int N, M       // Number of molecular species, Number of reactions
 t_stop, rp     // Simulation stopping time, Number of result points
 ens,           // Ensemble size
 t_ss,          // Steady state time-point for calculating covariances
 tau,           // Required for calculating dynamic correlations
 mRNA_1, mRNA_2, Prot_1, Prot_2 // M_x, M_y, P_x, P_y
 hist_t;        // Steady-state time at which histogram is collected

 double t = 0.0, norm_cnst = 0.0;
 double ** X_full;      // Matrix where each row is vector X at time-instant 'rp'.
                        //  This is the full output that we desire. Size = (rp * N)

 int * X, * X_initial // # Number of molecules of species at time t, Initial #
                // Size = (N)
 ** R_C_input,  // Input reaction structure, 1 indicates that the species is a reactant
                // Size = (N * M)
 ** R_C_output, // Output reaction structure, 1 indicates that the species is a product
                // Size = (N * M)
 * h_nu;        // vector of molecular reactant combinations of all reactions
                // Size = (M)
 double * c_nu, // vector containing rate constants for all reactions
                // Size = (M)
 * a_nu,        // Vector of propensity functions of each reaction. Size = (M)
 * a_nu_sorted; // Sorting elements of a_nu in ascending order. Size = (M)
 int * index, * index_sorted; // used in the above sorting process

 double * Result_Points, time_stamps;
 double ** processed_data,
                // data related to two molecular species across the
                //  steady state region and across the ensemble
                //   (matrix of size {[rp*ens] * [2]}),
                //     used for calculating the time-covariance plot.
 * cov_avg;

 double r1, r2,         // random numbers drawn from a uniform distribution.
 double tau_random; int nu_random;

 static char time_buffer[TIME_SIZE];
 const struct tm *tm; size_t len; time_t now;

// INPUTS:
//-------------------------------------------------------------------------
 infile >> N >> M >> t_stop >> rp >> ens >> t_ss >> tau >> hist_t;
 for(i=0;i<N;i++)
  for(j=0;j<M;j++)
   { infile >> R_C_input[i][j]; infile >> R_C_output[i][j]; }
 for(i=0;i<M;i++)
  infile >> c_nu[i];
 for(i=0;i<N;i++)
  infile >> X_initial[i];
//-------------------------------------------------------------------------
```

```
for(i=0;i<rp;i++)
{Result_Points[i] = (((static_cast< double >(i))+1.0)*(static_cast< double >(t_stop)))
                          /(static_cast< double >(rp));
 time_stamps[i] = 0.0;
}

time_t seconds;        // Declare variable to hold seconds on clock
time(&seconds);        // Get value from system clock and place in seconds variable
srand((unsigned int) seconds); // Convert seconds to a unsigned integer
now = time (NULL);
tm = localtime (&now);
len = strftime (time_buffer, TIME_SIZE, "%d %B %Y %I:%M:%S %p", tm);
outfile << time_buffer;

for(ens_count=0;ens_count<ens;ens_count++) // ENSEMBLE OF SIMULATIONS
{
 for(i=0;i<N;i++)
        X[i] = X_initial[i];
 t = 0.0;
 spc_count = 0.0;
 counter = 0;
 for(rp_count=0;rp_count<rp;rp_count++) // PERFORMING EACH RUN OF THE SIMULATION
 {if ((t>(t_ss-1+spc_count))&&(spc_count<t_stop-t_ss+1))
        // DATA OF TWO SPECIES BETWEEN WHICH CORRELATIONS ARE EVALUATED
        {processed_data[spc_count][0+(ens_count*2)]=static_cast< double > (X[Prot_1-1]);
         processed_data[spc_count][1+(ens_count*2)]=static_cast< double > (X[Prot_2-1]);
         spc_count = spc_count + 1;
        }

        time_stamps[rp_count] = t;
        if ((t>hist_t)&&(counter==0))
         counter = 1;
        if (counter==1) // DRAWING THE HISTOGRAM OR STEADY-STATE DISTRIBUTION
         {output << X[mRNA_1-1], X[mRNA_2-1], X[Prot_1-1], X[Prot_2-1];
          counter = 2;
         }

        for(i=0;i<N;i++)
         X_full[count][i] = X_full[count][i] + (static_cast< double > (X[i]));

        do
        {for(i=0;i<M;i++)
          h_nu[i] = 1;

         h_nu = Calc_h_nu(h_nu, X, R_C_input, N, M);
         a_nu = Calc_a_nu(h_nu, c_nu, a_nu, M);

         for(i=0;i<M;i++)
         index[i] = i+1;
         index_sorted = Sort_Index(a_nu, a_nu_sorted, index, M);
         a_nu_sorted = Sort(a_nu, a_nu_sorted, M);

         r1 = Random_Uniform();
         if(r1==0.0)
          r1 = 0.5;
         r2 = Random_Uniform();

         tau_random = Get_Random_tau(a_nu, M, r1);
         nu_random = Get_Random_nu(a_nu_sorted, index_sorted, M, r2);
```

```
          X = Calc_X(X, R_C_input, R_C_output, nu_random, N);
          t = t + tau_random;
          }while(t < Result_Points[count]);
   }
  }

 outfile << cpu_time();

 count = 0; tau_random = 0.0;
 for(i=0;i<rp;i++)
 {outfile << Result_Points[i] << "\t";
  outfile << time_stamps[i] << "\t";
  for(j=0;j<N;j++)
   outfile << X_full[i][j]/ens << "\t"; // MEAN NUMBERS OF MOLECULES OF ALL SPECIES
 }

 cov_avg = new double[tau];
 cov_avg = Process_Data_and_Calc_Cov(processed_data, cov_avg, t_ss, rp, tau, ens);
 norm_cnst = Normalize_Cov(processed_data, t_ss, t_stop, ens);

 for(i=0;i<tau;i++)
  outfile << (cov_avg[i]/norm_cnst);  // CORRELATIONS BETWEEN THE TWO PROTEINS

 outfile << cpu_time();
}
```

The following Functions are called by the *main* program:

```
double Random_Uniform(void)
{r = ((double)rand() / ((double)(RAND_MAX)+(double)(1))); return(r);
   // r is a random floating point value in the range [0,1)
}
```

```
int * Calc_h_nu(int * h_nu, int * X, int ** R_C_input, int N, int M)
{temp = 0;
 for(i=0;i<M;i++)
  h_nu[i] = 1;

 for(i=0;i<M;i++)
  for(j=0;j<N;j++)
   if(R_C_input[j][i]!=0)
          h_nu[i] = h_nu[i] * Cmbn(X[j],R_C_input[j][i]);

 for(i=0;i<M;i++)
 {for(j=0;j<N;j++)
   temp = temp + R_C_input[j][i];
                                // This is to take into account those reactions,
                                //   where the product is created spontaneously.
  if(temp==0)
         h_nu[i] = 1;
  temp = 0;
 } return(h_nu);
}
```

```
double * Calc_a_nu(int * h_nu, double * c_nu, double * a_nu, int M)
{for(i=0;i<M;i++)
  a_nu[i] = (static_cast< double > (h_nu[i])) * c_nu[i];
 return(a_nu);
}
```

```
double Get_Random_tau(double * a_nu, int M, double r1)
{tau = 0.0, a = 0.0;
 for(i=0;i<M;i++)
  a = a + a_nu[i];
 if(a==0.0) return(0.0);
 else { tau = (1/a) * log(1/r1); return(tau); }
}
```

```
int Get_Random_nu(double * a_nu_sorted, int * index_sorted, int M, double r2)
{a=0.0, b=0.0;
 for(i=0;i<M;i++)
  a = a + a_nu_sorted[i];
 i = 0;
 while(b < (r2 * a))
 {  b = b + a_nu_sorted[i]; i++; }
 nu = index_sorted[i-1];
 if (i==0) return(0);
 else return(nu);
}
```

```
int * Calc_X(int * X, int ** R_C_input, int ** R_C_output, int nu_random, int N)
{j = nu_random;
 if(nu_random!=0)
  for(i=0;i<N;i++)
   X[i] = X[i] - R_C_input[i][j-1] + R_C_output[i][j-1];
 return(X);
}
```

```
int * Sort_Index(double * a_nu, double * a_nu_sorted, int * ind, int M)
{temp_index = 0;        temp = 0.0;
 for(i=0;i<M;i++)
  a_nu_sorted[i] = a_nu[i];

 for(i=0;i<(M-1);i++)
  for(j=(i+1);j<M;j++)
   if(a_nu_sorted[i]>a_nu_sorted[j])
         {temp = a_nu_sorted[j];
          a_nu_sorted[j] = a_nu_sorted[i];
          a_nu_sorted[i] = temp;

          temp_index = ind[j];
          ind[j] = ind[i];
          ind[i] = temp_index;
         } return(ind);
}
```

```
double * Sort(double * a_nu, double * a_nu_sorted, int M)
{temp = 0.0;
 for(i=0;i<M;i++)
  a_nu_sorted[i] = a_nu[i];
```

```
 for(i=0;i<(M-1);i++)
  for(j=(i+1);j<M;j++)
   if(a_nu_sorted[i]>a_nu_sorted[j])
         {temp = a_nu_sorted[j];
          a_nu_sorted[j] = a_nu_sorted[i];
          a_nu_sorted[i] = temp;
         }      return(a_nu_sorted);
}
```

```
double cpu_time (void) // cpu_time returns the current reading on the CPU clock.
{value = (double) clock() / (double) CLOCKS_PER_SEC; return (value);
}
```

```
double * Process_Data_and_Calc_Cov(double ** processed_data, double * cov_avg,
                                                  int t_ss, int rp, int tau, int ens)
{p = rp-(t_ss-1)-tau;
 count_ens = 0;
 for(j=0;j<p;j++)
 {for(i=0;i<ens;i++)
         new_data[i][0] = processed_data[j][2*i];
  for(k=0;k<tau;k++)
  {for(i=0;i<ens;i++)
          new_data[i][1] = processed_data[j+k][(2*i)+1];
         full_cov[k][j] = Calc_Raw_Covariance(new_data,ens);
  }
 }
 for(i=0;i<tau;i++)
 {cov_avg[i] = 0.0;
  for(j=0;j<p;j++)
   cov_avg[i] = cov_avg[i] + full_cov[i][j];
        cov_avg[i] = cov_avg[i]/p;
 }  return(cov_avg);
}
```

```
double Normalize_Cov(double ** processed_data, int t_ss, int t_stop, int ens)
  // CALCULATING THE NORMALIZATION CONSTANT
{cov1 = 0.0, cov2 = 0.0, norm_cnst = 0.0;
 for(j=0;j<t_stop-t_ss+1;j++)
 {for(i=0;i<ens;i++)
  { new_data[i][0] = processed_data[j][2*i];
    new_data[i][1] = processed_data[j][2*i];
  }
  full_cov[0][j] = Calc_Raw_Covariance(new_data,ens);
  for(i=0;i<ens;i++)
  { new_data[i][0] = processed_data[j][(2*i)+1];
    new_data[i][1] = processed_data[j][(2*i)+1];
  }
  full_cov[1][j] = Calc_Raw_Covariance(new_data,ens);
 }

 for(j=0;j<t_stop-t_ss+1;j++)
 { cov1 = cov1 + full_cov[0][j]; cov2 = cov2 + full_cov[1][j]; }
 cov1 = cov1/(t_stop-t_ss+1); cov2 = cov2/(t_stop-t_ss+1);
  norm_cnst = sqrt(cov1*cov2); return(norm_cnst);
}
```

```
double Calc_Raw_Covariance(double ** data, int s)
{cov = 0.0;
```

```
 for(j=0;j<2;j++)
 {mean[j] = 0.0;
  for(i=0;i<s;i++)
   mean[j] = mean[j] + data[i][j];
  mean[j] = mean[j]/(static_cast< double > (s));
 }

 for(i=0;i<s;i++)
  for(j=0;j<2;j++)
   dataminusmean[i][j] = data[i][j] - mean[j];
 j = 0;
 for(i=0;i<s;i++)
  cov = cov + (dataminusmean[i][j] * dataminusmean[i][j+1]);
 cov = cov/(static_cast< double > (s)); return(cov);
}
```

```
int Cmbn(int n, int r)   // Calculating nCr
{a = 1, b = 1, c = 0, nn = n, rr = r;
 if((n==0)||(n<r)) return(0);
 else if(r==0) return(1);
 else if(r==1) return(n);
 else
 {for(i=0;i<rr;i++) { a = a * nn; nn--; }
  for(i=0;i<rr;i++) { b = b * r; r--; }
  c = a/b;  return(c);
 }
}
```

```
int Factorial(int n)
{m=n, x=1;
 if((n==0)||(n<0)) return(1);
 else { for(i=0;i<m-1;i++) { x = x * n; n--; } return(x); }
}
```

A.2 Sample case:

The objective here is to compare the dynamic correlations obtained through the LNA formulation with those obtained by Monte Carlo simulations. Here, we shall compare the results of the elementary activator system $X \rightarrow Y$, where correlations are drawn between the proteins of the regulator node X and the regulated node Y, which is $\mathrm{Corr}[P_x(t), P_y(t+\tau)]$ evaluated at steady-state conditions using Equation (3.14). Since LNA is applicable only at steady-state, in the simulations, we shall evaluate the correlation function during such a time period (> 't_ss'). The time-evolution of mRNAs and proteins of the GRN are shown in Figures A.1 and A.2 respectively. The time-series corresponding to a single run of the algorithm is the stochastic time-evolution of these variables, whilst the average of 1000 runs of the algorithm is equivalent to deterministic time-evolution, which is described by a set of ODEs given in section 5.1.

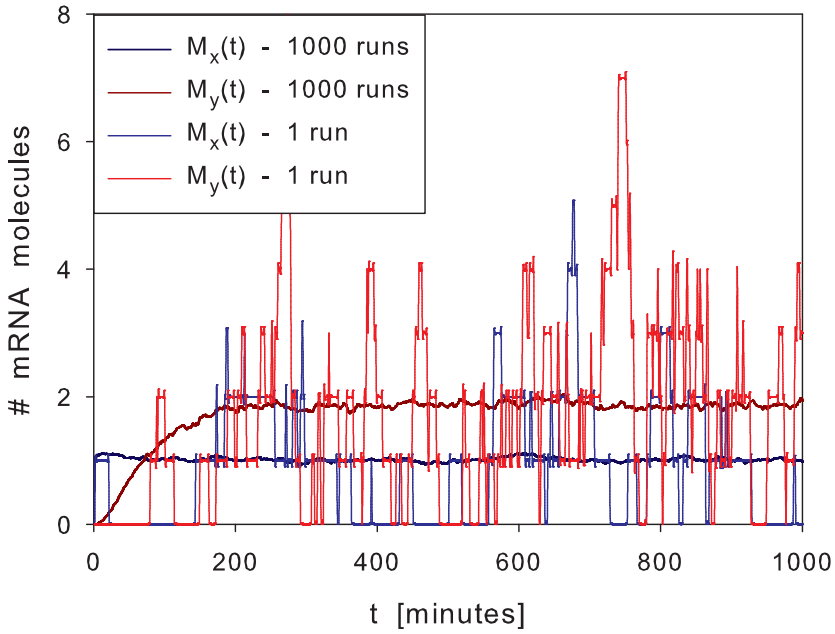

FIGURE A.1: Deterministic and stochastic time evolution of the mRNA species of the $X \rightarrow Y$ regulatory network. Average of 1000 runs of the simulation is considered to be equivalent to the deterministic time evolution obtained through the ODEs.

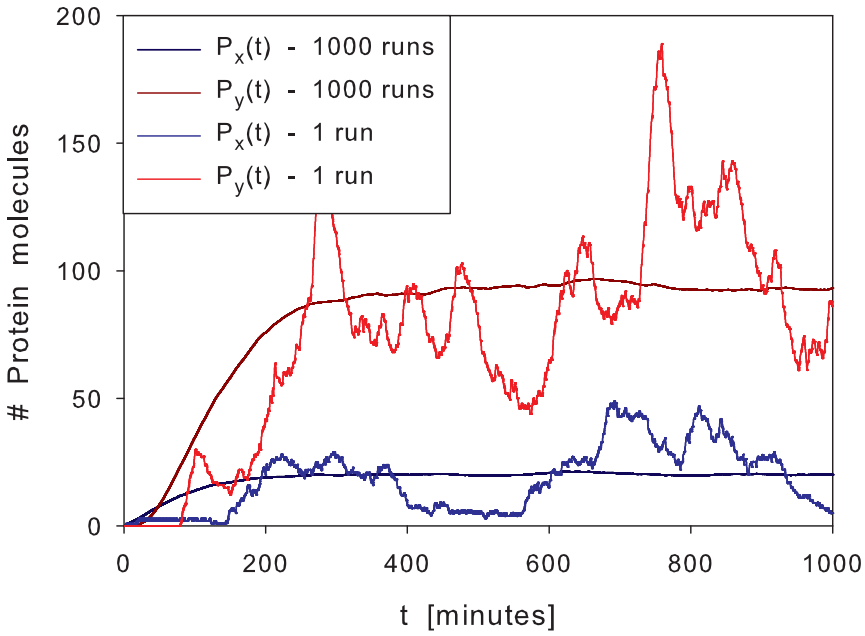

FIGURE A.2: Deterministic and stochastic time evolution of the protein species of the $X \rightarrow Y$ regulatory network.

Since the computational time required to simulate large values of molecular numbers, is very high, due to large values for the *propensity* functions (A.1) and consequently very low values for the time steps (A.10), we assume lower values of molecular numbers. The mean levels of mRNA and protein concentrations are assumed to be: $\langle M_x \rangle = 1.0\ nM$, $\langle M_y \rangle = 2.0\ nM$, $\langle P_x \rangle = 20.0\ nM$ and $\langle P_y \rangle = 100.0\ nM$. The reaction set for this system is given in Table 5.1, where the rate constants, which are the k_μ's in the simulation algorithm, for the above concentration levels are as follows:

$k^+_{M_x} = 0.0407\ nM\ min^{-1},\ k^+_{P_x} = 0.367\ min^{-1}$
$k^+_{M_y} = 0.308\ min^{-1},\ k^+_{P_y} = 0.88\ min^{-1}$

$k_{on} = 1.0\ nM^{-1} min^{-1},\ k_{off} = 20.0\ min^{-1}$

$k^-_{M_x} = 0.0407\ min^{-1},\ k^-_{M_y} = 0.0770\ min^{-1}$
$k^-_{P_x} = 0.0184\ min^{-1},\ k^-_{P_y} = 0.0176\ min^{-1}$

For the above set of values, the computational time for simulating 1000 runs of the regulatory system on a single PC is about 4 minutes. Also, the terms *concentration* and *number of molecules* are used interchangeably since the volume is assumed to be of 1 unit. From the reaction set of this GRN, it is clear that the production of M_x is spontaneous or in other words is transcribed from a gene that is constantly in the active/ON state. As discussed in 2.3, the steady-state distribution of such a species is Poissonian due to its spontaneous births and deaths. The distribution around the mean value of M_x is shown in Figure A.3. On the other hand, the mRNA of the downtream node M_y is modelled as being transcribed by a switching gene G_y which transits between the active and inactive states due to the binding and unbinding of the transcription factor P_x on to its regulatory region. Therefore, its distribution is a heavy-talied gamma as shown in Figure A.4. The issues on mRNA distribution were dealt with in section 2.2 with the aid of the model described in Figure 2.2.

Further in section 2.4, we discussed in brief on the shapes of the probability distributions of the proteins at steady-state conditions, for such models. Our simulation of the elementary activator system produces steady-state protein distributions, that are in line with these discussions. The distributions around the mean values of P_x and P_y are shown in Figures A.5 and A.6 respectively. While both are gamma distributed, the mean level of P_x being lower, has a long-tailed distribution as compared to that of P_y.

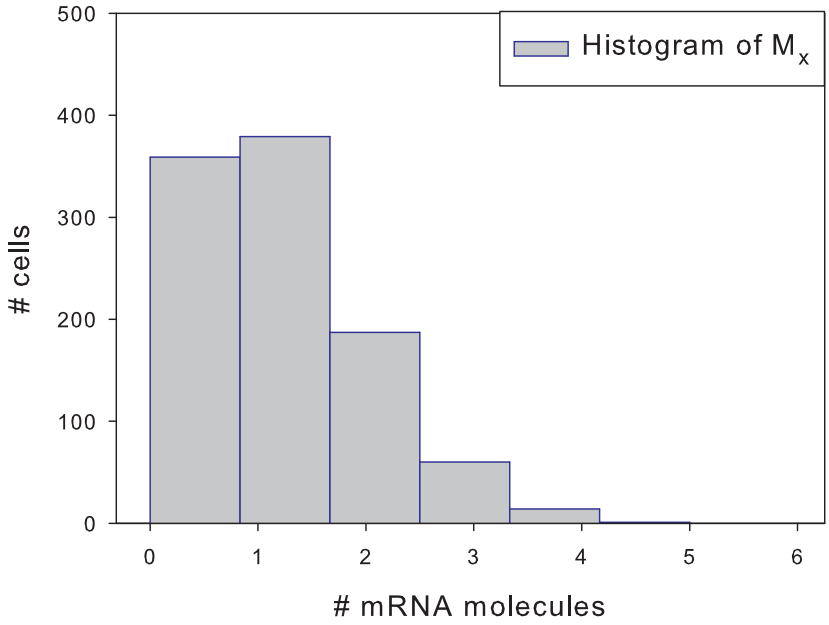

FIGURE A.3: Probability distribution or histogram of M_x at steady-state condition, for 1000 runs of the $X \rightarrow Y$ system, or equivalent number of cells incorporating the $X \rightarrow Y$ system.

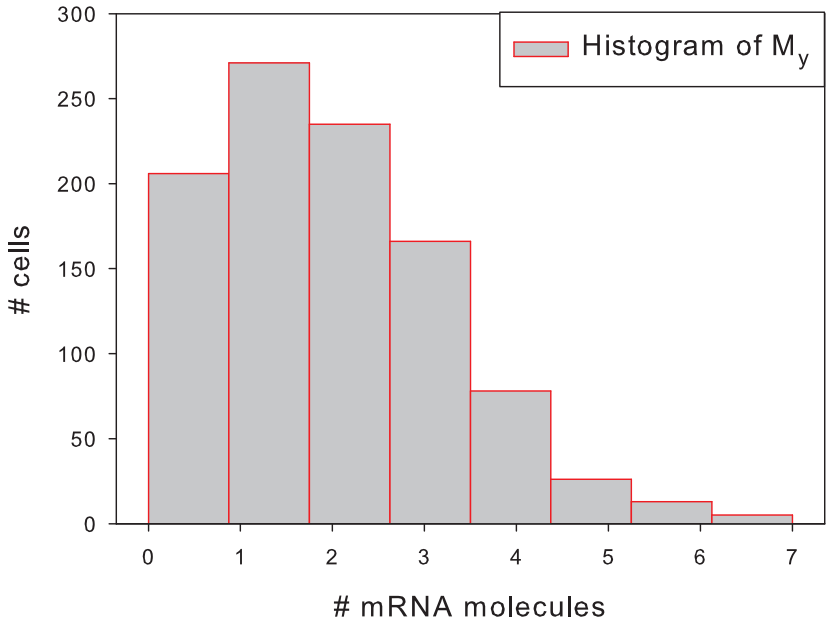

FIGURE A.4: Probability distribution or histogram of M_y at steady-state condition.

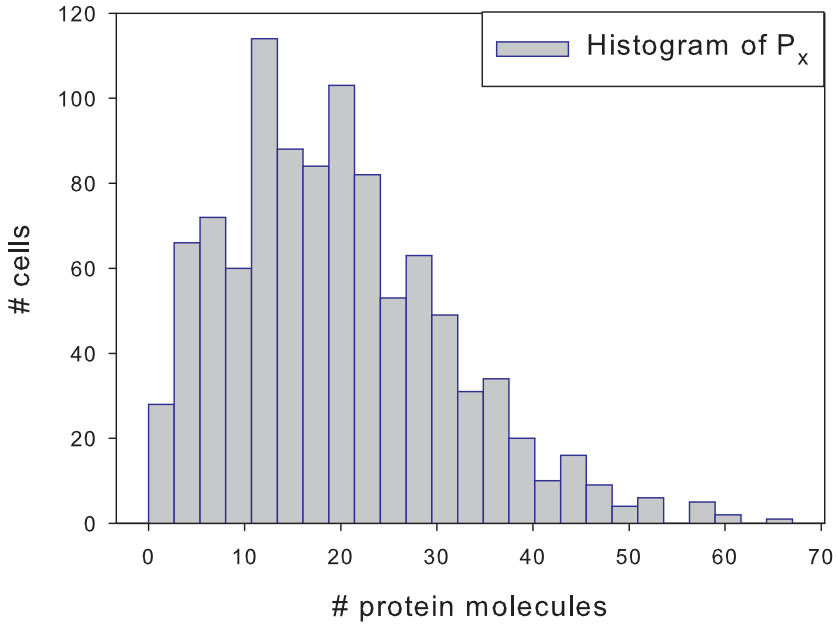

FIGURE A.5: Probability distribution or histogram of P_x at steady-state condition, for 1000 runs of the $X \rightarrow Y$ system, or equivalent number of cells incorporating the $X \rightarrow Y$ system.

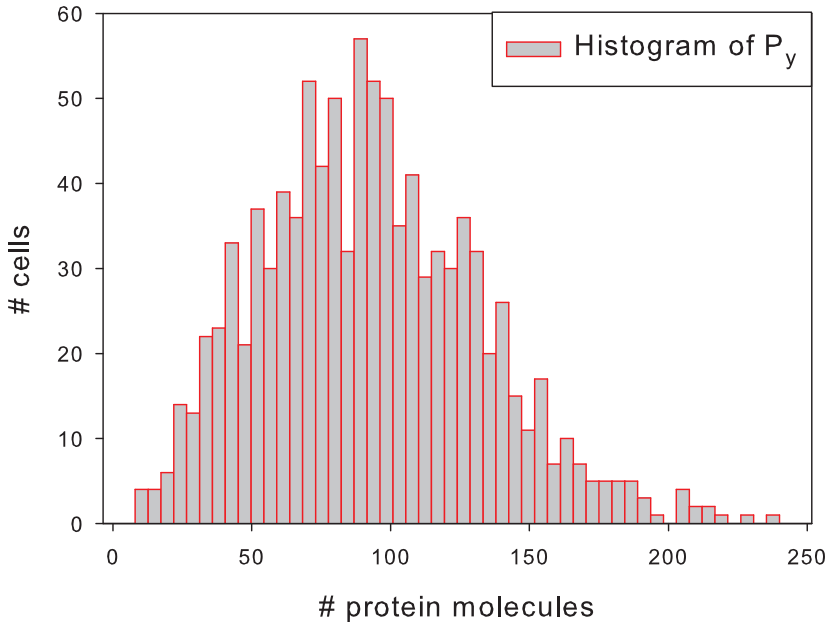

FIGURE A.6: Probability distribution or histogram of P_y at steady-state condition.

For such a regulatory system, we shall compare the dynamic correlations evaluated from the analytical formulation of chapter 3 with those obtained by simulations. Figure A.7 shows that the correlations obtained from simulations match closely with that of the analytical one, as the ensemble size increases. The evaluation of correlations as shown in the programming code, rests on the equivalence between the ensemble-averaging and time-averaging of such stochastic variables.

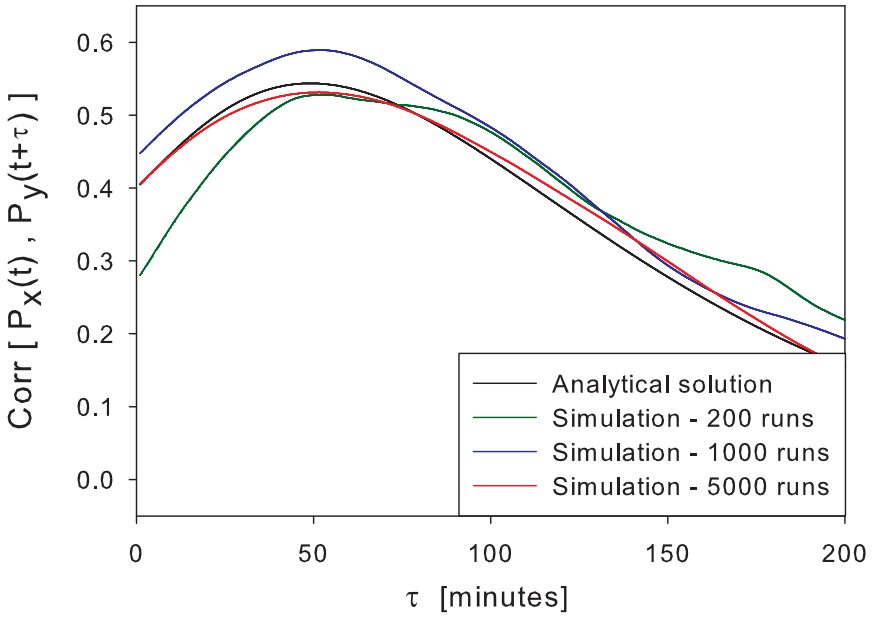

FIGURE A.7: Dynamic correlations between the proteins of the $X \rightarrow Y$ regulatory network, obtained through analytics and simulations.

Appendix B

Expression for τ^* and t_{resp} in the case of an Elementary Activator

Here we derive expressions for the times τ^* and t_{resp} in the case of an activator link where the connectivity between the molecular species can be represented as $M_x \rightarrow P_x \rightarrow M_y \rightarrow P_y$. The expressions that we derive here, correspond to those shown in Figure 5.4 of section 5.1.3. In section 3.3, on the single-gene, we derive the expressions for τ^* and t_{resp}. The derivations shown here follow exactly the same procedure adopted there. In the time-covariance functions , which are sum of exponentials raised to the power of the eigenevalues of the deterministic system, the exponential corresponding to the TF-DNA binding/unbinding rates $e^{-(k_{on}+k_{off})\tau}$ is neglected since $(k_{on} + k_{off})$ is usually much larger than the decay rates. Further, in the reduced expression for the covariances, we do the approximations $(k_{on} + k_{off} \gg k^-_{(M,P)}, k_{off} > k_{on})$ and arrive at the following expressions for the covariance between M_x and its adjacent species in the connectivity network, P_x.

$$\text{Cov } [M_x(t), P_x(t+\tau)] = \frac{k^+_{P_x} k^+_{M_x}}{k^-_{M_x}} \Bigg[- \frac{e^{-k^-_{M_x}\tau}}{(k^-_{M_x} - k^-_{P_x})} + \frac{2k^-_{M_x} e^{-k^-_{P_x}\tau}}{(k^{-2}_{M_x} - k^{-2}_{P_x})} \Bigg]$$

The expression for τ^* is obtained by partially differentiating the above expression *w.r.t* τ and equating it to zero.

$$\begin{aligned}
\frac{\partial}{\partial \tau} \text{Cov } [M_x(t), P_x(t+\tau)] &= 0 \\
\Rightarrow \; -2k^-_{P_x} k^-_{M_x} e^{-k^-_{P_x}\tau^*} + k^-_{M_x}(k^-_{M_x} + k^-_{P_x}) e^{-k^-_{P_x}\tau^*} &= 0 \\
e^{(k^-_{M_x} - k^-_{P_x})\tau^*} &= \frac{k^-_{M_x} + k^-_{P_x}}{2k^-_{P_x}} \\
\tau^* &= \frac{\ln\left[\frac{k^-_{M_x} + k^-_{P_x}}{2k^-_{P_x}}\right]}{(k^-_{M_x} - k^-_{P_x})} \qquad \text{(B.1)}
\end{aligned}$$

Denoting the mean steady-state level of M_x as $M_x(t_{ss})$, where t_{ss} is the steady state time-period, an instantaneous perturbation at time $t = 0$ in this mean value by an amount ΔM_x induces a deterministic response in $P_x(t)$. This response follows the deterministic rate equation but now, with the initial concentration of M_x being equal to $M_x(t_{ss}) + \Delta M_x$. On solving for the rate equation of the protein P_x, we get,

$$\begin{aligned} P_x(t) &= P_x(t_{ss}) + \frac{k^+_{P_x}\big[(\Delta M_x + M_x(t_{ss}))k^-_{M_x} - k^+_{M_x}\big]}{k^-_{M_x}(k^-_{M_x} - k^-_{P_x})}\Big[- e^{-k^-_{M_x}t} + e^{-k^-_{P_x}t}\Big] \\ P_x(t) &= P_x(t_{ss}) + \frac{k^+_{P_x}\big[\Delta M_x\big]}{(k^-_{M_x} - k^-_{P_x})}\Big[- e^{-k^-_{M_x}t} + e^{-k^-_{P_x}t}\Big] \end{aligned}$$

because $M_x(t_{ss}) = \langle M_x \rangle = k^+_{M_x}/k^-_{M_x}$, differentiating the above expression *w.r.t* the time variable t,

$$\frac{\partial P_x(t)}{\partial t} = \frac{k^+_{P_x}\big[\Delta M_x\big]}{(k^-_{M_x} - k^-_{P_x})}\Big[k^-_{M_x}e^{-k^-_{M_x}t} - k^-_{P_x}e^{-k^-_{P_x}t}\Big]$$

On equating this derivative to zero,

$$\begin{aligned} k^-_{M_x}e^{-k^-_{M_x}t_{resp}} &= k^-_{P_x}e^{-k^-_{P_x}t_{resp}} \\ t_{resp} &= \frac{\ln\left[\frac{k^-_{M_x}}{k^-_{P_x}}\right]}{(k^-_{M_x} - k^-_{P_x})} \end{aligned} \tag{B.2}$$

The difference between the two times is the same as in the case of the single gene system, and is:

$$t_{resp} - \tau^* = \frac{\ln\left[\frac{2k^-_{M_x}}{k^-_{M_x}+k^-_{P_x}}\right]}{(k^-_{M_x} - k^-_{P_x})} \tag{B.3}$$

Continuing on similar lines, the expressions for the two *times*, are derived in the case where the *output* element is now M_y in place of P_x. Note that the *input* element is still M_x. The covariance between these two molecular species is:

$$\begin{aligned} \text{Cov}\ [M_x(t), M_y(t+\tau)] = \frac{k^+_{M_y}k^+_{P_x}k^+_{M_x}}{k^-_{M_x}}\Bigg[&\frac{e^{-k^-_{M_x}\tau}}{(k^-_{M_x} - k^-_{M_y})(k^-_{M_x} - k^-_{P_x})} + \frac{2k^-_{M_x}e^{-k^-_{P_x}\tau}}{(k^-_{M_y} - k^-_{P_x})(k^{-2}_{M_x} - k^{-2}_{P_x})} \\ &- \frac{2k^-_{M_x}e^{-k^-_{M_y}\tau}}{(k^{-2}_{M_x} - k^{-2}_{M_y})(k^-_{M_y} - k^-_{P_x})}\Bigg] \end{aligned}$$

If $k^-_{M_y}$ is largest of the decay rates, its exponential and the related terms could be neglected.

$$\text{Cov } [M_x(t), M_y(t+\tau)] \approx \frac{k^+_{M_y}k^+_{P_x}k^+_{M_x}}{k^-_{M_x}}\left[\frac{e^{-k^-_{M_x}\tau}}{(k^-_{M_x}-k^-_{M_y})(k^-_{M_x}-k^-_{P_x})}+\frac{2k^-_{M_x}e^{-k^-_{P_x}\tau}}{(k^-_{M_y}-k^-_{P_x})(k^{-2}_{M_x}-k^{-2}_{P_x})}\right]$$

Differentiating *w.r.t* τ and equating to zero,

$$\begin{aligned}
\frac{\partial}{\partial\tau}\text{Cov } [M_x(t), M_y(t+\tau)] &= 0 \\
\Rightarrow \frac{k^-_{M_x}e^{-k^-_{M_x}\tau^*}}{(k^-_{M_x}-k^-_{M_y})(k^-_{M_x}-k^-_{P_x})} &\approx \frac{-2k^-_{M_x}k^-_{P_x}e^{-k^-_{P_x}\tau^*}}{(k^-_{M_y}-k^-_{P_x})(k^{-2}_{M_x}-k^{-2}_{P_x})} \\
e^{-k^-_{M_x}\tau^*}(k^-_{M_y}-k^-_{P_x})(k^-_{M_x}+k^-_{P_x}) &\approx 2k^-_{P_x}e^{-k^-_{P_x}\tau^*}(k^-_{M_y}-k^-_{M_x}) \\
e^{(k^-_{M_x}-k^-_{P_x})\tau^*} &\approx \frac{(k^-_{M_y}-k^-_{P_x})(k^-_{M_x}+k^-_{P_x})}{2k^-_{P_x}(k^-_{M_y}-k^-_{M_x})} \\
\tau^* &\approx \frac{\ln\left[\frac{(k^-_{M_y}-k^-_{P_x})(k^-_{M_x}+k^-_{P_x})}{2k^-_{P_x}(k^-_{M_y}-k^-_{M_x})}\right]}{(k^-_{M_x}-k^-_{P_x})}
\end{aligned} \tag{B.4}$$

Now, for a perturbation of ΔM_x in the mean steady-state value of M_x, the response in M_y is:

$$\begin{aligned}
M_y(t) &= M_y(t_{ss}) + \frac{k^+_{M_y}k^+_{P_x}\left[(\Delta M_x + M_x(t_{ss}))k^-_{M_x}-k^+_{M_x}\right]}{k^-_{M_x}} \\
&\quad\times\left[\frac{e^{-k^-_{M_x}t}}{(k^-_{M_x}-k^-_{M_y})(k^-_{M_x}-k^-_{P_x})}+\frac{e^{-k^-_{P_x}t}}{(k^-_{M_x}-k^-_{P_x})(k^-_{M_y}-k^-_{P_x})}\right. \\
&\quad\left.-\frac{e^{-k^-_{M_y}t}}{(k^-_{M_x}-k^-_{M_y})(k^-_{M_y}-k^-_{P_x})}\right] \\
&\approx M_y(t_{ss}) + \frac{k^+_{M_y}k^+_{P_x}\left[(\Delta M_x + M_x(t_{ss}))k^-_{M_x}-k^+_{M_x}\right]}{k^-_{M_x}} \\
&\quad\times\left[\frac{e^{-k^-_{M_x}t}}{(k^-_{M_x}-k^-_{M_y})(k^-_{M_x}-k^-_{P_x})}+\frac{e^{-k^-_{P_x}t}}{(k^-_{M_x}-k^-_{P_x})(k^-_{M_y}-k^-_{P_x})}\right]
\end{aligned}$$

Taking the derivative $\frac{\partial M_y(t)}{\partial t}$ and equating it to zero, we get the following equation for t_{resp}:

$$\begin{aligned}
\frac{-k^-_{M_x}e^{-k^-_{M_x}t_{resp}}}{(k^-_{M_x}-k^-_{M_y})(k^-_{M_x}-k^-_{P_x})} &= \frac{k^-_{M_x}e^{-k^-_{P_x}t_{resp}}}{(k^-_{M_x}-k^-_{P_x})(k^-_{M_y}-k^-_{P_x})} \\
t_{resp} &= \frac{\ln\left[\frac{k^-_{M_x}(k^-_{M_y}-k^-_{P_x})}{k^-_{P_x}(k^-_{M_y}-k^-_{M_x})}\right]}{(k^-_{M_x}-k^-_{P_x})}
\end{aligned} \tag{B.5}$$

$$t_{resp} - \tau^* = \frac{\ln\left[\frac{2k^-_{M_x}}{(k^-_{M_x}+k^-_{P_x})}\right]}{(k^-_{M_x} - k^-_{P_x})} \tag{B.6}$$

On the other hand, if the half-life of mRNA M_x is the smallest of all the molecular species, $k^-_{M_x}$ would consequently be the largest of the decay rates and therefore its exponential and the related terms could be neglected. The covariance term is now:

$$\text{Cov } [M_x(t), M_y(t+\tau)] = \frac{k^+_{M_y}k^+_{P_x}k^+_{M_x}}{k^-_{M_x}}\left[\frac{2k^-_{M_x}e^{-k^-_{P_x}\tau}}{(k^-_{M_y}-k^-_{P_x})(k^{-2}_{M_x}-k^{-2}_{P_x})} - \frac{2k^-_{M_x}e^{-k^-_{M_y}\tau}}{(k^{-2}_{M_x}-k^{-2}_{M_y})(k^-_{M_y}-k^-_{P_x})}\right]$$

Differentiating this expression for the covariance *w.r.t* τ and equating the resulting partial derivative to zero,

$$\begin{aligned}
\frac{\partial}{\partial\tau}\text{Cov } [M_x(t), M_y(t+\tau)] &= 0 \\
\Rightarrow \quad -2k^-_{M_x}k^-_{P_x}e^{-k^-_{P_x}\tau^*}(k^{-2}_{M_x}-k^{-2}_{M_y}) &\approx -2k^-_{M_x}k^-_{M_y}e^{-k^-_{M_y}\tau^*}(k^{-2}_{M_x}-k^{-2}_{P_x}) \\
e^{(k^-_{M_y}-k^-_{P_x})\tau^*} &\approx \frac{k^-_{M_y}(k^{-2}_{M_x}-k^{-2}_{P_x})}{k^-_{P_x}(k^{-2}_{M_x}-k^{-2}_{M_y})} \\
\tau^* &\approx \frac{\ln\left[\frac{k^-_{M_y}(k^{-2}_{M_x}-k^{-2}_{P_x})}{k^-_{P_x}(k^{-2}_{M_x}-k^{-2}_{M_y})}\right]}{(k^-_{M_y}-k^-_{P_x})}
\end{aligned} \tag{B.7}$$

Similarly, the expression for t_{resp} in the case, where $k^-_{M_x}$ is larger than the other decay rates, is obtained as follows:

$$\begin{aligned}
M_y(t) &\approx M_y(t_{ss}) + \frac{k^+_{M_y}k^+_{P_x}\left[(\Delta M_x + M_x(t_{ss}))k^-_{M_x} - k^+_{M_x}\right]}{k^-_{M_x}} \\
&\times \left[\frac{e^{-k^-_{P_x}t}}{(k^-_{M_x}-k^-_{P_x})(k^-_{M_y}-k^-_{P_x})} - \frac{e^{-k^-_{M_y}t}}{(k^-_{M_x}-k^-_{M_y})(k^-_{M_y}-k^-_{P_x})}\right]
\end{aligned}$$

Once again taking the derivative of this expression *w.r.t* t and equating it to 0, we get $\frac{\partial M_y(t)}{\partial t} = 0$ which gives:

$$\begin{aligned}
\frac{k^-_{P_x}e^{-k^-_{P_x}t}}{(k^-_{M_x}-k^-_{P_x})(k^-_{M_y}-k^-_{P_x})} &= -\frac{k^-_{M_y}e^{-k^-_{M_y}t}}{(k^-_{M_x}-k^-_{M_y})(k^-_{M_y}-k^-_{P_x})} \\
t_{resp} &= \frac{\ln\left[\frac{k^-_{M_y}(k^-_{M_x}-k^-_{P_x})}{k^-_{P_x}(k^-_{M_x}-k^-_{M_y})}\right]}{(k^-_{M_y}-k^-_{P_x})}
\end{aligned} \tag{B.8}$$

$$t_{resp} - \tau^* = \frac{\ln\left[\frac{(k^-_{M_x}+k^-_{M_y})}{(k^-_{M_x}+k^-_{P_x})}\right]}{(k^-_{M_y}-k^-_{P_x})} \tag{B.9}$$

We now look at the case where the *output* element is P_y, whilst M_x is still the *input* element. The dynamic covariance between them are:

$$\begin{aligned}\text{Cov } [M_x(t), P_y(t+\tau)] = \frac{k^+_{P_y}k^+_{M_y}k^+_{P_x}k^+_{M_x}}{k^-_{M_x}}\Bigg[&\frac{-e^{-k^-_{M_x}\tau}}{(k^-_{M_x}-k^-_{M_y})(k^-_{M_x}-k^-_{P_x})(k^-_{M_x}-k^-_{P_y})}\\ &-\frac{2k^-_{M_x}e^{-k^-_{P_x}\tau}}{(k^-_{M_y}-k^-_{P_x})(k^{-2}_{M_x}-k^{-2}_{P_x})(k^-_{P_x}-k^-_{P_y})}\\ &-\frac{2k^-_{M_x}e^{-k^-_{M_y}\tau}}{(k^{-2}_{M_y}-k^{-2}_{M_x})(k^-_{M_y}-k^-_{P_x})(k^-_{M_y}-k^-_{P_y})}\\ &+\frac{2k^-_{M_x}e^{-k^-_{P_y}\tau}}{(k^-_{M_y}-k^-_{P_y})(k^-_{P_x}-k^-_{P_y})(k^{-2}_{M_x}-k^{-2}_{P_y})}\Bigg]\end{aligned}$$

Neglecting terms related to $e^{-k^-_{M_y}}$ and $e^{-k^-_{M_x}}$ since the decay rates $k^-_{M_y}$ and $k^-_{M_x}$ are large compared to the decay rates of the proteins. The covariance in now approximated as:

$$\begin{aligned}\text{Cov } [M_x(t), P_y(t+\tau)] \approx k^+_{M_y}k^+_{P_x}k^+_{P_y}\langle M_x\rangle\Bigg[&\frac{-2k^-_{M_x}e^{-k^-_{P_x}\tau}}{(k^-_{M_y}-k^-_{P_x})(k^{-2}_{M_x}-k^{-2}_{P_x})(k^-_{P_x}-k^-_{P_y})}\\ &+\frac{2k^-_{M_x}e^{-k^-_{P_y}\tau}}{(k^-_{M_y}-k^-_{P_y})(k^-_{P_x}-k^-_{P_y})(k^{-2}_{M_x}-k^{-2}_{P_y})}\Bigg]\end{aligned}$$

Differentiating this expression for the covariance *w.r.t* τ and equating the partial derivative to zero,

$$\begin{aligned}\frac{\partial}{\partial\tau}\text{Cov } [M_x(t), P_y(t+\tau)] &= 0\\ \Rightarrow \frac{2k^-_{M_x}k^-_{P_x}e^{-k^-_{P_x}\tau^*}}{(k^-_{M_y}-k^-_{P_x})(k^{-2}_{M_x}-k^{-2}_{P_x})(k^-_{P_x}-k^-_{P_y})} &\approx \frac{2k^-_{M_x}k^-_{P_y}e^{-k^-_{P_y}\tau^*}}{(k^-_{M_y}-k^-_{P_y})(k^-_{P_x}-k^-_{P_y})(k^{-2}_{M_x}-k^{-2}_{P_y})}\\ e^{(k^-_{P_y}-k^-_{P_x})\tau^*} &\approx \frac{k^-_{P_y}(k^-_{M_y}-k^-_{P_x})(k^{-2}_{M_x}-k^{-2}_{P_x})}{k^-_{P_x}(k^-_{M_y}-k^-_{P_y})(k^{-2}_{M_x}-k^{-2}_{P_y})}\\ \tau^* &\approx \frac{\ln\left[\frac{k^-_{P_y}(k^-_{M_y}-k^-_{P_x})(k^{-2}_{M_x}-k^{-2}_{P_x})}{k^-_{P_x}(k^-_{M_y}-k^-_{P_y})(k^{-2}_{M_x}-k^{-2}_{P_y})}\right]}{(k^-_{P_y}-k^-_{P_x})}\end{aligned} \tag{B.10}$$

The expression for the response in the mean value of P_y is,

$$\begin{aligned}
P_y(t) &= P_y(t_{ss}) + \frac{k^+_{P_y} k^+_{M_y} k^+_{P_x} \left[(\Delta M_x + M_x(t_{ss}))k^-_{M_x} - k^+_{M_x}\right]}{k^-_{M_x}} \\
&\quad \times \left[- \frac{e^{-k^-_{M_x} t}}{(k^-_{M_x} - k^-_{M_y})(k^-_{M_x} - k^-_{P_x})(k^-_{M_x} - k^-_{P_y})} \right. \\
&\quad + \frac{e^{-k^-_{P_x} t}}{(k^-_{M_x} - k^-_{P_x})(k^-_{M_y} - k^-_{P_x})(k^-_{P_y} - k^-_{P_x})} \\
&\quad + \frac{e^{-k^-_{M_y} t}}{(k^-_{M_x} - k^-_{M_y})(k^-_{M_y} - k^-_{P_x})(k^-_{M_y} - k^-_{P_y})} \\
&\quad \left. - \frac{e^{-k^-_{P_y} t}}{(k^-_{M_x} - k^-_{P_y})(k^-_{M_y} - k^-_{M_y})(k^-_{P_y} - k^-_{P_x})} \right] \\
&\approx P_y(t_{ss}) + \frac{k^+_{P_y} k^+_{M_y} k^+_{P_x} \left[(\Delta M_x + M_x(t_{ss}))k^-_{M_x} - k^+_{M_x}\right]}{k^-_{M_x}} \\
&\quad \times \left[\frac{e^{-k^-_{P_x} t}}{(k^-_{M_x} - k^-_{P_x})(k^-_{M_y} - k^-_{P_x})(k^-_{P_y} - k^-_{P_x})} \right. \\
&\quad \left. - \frac{e^{-k^-_{P_y} t}}{(k^-_{M_x} - k^-_{P_y})(k^-_{M_y} - k^-_{M_y})(k^-_{P_y} - k^-_{P_x})} \right]
\end{aligned}$$

Taking the derivative $\frac{\partial P_y(t)}{\partial t}$ and equating it to zero, we get the following equation for t_{resp}:

$$\begin{aligned}
\frac{-k^-_{P_x} e^{-k^-_{P_x} t_{resp}}}{(k^-_{M_x} - k^-_{P_x})(k^-_{M_y} - k^-_{P_x})(k^-_{P_y} - k^-_{P_x})} &= \frac{-k^-_{P_y} e^{-k^-_{P_y} t_{resp}}}{(k^-_{M_x} - k^-_{P_y})(k^-_{M_y} - k^-_{M_y})(k^-_{P_y} - k^-_{P_x})} \\
t_{resp} &= \frac{\ln\left[\frac{k^-_{P_y}(k^-_{M_x} - k^-_{P_x})(k^-_{M_y} - k^-_{P_x})}{k^-_{P_x}(k^-_{M_x} - k^-_{P_y})(k^-_{M_y} - k^-_{P_y})}\right]}{(k^-_{P_y} - k^-_{P_x})} \qquad \text{(B.11)}
\end{aligned}$$

$$t_{resp} - \tau^* \approx \frac{\ln\left[\frac{(k^-_{M_x} + k^-_{P_y})}{(k^-_{M_x} + k^-_{P_x})}\right]}{(k^-_{P_y} - k^-_{P_x})} \qquad \text{(B.12)}$$

Bibliography

Uri Alon. Network motifs: theory and experimental approaches. *Nature Reviews Genetics*, 8(6):450–461, Jun 2007.

Yoav Arava, Yulei Wang, John D Storey, Chih Long Liu, Patrick O Brown, and Daniel Herschlag. Genome-wide analysis of mrna translation profiles in *Saccharomyces cerevisiae*. *PNAS*, 100(7):3889–3894, 2003.

Adam Arkin, John Ross, and Harley H. McAdams. Stochastic kinetic analysis of developmental pathway bifurcation in phage λ-infected *Escherichia coli* cells. *Genetics*, 149:1633–1648, Aug 1998.

D. W. Austin, M. S. Allen, J. M. McCollum, R. D. Dar, J. R. Wilgus, G. S. Sayler, N. F. Samatova, C. D. Cox, and M. L. Simpson. Gene network shaping of inherent noise spectra. *Nature Letters*, 439:608–611, Feb 2006.

A Bar-Even, J Paulsson, N Maheshri, M Carmi, E O'Shea, Y Pilpel, and N Barkai. Noise in protein expression scales with natural protein abundance. *Nature Genetics*, 38:636–643, 2006.

Ziv Bar-Joseph, Georg K Gerber, Tong Ihn Lee, Nicola J Rinaldi, Jane Y Yoo, Franois Robert, D Benjamin Gordon, Ernest Fraenkel, Tommi S Jaakkola, Richard A Young, and David K Gifford. Computational discovery of gene modules and regulatory networks. *Nature Biotechnology*, 21:1337–1342, October 2003.

N. Barkai and S. Leibler. Biological rhythms: circadian clocks limited by noise. *Nature*, 403:267–268, 2000.

Attila Becskei, Benjamin B Kaufmann, and Alexander van Oudenaarden. Contributions of low molecule number and chromosomal positioning to stochastic gene expression. *Nature Genetics*, 37:937 – 944, August 2005.

Attila Becskei and Luis Serrano. Engineering stability in gene networks by autoregulation. *Nature*, 405:590–593, June 2000.

Archana Belle, Amos Tanay, Ledion Bitincka, Ron Shamir, and Erin K. O'Shea. Quantification of protein half-lives in the budding yeast proteome. *PNAS*, 103(35): 13004–13009, 2006.

C. Bornaes and M. W. Ignjatovic et al. A regulatory element in the cha1 promoter which confers inducibility by serine and threonine on *Saccharomyces cerevisiae* genes. *Molecular and Cellular Biology*, 13(12):7604–7611, 1993.

Nicolas E. Buchler, Ulrich Gerland, and Terence Hwa. On schemes of combinatorial transcription logic. *PNAS*, 100(9):5136–5141, 2003.

R. Bundschuh, F. Hayot, and C. Jayaprakash. The role of dimerization in noise reduction of simple genetic networks. *Journal of Theoretical Biology*, 220(2):261 – 269, 2003. ISSN 0022-5193.

Long Cai, Nir Friedman, and X. Sunney Xie. Stochastic protein expression in individual cells at the single molecule level. *Nature*, 440:358–362, March 2006.

T. Chen, H.L. He, and G.M. Church. Modeling gene expression with differential equations. *Proceedings of the Pacific Symposium on Biocomputing*, 21:29–40, 1999.

Tianjiao Chu, Clark Glymour, Richard Scheines, and Peter Spirtes. A statistical problem for inference to regulatory structure from associations of gene expression measurements with microarrays. *Bioinformatics*, 19(9):1147–1152, 2003.

Stephen Cooper. Is whole-culture synchronization biology's 'perpetual-motion machine'? *Trends in Biotechnology*, 22(6):266 – 269, 2004a. ISSN 0167-7799.

Stephen Cooper. Rejoinder: whole-culture synchronization cannot, and does not, synchronize cells. *Trends in Biotechnology*, 22(6):274 – 276, 2004b. ISSN 0167-7799.

Stephen Cooper and Kerby Shedden. Microarray analysis of gene expression during the cell cycle. *Cell & Chromosome*, 2(1):1, 2003. ISSN 1475-9268.

Stephen Cooper and Kerby Shedden. Microarrays and the relationship of mrna variation to protein variation during the cell cycle. *Journal of Theoretical Biology*, 249(3):574 – 581, 2007. ISSN 0022-5193.

Chris D. Cox, James M. McCollum, Michael S. Allen, Roy D. Dar, and Michael L. Simpson. Using noise to probe and characterize gene circuits. *PNAS*, 105(31):10809–10814, Aug 2008.

David Danks and Clark Glymour. Linearity properties of bayes nets with binary variables. In *Proceedings of the 17th Annual Conference on Uncertainty in Artificial Intelligence (UAI-01)*, pages 98–104, San Francisco, CA, 2001. Morgan Kaufmann.

Eric H. Davidson, Jonathan P. Rast, Paola Oliveri, Andrew Ransick, Cristina Calestani, Chiou-Hwa Yuh, Takuya Minokawa, Gabriele Amore, Veronica Hinman, Csar Arenas-Mena, Ochan Otim, C. Titus Brown, Carolina B. Livi, Pei Yun Lee, Roger Revilla, Alistair G. Rust, Zheng jun Pan, Maria J. Schilstra, Peter J. C. Clarke, Maria I. Arnone, Lee Rowen, R. Andrew Cameron, David R. McClay, Leroy Hood, and Hamid

Bolouri. A genomic regulatory network for development. *Science*, 295(5560):1669–1678, 2002.

C. R. Vazquez de Aldana, M. J. Marton, and A. G. Hinnebusch. Gcn20, a novel atp binding cassette protein, and gcn1 reside in a complex that mediates activation of the eif-2α kinase gcn2 in amino acid-starved cells. *EMBO Journal*, 14(13):31843199, July 1995.

Michiel J. L. de Hoon, Seiya Imoto, Kazuo Kobayashi, Naotake Ogasawara, and Satoru Miyano. Inferring gene regulatory networks from time-ordered gene expression data of bacillus subtilis using differential equations. In *Pacific Symposium on Biocomputing*, pages 17–28, 2003.

A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum likelihood from incomplete data via the em algorithm. *Journal of the Royal Statistical Society: Series B (Methodological)*, 39(1):1–38, 1977.

Yann Dublanche, Konstantinos Michalodimitrakis, Nico Kmmerer, Mathilde Foglierini, and Luis Serrano. Noise in transcription negative feedback loops: simulation and experimental analysis. *Molecular Systems Biology*, 2:41, 2006.

M. J. Dunlop, R. S. Cox III, J. H. Levine, R. M. Murray, and M. B. Elowitz. Regulatory activity revealed by dynamic correlations in gene expression noise. *Nature Genetics*, 40:1493–1498, December 2008.

Michael B. Eisen, Paul T. Spellman, Patrick O. Brown, and David Botstein. Cluster analysis and display of genome-wide expression patterns. *Proceedings of the National Academy of Sciences of the United States of America*, 95(25):14863–14868, 1998.

J. Elf and M. Ehrenberg. Fast evaluation of fluctuations in biochemical networks with the linear noise approximation. *Genome Research*, 13:2475–2484, 2003.

Michael B. Elowitz and Stanislas Leibler. A synthetic oscillatory network of transcriptional regulators. *Nature*, 403:335–338, 2000.

Michael. B. Elowitz, Arnold. J. Levine, Eric. D. Siggia, and Peter. S. Swain. Stochastic gene expression in a single cell. *Science*, 297:1183–1186., 2002.

R. Feiler, R. Bjornson, K. Kirschfeld, D. Mismer, G. M. Rubin, D. P. Smith, M. Socolich, and C. S. Zuker. Ectopic expression of ultraviolet-rhodopsins in the blue photoreceptor cells of drosophila: visual physiology and photochemistry of transgenic animals. *The Journal of Neuroscience*, 12(10):3862–3868, October 1992.

Stephen P. A. Fodor, Richard P. Rava, Xiaohua C. Huang, Ann C. Pease, Christopher P. Holmes, and Cynthia L. Adams. Multiplexed biochemical assays with biological chips. *Nature*, 364:555–556, August 1993.

Nir Friedman, Michal Linial, Iftach Nachman, and Dana Pe'er. Using bayesian networks to analyze expression data. *Journal of Computational Biology*, 7(3-4):601–620, 2000.

Timothy S. Gardner, Charles R. Cantor, and James J. Collins. Construction of a genetic toggle switch in *Escherichia coli*. *Nature*, 403:339–342, January 2000.

Audrey P. Gasch, Paul T. Spellman, Camilla M. Kao, Orna Carmel-Harel, Michael B. Eisen, Gisela Storz, David Botstein, and Patrick O. Brown. Comprehensive identification of cell cycle-regulated genes of the yeast *Saccharomyces cerevisiae* by microarray hybridization. *Molecular Biology of the Cell*, 11(12):4241–4257, december 2000.

J. Gebert, N. Radde, and G.-W. Weber. Modeling gene regulatory networks with piecewise linear differential equations. *European Journal of Operational Research*, 181(3): 1148 – 1165, 2007. ISSN 0377-2217.

S. Ghaemmaghami and Won-Ki Huh et al. Global analysis of protein expression in yeast. *Nature*, 425:737–741, 2003.

Michael A. Gibson and Jehoshua Bruck. Efficient exact stochastic simulation of chemical systems with many species and many channels. *Journal of Physical Chemistry A*, 104 (9):1876–1889, March 2000.

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

Daniel T. Gillespie. Approximating the master equation by fokker-planck-type equations for single-variable chemical systems. *Journal of Chemical Physics*, 72(10):5363–5370, May 1980.

Daniel T. Gillespie. The multivariate langevin and fokker-planck equations. *American Journal of Physics*, 64(10):1246–1257, Oct 1996.

Daniel T. Gillespie. The chemical langevin equation. *Journal of Chemical Physics*, 113 (1):297–306, Jul 2000.

Daniel T. Gillespie. Approximate accelerated stochastic simulation of chemically reacting systems. *Journal of Chemical Physics*, 115(4):1716–1733, July 2001.

Ido Golding, Johan Paulsson, Scott M. Zawilski, and Edward C. Cox. Real-time kinetics of gene activity in individual bacteria. *Cell*, 123:1025–36, 2005.

Nicholas J. Guido, Xiao Wang, David Adalsteinsson, David McMillen, Jeff Hasty, Charles R. Cantor, Timothy C. Elston, and J. J. Collins. A bottom-up approach to gene regulation. *Nature*, 439(7078):856–860, February 2006. ISSN 1476-4687.

S. Holmberg and P. Schjerling. Cha4p of *Saccharomyces Cerevisiae* activates transcription via serine/threonine response elements. *Genetics*, 144(2):467ñ478, 1996.

M. De Hoon, S. Imoto, and S. Miyano. Inferring gene regulatory networks from time-ordered gene expression data using differential equations. *Lecture Notes in Computer Science*, 2534:267274, 2002.

Sara Hooshangi, Stephan Thiberge, and Ron Weiss. Ultrasensitivity and noise propagation in a synthetic transcriptional cascade. *Proceedings of the National Academy of Sciences of the United States of America*, 102(10):3581–3586, 2005.

Piers J. Ingram, Michael P. H. Stumpf, and Jaroslav Stark. Nonidentifiability of the source of intrinsic noise in gene expression from single-burst data. *PLoS Computational Biology*, 4(10):e1000192, 2008.

Franois Jacob and Jacques Monod. Genetic regulatory mechanisms in the synthesis of proteins. *Journal of Molecular Biology*, 3:318–356, 1961.

Shiraz Kalir, Shmoolik Mangan, and Uri Alon. A coherent feed-forward loop with a sum input function prolongs flagella expression in *Escherichia coli*. *Molecular Systems Biology*, pages 1–5, March 2005.

Joel Keizer. *Statistical Thermodynamics of Nonequilibrium Processes*. Springer-Verlag, 1987.

T B Kepler and T C Elston. Stochasticity in transcriptional regulation: origins, consequences, and mathematical representations. *Biophysical Journal*, 81(6):31163136, December 2001.

Sun Yong Kim, Seiya Imoto, and Satoru Miyano. Inferring gene networks from time series microarray data using dynamic bayesian networks. *Briefings in Bioinformatics*, 4(3):228–235, September 2003.

O. Kobayashi, H. Suda, T. Ohtani, and H. Sone. Molecular cloning and analysis of the dominant flocculation gene flo8 from *Saccharomyces cerevisiae*. *Molecular Genetics and Genomics*, 251(6):707–715, July 1996.

O. Kobayashi, H. Yoshimoto, and H. Sone. Analysis of the genes activated by the flo8 gene in *Saccharomyces cerevisiae*. *Current Genetics*, 36(5):256–261, 1999.

Jean-Christophe Leloup and Albert Goldbeter. A model for circadian rhythms in drosophila incorporating the formation of a complex between the per and tim proteins. *Journal of Biological Rhythms*, 13(1):70–87, 1998.

Robert J. Lipshutz, Stephen P.A. Fodor, Thomas R. Gingeras, and David J. Lockhart. High density synthetic oligonucleotide arrays. *Nature*, 21:20–24, January 1999.

Shi V. Liu. Debating cell-synchronization methodologies: further points and alternative answers. *Trends in Biotechnology*, 23(1):9 – 10, 2005. ISSN 0167-7799.

Shi V. Liu. Method and apparatus for producing age-synchronized cells, US patent US6767734B, July 2004.

Hedia Maamar, Arjun Raj, and David Dubnau. Noise in gene expression determines cell fate in *Bacillus subtilis*. *Science*, 317:526–529, July 2007.

S. Mangan, S. Itzkovitz, A. Zaslaver, and U. Alon. The incoherent feed-forward loop accelerates the response-time of the gal system of escherichia coli. *Journal of Molecular Biology*, 356(5):1073 – 1081, 2006. ISSN 0022-2836.

Umberto M. Marconi, Andrea Puglisi, Lamberto Rondoni, and Angelo Vulpiani. Fluctuation-dissipation: Response theory in statistical physics. *Physics Reports*, 461 (4-6):111–195, June 2008.

Harley H. McAdams and Adam Arkin. Stochastic mechanisms in gene expression. *PNAS, Biochemistry*, 94:814–819, February 1997.

R. Milo and S. Shen-Orr et al. Network motifs: Simple building blocks of complex networks. *Science*, 298:824–827, 2002.

Pedro T. Monteiro and Nuno D. Mendes et al. Yeastract-discoverer: new tools to improve the analysis of transcriptional regulatory associations in *Saccharomyces cerevisiae*. *Nucl. Acids Res.*, 36:D132–136, 2008.

Dale Muzzey and Alexander van Oudenaarden. Quantitative time-lapse flourescence microscopy in single cells. *Annual Review of Cell and Developmental Biology*, 25(1): 13.1–13.27, 2009.

Iftach Nachman, Aviv Regev, and Sharad Ramanathan. Dissecting timing variability in yeast meiosis. *Cell*, 131(3):544–556, November 2007.

Mariana Nacht, Anne T. Ferguson, Wen Zhang, Joseph M. Petroziello, Brian P. Cook, Yu Hong Gao, Sharon Maguire, Deborah Riley, George Coppola, Gregory M. Landes, Stephen L. Madden, and Saraswati Sukumar. Combining Serial Analysis of Gene Expression and Array Technologies to Identify Genes Differentially Expressed in Breast Cancer. *Cancer Res*, 59(21):5464–5470, 1999.

R.B. Nelson. Simplified calculation of eigenvector derivatives. *AIAA Journal*, 14(9): 1201–1205, Sep 1976.

John R S Newman, Sina Ghaemmaghami, Jan Ihmels, David K Breslow, Matthew Noble, Joseph L DeRisi, and Jonathan S Weissman. Single-cell proteomic analysis of s. cerevisiae reveals the architecture of biological noise. *Nature*, 441:840–846, 2006.

Aaron Novick and Milton Weiner. Enzyme induction as an all-or-none phenomenon. *Proceedings of the National Academy of Sciences of the United States of America*, 43 (7):553–566, 1957.

Lars Onsager. Reciprocal relations in irreversible processes. i. *Phys. Rev.*, 37(4):405–426, Feb 1931.

E. M. Ozbudak, M. Thattai, I. Kurtser, A. D. Grossman, and A. van Oudenaarden. Regulation of noise in the expression of a single gene. *Nature Genetics*, 31:69–73, May 2002.

Willmar D. Patino, Omar Y. Mian, and Paul M. Hwang. Serial Analysis of Gene Expression: Technical Considerations and Applications to Cardiovascular Biology. *Circ Res*, 91(7):565–569, 2002.

Johan Paulsson. Summing up the noise in gene networks. *Nature*, 427(415-18), 2004.

Johan Paulsson. Models of stochastic gene expression. *Physics of Life Reviews*, 2: 157–175, 2005.

Juan M. Pedraza and A. van Oudenaarden. Noise propagation in gene networks. *Science*, 307:1965–1969, 2005.

Bruno-Edouard Perrin, L. Ralaivola, A. Mazurie, S. Bottani, J. Mallet, and F. d'Alche Buc. Gene networks inference using dynamic bayesian networks. *Bioinformatics*, 19(Suppl.2):ii138–ii148, 2003.

Mark Ptashne and Alexander Gann. Transcriptional activation by recruitment. *Nature*, 386:569 – 577, april 1997.

A. Raj, P. van den Bogaard, S. A. Rifkin, A. van Oudenaarden, , and S. Tyagi. Imaging individual mrna molecules using multiple singly labeled probes. *Nature Methods*, 9: 877–879, 2008.

Arjun Raj, Charles S. Peskin, Daniel Tranchina, Diana Y. Vargas, and Sanjay Tyagi. Stochastic mrna synthesis in mammalian cells. *PLoS Biology*, 4:e309(10):1707–1719, September 2006.

Jonathan M. Raser and Eric K. O'Shea. Control of stochasticity in eukaryotic gene expression. *Science*, 304(5678):1811–1814, 2004.

L.C. Rogers. Derivatives of eigenvalues and eigenvectors. *AIAA Journal*, 8(5):943–944, May 1970.

Nitzan Rosenfeld and Uri Alon. Response delays and the structure of transcription networks. *Journal of Molecular Biology*, 329(4):645 – 654, 2003. ISSN 0022-2836.

Nitzan Rosenfeld, Jonathan W. Young, Uri Alon, Peter S. Swain, and Michael B. Elowitz. Gene regulation at the single-cell level. *Science*, 307(5717):1962–1965, 2005.

E. Sakamoto and H. Iba. Inferring a system of differential equations for a gene regulatory network by using genetic programming. *Proceedings of the Congress on Evolutionary Computation*, page 720726, 2001.

Katsuhiko Sato, Yoichiro Ito, Tetsuya Yomo, and Kunihiko Kaneko. On the relation between fluctuation and response in biological systems. *Proceedings of the National Academy of Sciences of the United States of America*, 100(24):14086–14090, 2003.

Mark Schena, Dari Shalon, Ronald W. Davis, and Patrick O. Brown. Quantitative monitoring of gene expression patterns with a complementary dna microarray. *Science*, 270(5235):467–470, 1995.

Eran Segal, Michael Shapira, Aviv Regev, Dana Pe'er, David Botstein, Daphne Koller, and Nir Friedman. Module networks: identifying regulatory modules and their condition-specific regulators from gene expression data. *Nature Genetics*, 34:166–176, May 2003.

Kerby Shedden and Stephen Cooper. Analysis of cell-cycle gene expression in *Saccharomyces cerevisiae* using microarrays and multiple synchronization methods. *Nucl. Acids Res.*, 30(13):2920–2929, 2002.

Shai S. Shen-Orr, Ron Milo, Shmoolik Mangan, and Uri Alon. Network motifs in the transcriptional regulation network of *Escherichia coli*. *Nature Genetics*, 31:64–68, May 2002.

Alex Sigal, Ron Milo, Ariel Cohen, Naama Geva-Zatorsky, Yael Klein, Inbal Alaluf, Naamah Swerdlin, Natalie Perzov, Tamar Danon, Yuvalal Liron, Tal Raveh, Anne E Carpenter, Galit Lahav, and Uri Alon. Dynamic proteomics in individual human cells uncovers widespread cell-cycle dependence of nuclear proteins. *Nature Methods*, 3: 525–531, June 2006a.

Alex Sigal, Ron Milo, Ariel Cohen, Naama Geva-Zatorsky, Yael Klein, Yuvalal Liron, Nitzan Rosenfeld, Tamar Danon, Natalie Perzov, and Uri Alon. Variability and memory of protein levels in human cells. *Nature*, 444:643–646, November 2006b.

Alexander Soukas, Paul Cohen, Nicholas D. Socci, and Jeffrey M. Friedman. Leptin-specific patterns of gene expression in white adipose tissue. *Genes & Development*, 14(8):963–980, 2000.

T. Speck and U. Seifert. Restoring a fluctuation-dissipation theorem in a nonequilibrium steady state. *Europhys. Lett.*, 74(3):391–396, 2006.

Paul T. Spellman and Gavin Sherlock. Final words: cell age and cell cycle are unlinked. *Trends in Biotechnology*, 22(6):277 – 278, 2004a. ISSN 0167-7799.

Paul T. Spellman and Gavin Sherlock. Reply: whole-culture synchronization - effective tools for cell cycle studies. *Trends in Biotechnology*, 22(6):270 – 273, 2004b. ISSN 0167-7799.

Paul T. Spellman, Gavin Sherlock, Michael Q. Zhang, Vishwanath R, Kirk Anders, Michael B. Eisen, Patrick O. Brown, David Botstein, and Bruce Futcher. Comprehensive identification of cell cycle-regulated genes of the yeast *Saccharomyces cerevisiae* by microarray hybridization. *Molecular Biology of the Cell*, 9:3273–3297, december 1998.

David Sprinzak and Michael B. Elowitz. Reconstruction of genetic circuits. *Nature Reviews*, 438:443–448, November 2005.

Peter S. Swain, Michael B. Elowitz, and Eric D. Siggia. Intrinsic and extrinsic contributions to stochasticity in gene expression. *PNAS*, 99(20):1279512800, October 2002.

Yi Tao, Xiudeng Zheng, and Yuehua Sun. Effect of feedback regulation on stochastic gene expression. *Journal of Theoretical Biology*, 247:827–836, 2007.

Miguel C. Teixeira and P. Monteiro et al. The yeastract database: a tool for the analysis of transcription regulatory associations in *Saccharomyces cerevisiae*. *Nucl. Acids Res.*, 34:D446–451, 2006.

Mukund Thattai and Alexander van Oudenaarden. Intrinsic noise in gene regulatory networks. *PNAS*, 98(15):8614–8619, July 2001.

Mukund Thattai and Alexander van Oudenaarden. Attenuation of noise in ultrasensitive signaling cascades. *Biophysical Journal*, 82(6):2943–2950, June 2002.

Ryota Tomioka, Hidenori Kimura, Tetsuya J. Kobayashi, and Kazuyuki Aihara. Multivariate analysis of noise in genetic regulatory networks. *Journal of Theoretical Biology*, 229:501–521, 2004.

Roger Y. Tsien. The green fluorescent protein. *Annual Review of Biochemistry*, 67(1): 509–544, 1998.

N. G. van Kampen. *Stochastic Processes in Physics and Chemistry*. Elsevier, North-Holland Personal Library, third edition, 2007.

Diana Y. Vargas, Arjun Raj, Salvatore A. E. Marras, Fred Russell Kramer, and Sanjay Tyagi. Mechanism of mrna transport in the nucleus. *Proceedings of the National Academy of Sciences of the United States of America*, 102(47):17008–17013, 2005.

Victor E. Velculescu, Lin Zhang, B. Vogelstein, and K. W. Kinzler. Serial analysis of gene expression. *Science*, 270(5235):484–487, October 1995.

Jos M. G. Vilar, Calin C. Guet, and Stanislas Leibler. Modeling network dynamics: the lac operon, a case study. *J. Cell Biol.*, 161(3):471–476, 2003.

Yulei Wang, Chih Long Liu, John D. Storey, Robert J. Tibshirani, Daniel Herschlag, and Patrick O. Brown. Precision and functional specificity in mrna decay. *PNAS*, 99 (9):5860–5865, 2002.

Mathias F. Wernet, Esteban O. Mazzoni, Arzu elik, Dianne M. Duncan, Ian Duncan, and Claude Desplan. Stochastic spineless expression creates the retinal mosaic for colour vision. *Nature*, 440:174–180, March 2006.

Gad Yagil and Ezra Yagil. On the relation between effector concentration and the rate of induced enzyme synthesis. *Biophysical Journal*, 11(1):11–27, January 1971.

Ji Yu, Jie Xiao, Xiaojia Ren, Kaiqin Lao, and Sunney X. Xie. Probing gene expression in live cells, one protein molecule at a time. *Science*, 311(5767):1600–1603, 2006.

Etay Ziv, Ilya Nemenman, and Chris H. Wiggins. Optimal signal processing in small stochastic biochemical networks. *PLoS ONE*, 2(10):e1077, 10 2007.