Correlated fluctuations carry signatures of gene regulatory network dynamics

Vijayanarasimha H. Pakka, Adam Prügel-Bennett, Srinandan Dasmahapatra[*]

School of Electronics and Computer Science, University of Southampton, Southampton SO17 1BJ, United Kingdom

Abstract

The dynamics of transcriptional control involve small numbers of molecules and result in significant fluctuations in protein and mRNA concentrations. The correlations between these intrinsic fluctuations then offer, via the fluctuation dissipation relation, the possibility of capturing the systems response to external perturbations, and hence the nature of the regulatory activity itself. We show that for simple regulatory networks of activators and repressors, the correlated fluctuations between molecular species show distinct characteristics for changes in regulatory mechanism and for changes to the topology of causal influence. Here, we do a stochastic analysis and derive time-dependent correlation functions between molecular species of regulatory networks and present analytical and numerical results on peaks and delays in correlations between proteins within networks. Upon using these values of peaks and delays as a 2-dimensional feature space, we find that different regulatory mechanisms separate into distinct clusters. This indicates that experimentally observable pairwise correlations can distinguish between gene regulatory networks.

Keywords: transcriptional regulation, gene regulatory networks, correlated fluctuations, fluctuation-dissipation relations, master equation

[*]Corresponding author
  Email addresses: `vhp05r@ecs.soton.ac.uk` (Vijayanarasimha H. Pakka), `apb@ecs.soton.ac.uk` (Adam Prügel-Bennett), `sd@ecs.soton.ac.uk` (Srinandan Dasmahapatra)

1. Introduction

A gene regulatory network (GRN) summarizes the causal influences between genes that regulate the expression of proteins in cells. While the set of genes constitute the nodes of this network, the regulatory activity between genes is represented by its edges. The biochemical steps involved in this activity is inherently stochastic in nature, and fluctuating numbers of mRNA and protein levels have been observed in clonal populations. Many interesting biological phenomena have been attributed to these fluctuations (Arkin et al., 1998; Raser and O'Shea, 2005; Raj and van Oudenaarden, 2008), which include phenotypic variation within an isogenic population of cells or their influence on switching between cell fates. Such examples include the lysis/lysogeny switch in bacteriophage-λ (Arkin et al., 1998) driven by fluctuations in Cro_2 and CI_2; competent-vegetative switching in *B. subtilis* cells (Maamar et al., 2007); and in the regulatory circuit responsible for colour vision in *Drosophila*, where cells commit to either yellow(70%) or pale(30%) ommatidial subtypes (Wernet et al., 2006). Stochastic effects in transcription factor proteins affect the regulation of the genes whose *cis*-regulatory upstream sequence it binds to; this is exemplified in the variability in the time of onset of early meiosis genes in diploid yeast cells subjected to nutritional deprivation linked to fluctuations in the level of Ime1 proteins(Nachman et al., 2007). Cell-cell variability due to transcriptional noise has been shown to be advantageous in adverse environmental conditions in (Blake et al., 2006). While such examples establish the salience of stochasticity in cell biology, how fluctuations are organized and propagate within regulatory networks remains an important issue, and whether different networks leave identifiable traces on the measured fluctuations in protein numbers.

Simple models of gene regulation (Kepler and Elston, 2001) identify experimental interactions that outline the contributions of the core processes in variability of gene expression (Ozbudak et al., 2002). Quantifying molecular fluctuations by such formulations allows one to study the effect of various network topologies, regulatory mechanisms and parameters on the stationary statistics of variables in GRNs (Tao et al., 2007; Tomioka et al., 2004; Pedraza and van Oudenaarden, 2005). Such observations make it possible for the quantification of noise to reveal the regulatory links between genes (Austin et al., 2006; Cox et al., 2008). For example, (Cox et al., 2008) consider single-gene models seeking to characterize regulatory networks by their specific features of their stationary auto-correlations, *viz.* the coefficient of variation and the time at which auto-correlations decay to half their maximum value. However, a crucial element missing in such studies is

the temporal aspect. Since the entire process of gene expression is dynamic in nature, tracking the fluctuations in species numbers over time would yield significant information. 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 temporal correlations that are more likely to be characteristic of the regulatory mechanisms, structure and parameter values of the system. Such studies are possible by tracking of mRNA/protein numbers over time in individual cells, *e.g.,* (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., 2006). That is the object of our study.

A number of papers have studied the single-gene case in great detail (Paulsson, 2004; Lei, 2009; Shahrezaei and Swain, 2008) and have investigated the noise properties of multi-gene networks (Tomioka et al., 2004; Pedraza and van Oudenaarden, 2005). An important difference between these investigations and our current work lies in our reliance on time-delayed correlation functions, not equal time ones and Appendix C presents the simplest case of a single gene in some detail. The motivation behind looking at time delayed correlations lies in linear response theory (van Kampen, 2007).

Given the stochasticity of the very processes that constitute regulatory responses to stimuli, it is likely that correlations between fluctuations will illuminate the dynamics of regulatory interactions. Indeed the intuition behind Onsager's regression hypothesis (Onsager, 1931) is that the regression to an equilibrium state after a short-lived external perturbation or an intrinsic fluctuation are identical, which has been developed further in several fluctuation-dissipation theorems even away from equilibrium (Keizer, 1987; Speck and Seifert, 2006). Regulation of expression determines responses to signals which are organized by the cell to form functional pathways. In this paper, 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.

To achieve that, we study the time-dependent correlation functions between proteins (such as a transcription factor and a downstream target) in a network expressed in a sum-of-exponentials form. We expect distinct behavioural patterns in the correlated fluctuations for *variations* in the regulatory networks. Variations are introduced in elementary fragments of regulatory networks by employing three strategies: (*i*) adopting different regulatory mechanisms; activation/repression with or without co-operative mechanisms for the same topology

3

of dependence between genes, (*ii*) introduction of additional genes into the network, and (*iii*) adopting different sets of values for the reaction rate constants. They amount to varying the regulatory mechanisms, the network structure and the parameter values respectively, and hence form the basis for deriving qualitative relationships between these attributes and features extracted from the fluctuation properties of the network, which we summarize in Figure 3 below. Therefore, by tracking the species numbers in single cells over time, one could predict the type of regulatory activity present in a GRN, as demonstrated here for the case of small networks.

After setting up the machinery of stochastic processes that we need to compute correlation functions in Section 2, we present specific details of the models and the values that we choose for the rate constants in Section 3. Section 4 establishes the dependences of correlations on parameter values in different networks to set the stage for the comparison presented in Section 5.

2. Dynamic Correlation Functions

The relation between the macroscopic dynamics and the fluctuation properties of the system is formalized by the Fluctuation-Dissipation Theorem (FDT). In statistical physics, the FDT states that the linear response of an equilibrium (or steady state) (Keizer, 1987) system to external perturbations is obtained from the correlation functions of the fluctuations in the unperturbed 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 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 (CME) (van Kampen, 2007), where the CME is simplified to the Linear Noise Approximation (LNA). This LNA is equivalent to the Fokker-Planck (F-P) equation for these processes (Gillespie, 2000), whose solution are the distributions of the fluctuations around the average concentrations of the molecules. Here, we derive the dynamic correlations between these molecular fluctuations, via the CME, and on the way observe the emergence of a relationship between the dynamics of the averages and the fluctuating part of the distributions.

Supposing that $\mathbf{X}(t)$ is a vector of the number of molecular species in the system and that the evolution of this vector is a jump-type Markov process, then the time-evolution of the joint probability distribution of all species of a spatially

homogeneous chemical system is given by the CME

$$\frac{\partial}{\partial t}P(\mathbf{X}, t) = \sum_{j=1}^{M} \left[P(\mathbf{X} - \boldsymbol{v}_j, t)a_j(\mathbf{X} - \boldsymbol{v}_j) - P(\mathbf{X}, t)a_j(\mathbf{X}) \right], \tag{1}$$

where the propensity function a_j gives probability for one out of a possible M reactions r_j ($1 \leq j \leq M$) to occur in time $[t, t + dt)$ in volume Ω, thus changing the number of s_i molecules by an amount v_{ij} called the stoichiometry. The initial condition for the system to be $P(\mathbf{X}, t_0) = \delta(\mathbf{X} - \mathbf{X_0})$. Upon multiplying eq. (1) by $\mathbf{X}$ and $\mathbf{X}\mathbf{X}^T$, and summing over all configurations $\mathbf{X}$ after making the substitution $\mathbf{X} = \mathbf{X} - \boldsymbol{v}_j + \boldsymbol{v}_j$, we get the following expressions for the evolution of the mean values:

$$\frac{d\langle X_i \rangle}{dt} = \sum_{j=1}^{M} \langle v_{ij}a_j(\mathbf{X}) \rangle. \tag{2}$$

and covariances:

$$\frac{d\operatorname{Cov}(X_i, X_j)}{dt} = \sum_{k=1}^{M} \left[\langle X_i v_{jk}a_k(\mathbf{X}) \rangle + \langle X_j v_{ik}a_k(\mathbf{X}) \rangle + \langle v_{ik}v_{jk}a_k(\mathbf{X} - \boldsymbol{v}_k) \rangle \right] \tag{3}$$

where $\operatorname{Cov}(X_i, X_j) = \langle X_i X_j \rangle - \langle X_i \rangle \langle X_j \rangle$, so we have subtracted the contributions of the evolution of the means.

These equations do not close on themselves if the dependence of the propensities on particle numbers is non-linear. The rate equations for the averages of the concentrations $x_i = X_i/\Omega$ obey mass action kinetics of the biochemical system,

$$\frac{d\langle x_i \rangle}{dt} = \sum_{j=1}^{M} v_{ij}R_j(\langle \mathbf{x} \rangle) \tag{4}$$

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. Linearizing about the steady states $\langle \mathbf{x} \rangle^s$ which satisfy $\frac{d}{dt}\langle \mathbf{x} \rangle = 0$ in (4), we find that the deviations $\delta\mathbf{x}(t) := \mathbf{x}(t) - \langle \mathbf{x} \rangle^s$ are given by

$$\left(\frac{d}{dt} \right) \delta\mathbf{x} = \mathbf{A}\delta\mathbf{x}, \tag{5}$$

where $\mathbf{A}$ is the Jacobian matrix with elements:

$$A_{ik} = \frac{\partial}{\partial \langle x_k \rangle} \left(\sum_{j=1}^{M} v_{ij}R_j(\langle \mathbf{x} \rangle) \right) \Bigg|_{\langle \mathbf{x} \rangle = \langle \mathbf{x} \rangle^s}. \tag{6}$$

At steady state, the means in the kinetic rates R_j are time-independent, and so are the elements of $\mathbf{A}$ and

$$\mathbf{BB}^T = \boldsymbol{\nu}\,\mathbf{diag}\,(\mathbf{R}(\langle\mathbf{x}^s\rangle))\,\boldsymbol{\nu}^T,$$

yielding the closed form equation for the covariances (from eq. 3):

$$\frac{\partial}{\partial t}\mathbf{C} = \mathbf{AC} + \mathbf{CA}^T + \mathbf{BB}^T.$$

The deviations from steady state are thus governed by a multi-variate Ornstein-Uhlenbeck process (van Kampen, 2007) and we use standard techniques to solve for the two-time correlations $\mathbf{C}(t, t+\tau)$:

$$\langle\,\delta\mathbf{x}(t+\tau)\,\delta\mathbf{x}(t)^T\,\rangle = \mathbf{Y}(\tau).\mathbf{C}(t),\ \text{with} \tag{7}$$
$$\frac{d\mathbf{Y}(\tau)}{d\tau} = \mathbf{A}.\mathbf{Y}(\tau), \tag{8}$$

where the Green's function $\mathbf{Y}(\tau)$ with $\mathbf{Y}(0) = 1$ determines the evolution of the deviations from steady state. The variable t takes values in the regime where steady state has been reached and hence the stationary covariance $\mathbf{C}(t)$ is t-independent. Using the independence and bi-orthogonality properties of the eigenvectors of the non-symmetric matrix $\mathbf{A}$, it is possible (van Kampen, 2007; Elf and Ehrenberg, 2003) to write the time-covariance function between two molecular species x_i and x_j as:

$$\langle\,\delta x_i(t+\tau)\,\delta x_j(t)\,\rangle = \sum_k Y_{ik}(\tau)C_{kj} = \sum_k\sum_l e^{\lambda_l\tau}U^T_{lk}V_{il}C_{kj}. \tag{9}$$

Here, the elements of $\mathbf{V}$ and $\mathbf{U}^T$, the matrices comprising of the right and left eigenvectors of $\mathbf{A}$ respectively and the eigenvalues λ_i are explicitly written out.

For effective comparison of results from different GRNs, we use time-delayed correlations by normalizing the above covariances. 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.

$$\text{Corr}\,[\delta x_i(t+\tau), \delta x_j(t)] = \frac{\langle\,\delta x_i(t+\tau)\,\delta x_j(t)\,\rangle}{\sqrt{\langle\,(\delta x_i(t+\tau))^2\,\rangle\,\langle\,(\delta x_j(t))^2\,\rangle}} \tag{10}$$

The form of the covariance function brings out the existing relation between the covariances and the averaged responses of the system. The system Jacobian $\mathbf{A}$ and diffusion matrix $\mathbf{BB}^T$ are derived through the stoichiometry $\boldsymbol{\nu}$ and reaction

6

rates **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. These have been separated into contributions characteristic of individual genes in the network (intrinsic noise) and those from various upstream or system-wide elements such as polymerase fluctuations (extrinsic noise) in the literature (Raser and O'Shea, 2005).

From the dynamic correlations we shall extract features for possible identification of the type of regulation present between pairs of genes for likely signatures of network structure from the time-series data from a GRN. To demonstrate the above, we consider regulatory mechanisms such as activation and repression via monomers/dimers and network mechanisms such as cascades, dual activators and feedforward loops. A simple and effective way to model these networks is by choosing a set of elementary reactions that capture the significant steps of the regulatory process. 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. By employing fewer molecular species we intend to reduce the complexity of the model, while retaining the salient features of the complete model.

2.1. Features of time-dependent protein correlations

Positive regulation between any two genes of a network not only induces positive covariances between their respective proteins but also gives a characteristic shape to them. This can be seen in the covariance function in the case of an elementary activator $X \rightarrow Y$ as shown in Figure 1(b), and detailed derivations of which are given in Appendix A. The normalized $\text{Corr}[P_x(t), P_y(t + \tau)]$ function has exactly the same shape as the covariance. The defining *features* of these functions are (i) the stationary covariance $\text{Cov}(\tau = 0)$, (ii) the maximum or peak covariance $\text{Cov}^* = \text{Cov}(\tau = \tau^*)$ or, the maximum value of the time-dependent correlation Corr^* and (iii) τ^*, which is the time taken for the covariance to reach this peak, as indicated in Figure 1(b). (Note that in the following, we assume a single maximum for the covariance as a function of time; it is true for the examples presented in this paper.) It shall be seen in the coming sections that these features are sensitive to the various attributes of the system such as the regulatory mechanism, network structure and the rate constants, and therefore act as tools for the identification of these attributes. For example, in Appendix C we show how the feature τ^* is sensitive in the rate constants by deriving its expression in the case of a single gene.

In the case of $X \to Y$, $\mathrm{Corr}[P_x(t), P_y(t+\tau^*)] = 0.8$ and $\tau^* = 49$ minutes. Physically, this means that at equilibrium, the deviations δP_y from the average protein concentration $\langle P_y \rangle$ are influenced mostly by δP_x occuring 49 minutes earlier to it. This time delay is the 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 not indicate of any such causality. This is supported by the 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. Therefore, stationary correlation alone cannot in effect predict the direction of the *causal* link between 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 reason behind the non-zero value of the stationary correlations is the slow decaying auto-correlation function of P_x, due mainly to proteolysis. 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 $(> t_1)$.

Mathematically, the shape of the dynamic covariance curve is due to its functional form $\sum_{l=1}^{N} h_l e^{\lambda_l \tau}$, where the h's are some functions in the eigenvectors of the Jacobian. In the case of a single activator (Appendix A), the eigenvalues are the four decay rates, hence the covariance function is the sum of exponentials in time raised to the negative power of these decay rates. However, since the co-efficients of these exponentials are different functions of the eigenvectors, their values are different, leading to the characteristic shape of their resulting sum. These exponentials normalized between ± 1 and 0 are shown in Figure 1(a).

2.2. Comparison with Results from Stochastic Simulations

The objective here is to compare the dynamic correlations obtained through the analytical formulation with those obtained by Monte Carlo simulations (Gillespie, 1976). We do this in the case of the elementary activator system $X \to Y$, where correlations are drawn between the proteins P_x and P_y at steady-state conditions. At low copy numbers of molecules the question of the applicability of analytical results come to fore, which we address here by assuming the following mean concentration levels of: $\langle M_x \rangle = 1.0 \; nM$, $\langle M_y \rangle = 2.0 \; nM$, $\langle P_x \rangle = 20.0 \; nM$ and $\langle P_y \rangle = 100.0 \; nM$. The values for the rate constants are chosen accordingly: $k_{M_x}^+ = 0.0407 \; nM \; min^{-1}$, $k_{P_x}^+ = 0.367 \; min^{-1}$, $k_{M_y}^+ = 0.308 \; min^{-1}$, $k_{P_y}^+ = 0.88 \; min^{-1}$, $k_{on} = 1.0 \; nM^{-1} min^{-1}$, $k_{off} = 20.0 \; min^{-1}$, $k_{M_x}^- = 0.0407 \; min^{-1}$, $k_{M_y}^- = 0.0770 \; min^{-1}$, $k_{P_x}^- = 0.0184 \; min^{-1}$, $k_{P_y}^- = 0.0176 \; min^{-1}$. The ensemble size of
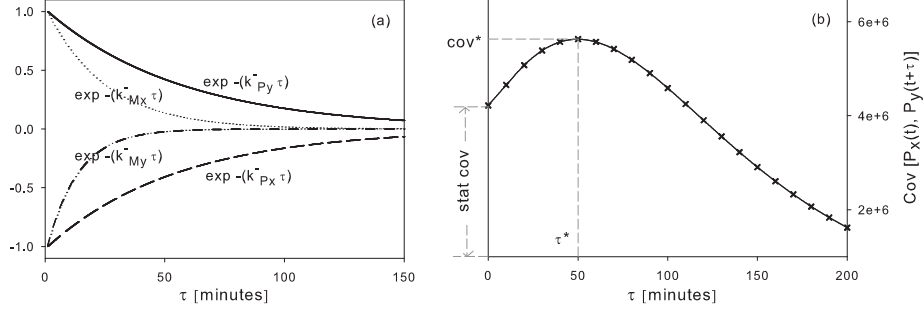
Figure 1: (a) The normalized exponentials in time raised to the negative power of the decay rates, in the case of the elementary activator system. The normalization is performed by dividing the exponentials with their respective co-efficients which are 4.18×10^8 for $k_{P_y}^-$, 1.0×10^7 for $k_{M_x}^-$, -6.34×10^5 for $k_{M_y}^-$, and -4.23×10^8 for $k_{P_x}^-$. (b) The sum of the four exponentials along with their co-efficients results in the dynamic covariances.

the system is equivalent to the number of runs of the simulation program. In Figure 2 we notice that as the ensemble size increases, the correlations obtained from simulations match closely with that of the analytical one. The evaluation of correlations from simulated runs, rests on the equivalence between the ensemble-averaging and time-averaging of the stochastic variables.
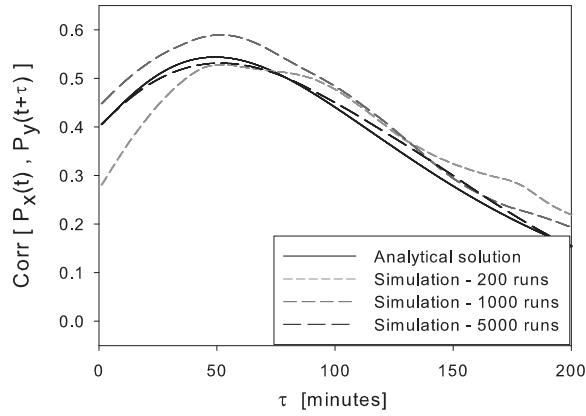

Figure 2: Dynamic correlations between the proteins of the $X \to Y$ regulatory network, obtained through analytics and simulations.

3. Clustering of regulatory networks by correlation features

We investigate the effect that various rate constants have on the features $Corr^*$ and τ^* of the dynamic correlation functions between the regulated P_y and regulator P_x proteins. While stationary correlations, especially in the case of elementary networks, have been the object of study in recent years, we probe the variation in the temporal aspect of the correlations and show their effectiveness as network signatures. In section 5 we show the effect of rate constants on stationary correlation in the case of $X \rightarrow Y$ system. In section 4 we demonstrate the effect of rate constants on dynamic correlations in the case of different regulatory mechansims. The results of this section and section 4 provide sufficient evidence to show that the features of dynamic correlation functions could be effectively used as signatures in identifying regulatory mechanisms of regulatory networks.

Features of the dynamic correlations such as the maximum value of the 2-point correlation, $Corr^*$, and the time τ^* taken to reach it are plotted in the case of different networks. Clustering of these features for variations in rate constants enables one to identify the regulatory mechanism and structure of the network. As a summary of the results obtained, the effectiveness of dynamic correlations in characterizing small gene regulatory networks is seen in the Figure 3. The following section goes into each cluster of points for a network in more detail. Here we point out that, for example, the top two panels indicate that the movement along the time-delay feature τ^* is influenced not by the decay rate of the transcription factor but that of its target. In the case of co-operative activation $X \rightarrow Y \leftarrow Z$ variations were observed in the peak correlation $Corr^*$ whilst leaving unaltered the time delay τ^*, for changes in the rate constants. The clusters corresponding to the cascading activator networks (cascade of activators and cascade of repressors) have a larger value of τ^*, and differ from each other in the magnitude of peak correlations $Corr^*$. By combining the above networks one could compute correlations that capture signatures of more complex regulatory mechanisms such as the coherent and incoherent feedforward networks that have been much studied in the literature (Alon, 2007). Once again these complex networks are clearly identifiable with the aid of the correlation features. Finally, the effect of dimerization is clearly noticed with increased $Corr^*$ values. Due to changes in the shape of these correlation plots, the use of integral over time τ could also act as an additional feature for clustering purposes. However, this is a topic for further investigation.

We have demonstrated that at steady-state conditions dynamic correlations of protein fluctuations could be used to gain insight into the microscopic mechanisms of gene regulation. This means that a lot could be learnt by tracking the
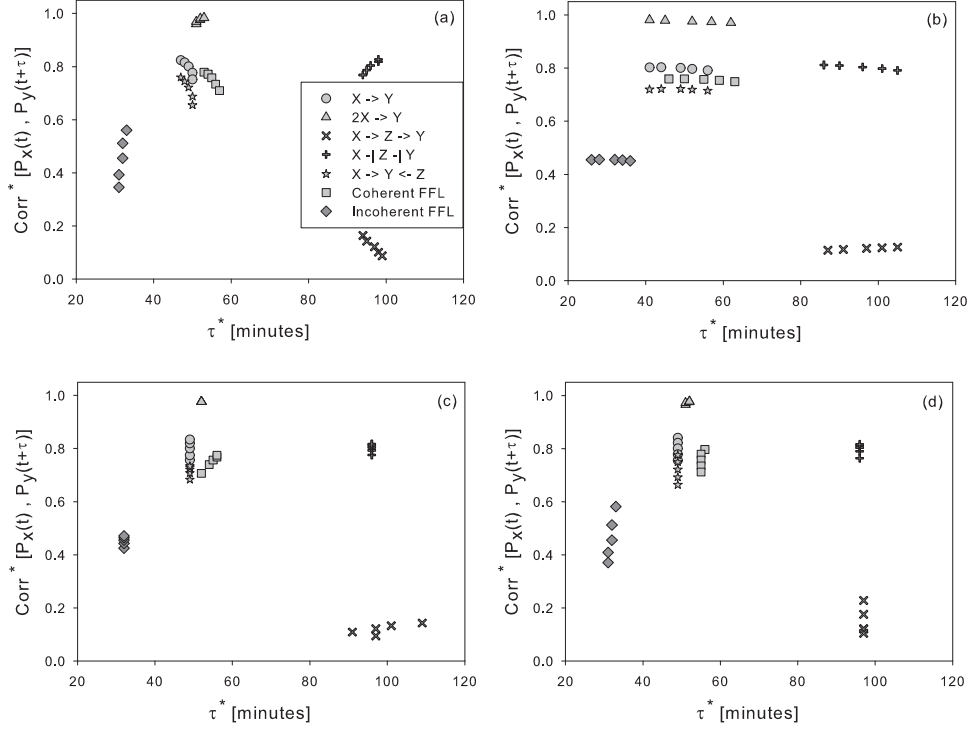
Figure 3: The inset in (a) depicts the networks used in this paper: $(X \rightarrow Y)$ the single activator; $(2X \rightarrow Y)$ a single activator network, but with X coding for a protein which performs the activation after forming a homodimer; $(X \rightarrow Z \rightarrow Y)$ which describes a cascade of interactions, X upregulating Z which upregulates Y; $(X\dashv Z\dashv Y)$ cascading repressors; $(X \rightarrow Y \leftarrow Z)$ the combinatorial action of genes X and Z on Y; (coherent FFL) which combines $X \rightarrow Y \leftarrow Z$ and $X \rightarrow Z$; (incoherent FFL) which combines $X \rightarrow Y \vdash Z$ and $X \rightarrow Z$ (a) $(t_{1/2})_{P_x}$ is varied in steps of $[40, 50, \text{basevalue}, 90, 120]$ minutes, where the basevalue corresponds to a half-life of 65 minutes. (b) Similarly, $(t_{1/2})_{P_y}$ is varied in steps of $[40, 50, \text{basevalue}, 90, 120]$ minutes, where the basevalue corresponds to a half-life of 70 minutes. (c) The transcription rate or the basal transcription rate (in case of repression) $k^+_{M_y}$ is varied as $[0.6, 0.8, 1.0, 1.2, 1.4] \times \text{basevalue}$. (d) The rate of protein translation $k^+_{P_x}$ is varied in steps as $[0.6, 0.8, 1.0, 1.2, 1.4] \times \text{basevalue}$.

11

evolution of mRNA or protein numbers in single cells over time. The most revealing information that can be expected out of such time-series data is the type of regulatory activity present between pairs of genes. As an example, given the information about the lifetimes of proteins and mRNAs of X and Y, the two-time correlations could infer the presence or absence of an intermediary gene. We also demonstrated the effect that various reaction rate constants have on the dynamic correlations between the proteins P_x and P_y, in the case of different regulatory systems. By adopting a learning methodology, the framework of sensitivity analysis could be used for estimating the parameters of a regulatory system given the dynamic correlations between its species. Experimentally this would be equivalent to altering the reaction rates, as done in (Ozbudak et al., 2002) where transcription and translation rates were varied synthetically, and tracking the time-evolution of species in the altered system to yield data for such novel learning approaches.

4. Effect of parameters on dynamic correlations in the case of simple regulatory networks

Fragments of a GRN that represent different regulatory mechanisms such as repression, regulation via dimers, cascading activation, co-operative activation, and so on, not only display different dynamic characteristics in the determinitic domain but also show unique characteristics in their molecular fluctuations. These differences in their fluctuations are captured effectively via the dynamic correlation functions. Some of the elementary networks and their response in correlated fluctuations for a few rate constants are shown below. The detailed set of reactions for the case of an elementary activator are given in Appendix A and for other networks in Appendix B.
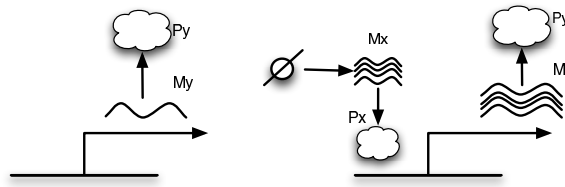
4.1. Elementary Activator

Figure 4: Schematic represention of the elementary activator system. The action of the activator protein P_x increases the amount of M_y transcript produced compared to the basal levels depicted on the left.

12

For time-independent (equal time at steady-state) fluctuations, it has been shown that some rate constants exert influence on both means and variances of protein levels (Thattai and van Oudenaarden, 2001). Our intention here is to study the effect on fluctuations while the mean concentration levels are held constant at the values given in Table A.2. For example, supposing that the pair $[k_{M_x}^+, k_{M_x}^-]$ are at their base values, deviations 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(a). This is suggestive of possible ways in which biological networks could alter their fluctuation properties without affecting the mean concentration levels of proteins and mRNAs.
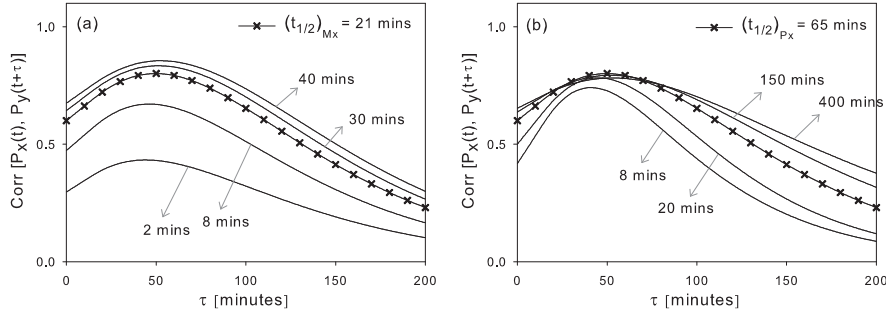

Figure 5: (a) The dynamic correlation between P_x and P_y is sensitive to changes in the production and decay rates of M_x. The base value for the half-life of M_x is 21 minutes. $k_{M_x}^-$ and $k_{M_x}^+$ are varied simultaneously so that their ratio $\langle M_x \rangle = k_{M_x}^+/k_{M_x}^-$ remains constant. 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 $\langle P_x \rangle = \langle M_x \rangle k_{P_x}^+/k_{P_x}^-$ remains constant. While there is not much difference in Corr*, τ^* decreases from 55 to 45 minutes for increase in these parameters. The half-life of M_x and the rate constant $k_{M_x}^-$ are related as follows: $k_{M_x}^- = ln(2) \times [1/(t_{1/2})_{M_x} + 1/t_{double}]$, where $t_{double} = 90$ minutes. We use this definition (as in eq. (A.6)) in the other figures as well.

From Figures 5 and 6 we observe that the feature τ^* is sensitive not only to the decay rates but also in the transcription and translation rates. However, the following analysis makes it clear that τ^* is predominatly sensitive only to the four decay rates. The system has five eigenvalues, 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

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

At steady state, mean value of $G_y = \frac{k_{off}}{k_{off}+k_{on}P_x}$. Therefore $b = \left(G_y + \frac{K_D}{G_y}\right)k_{on} + k_{P_x}^-$

and $c = k_{off} k_{P_x}^- / G_y$. Since K_D is usually of the order of hundreds of nM whereas G_y is less than 1 nM, $b \approx \frac{k_{off}}{G_y} + k_{P_x}^-$. The roots of the quadratic equation are now $-k_{P_x}^-$ and the much larger $-k_{off}/G_y$. The contribution of the large eigenvalues to the sum of exponentials form for the covariance function decays rapidly, which was why the exponential corresponding to $-k_{off}/G_y$, in the Figure 1(a) was not shown.
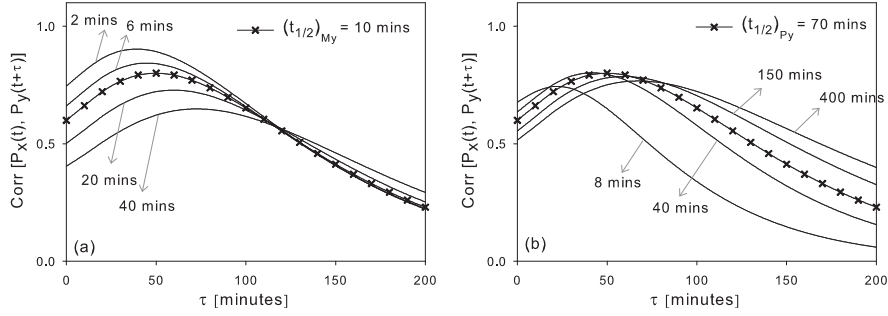

Figure 6: (a) Transcription rate constant $k_{M_y}^+$, and $k_{M_y}^-$ are varied simultaneously in steps, such that $\langle M_y \rangle$ remains constant. While Corr* does change by an appreciable amount, there is a huge decrease in τ^*, from 87 to 43 minutes. (b) Similar is the case for variations 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.

As an extension, only those decay rates that correspond to Y induce more sensitivity in τ^* as shown in Figures 6(a) and 6(b). 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, thus increasing P_y levels at constant rate of production, it becomes difficult to respond to those protein molecules that came into existence due to transcription events triggered by a specific set of P_x molecules. Hence, the influence of any small perturbation in transcription factor quantities at steady state on its target proteins is diminished and delayed (increasing τ^*).

With regard to changes in K_D, the protein correlations remain unchanged, except for $K_D < 200$ nM when there is progressive reduction only in Corr* while τ^* still remains at its base value of 49 minutes. The insensitivity of τ^* to the binding/unbinding reactions is because they are much faster that the other elementary reactions. Further, the correlations are totally insensitive to cases where k_{on} and k_{off} take on different values such that their ratio $K_D = k_{off}/k_{on}$ remains invariant.
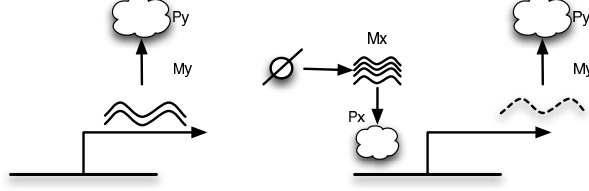
Figure 7: Schematic represention of the elementary repressor system. The basal transcription levels on the left is diminished by the action of the repressor.

4.2. Elementary Repressor

In the elementary fragment $X \dashv Y$ the sensitivity of the dynamic correlations between the repressor P_x and the repressed protein P_y to different values of protein half-life is shown in Figure 8. Note the close similarity between these correlations and those in Figure 6(b). Whilst a change in the regulatory mechanism has influenced the dynamic correlations such that they are now negative in value, the sensitivities of the correlations *w.r.t* the parameters, are very similar to those of the activator case. This similarity 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.
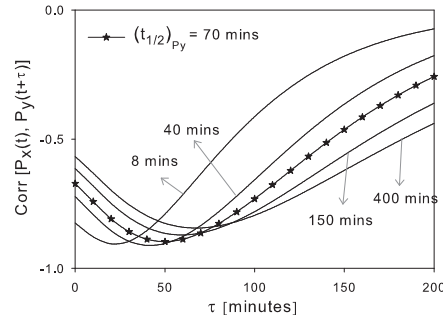
Figure 8: Elementary repression $X \dashv Y$. $[k_{P_y}^+, k_{P_y}^-]$ are varied simultaneously so that the mean concentration level of $P_y = M_y k_{P_y}^+ / k_{P_y}^-$ remains constant. While there is an increase in Corr*, τ^* decreases from 102 to 36 minutes for step-wise increase in these parameters.

4.3. Activation via Dimerization

In the case of activation via dimers $2X \rightarrow Y$, for short half-lives of P_x (about 8 minutes), the effect of dimerization begins to wear off (Figure 10(a)), as P_x

15

Figure 9: Schematic representation of the activator system where dimers are involved. As before, the figure on the left indicates basal transcription.
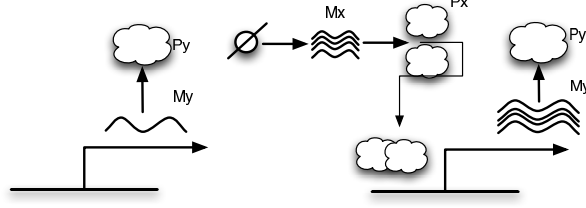
molecules are now being produced and degraded faster (due to increase in values of $[k_{P_x}^+, k_{P_x}^-]$) and thus not able to contribute to dimerization. On the other hand, in Figure 10(b) we do not see any such effect upon reductio in the half-life of P_y due to its non-participation in the dimerization process.
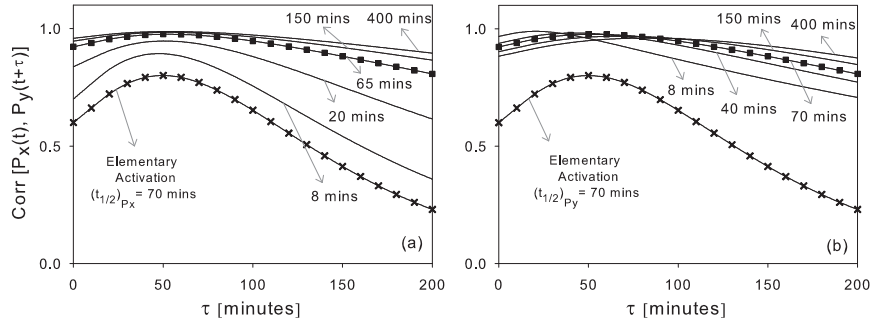

Figure 10: Activation via Dimers $2X \rightarrow Y$. (a) $k_{P_x}^-$ is varied in steps from its base value of 0.0184 min^{-1} that corresponds to a half-life of $(t_{1/2})_{P_x} = 65$ minutes. $k_{P_x}^+$ is simultaneously varied such that $\langle P_x \rangle$ remains unchanged. For increase in the 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 11. 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 localized peak 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 corresponding effect on P_y over a long period of time, *i.e.*, the effect of
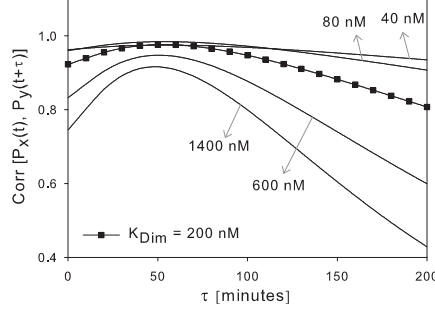
16

Figure 11: Activation via dimers $2X \rightarrow Y$. The base value of the dimer dissociation constant K_{Dim} is fixed at $200\,nM$, and is further assigned values of $40, 80, 600$ and $1400\,nM$. An equivalent variation in the value of $k_{M_y}^+$ is done such that $\langle M_y \rangle$ remains unchanged. For increase in the values of $[k_{M_y}^+, K_{Dim}]$, τ^* decreases marginally from 53 to 47 minutes.

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 $\text{Cov}[P_{x_2}(t), P_y(t + \tau)]$ and $\text{Cov}[P_x(t), P_y(t + \tau)]$ 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.

4.4. Cascading Activation

In the case of the three-gene cascaded activation $X \rightarrow Z \rightarrow Y$, the mRNA and protein decay rates of the intermediate gene Z have an effect on the τ^* of the correlation plots (Figure 12(b)). This is due to the fact that while four of the eight eigenvalues of the system are $[-k_{M_x}^-, -k_{M_z}^-, -k_{M_y}^-, -k_{P_y}^-]$, the other four are the roots to two quadratic equations and are approximated as $[-k_{P_x}^-, -k_{P_z}^-, -\frac{k_{r_1}}{G_z}, -\frac{k_{r_2}}{G_y}]$. Therefore the element Z further sensitizes the temporal characteristics of the correlations between the input and output elements of the GRN. In this regard it is interesting to note that, for the hypothetical case of mRNA and protein half-lives of Z of around 1 minute each, $\tau_{new}^* \approx 49$ as in the case of $X \rightarrow Y$, suggesting that for rapidly decaying mRNA/proteins, the intermediate gene-node ceases to influence the dynamic character of the correlations. Another form of cascaded activation $X \dashv Z \dashv Y$ throws up similar behaviour except that now the magnitude of correlations are higher.
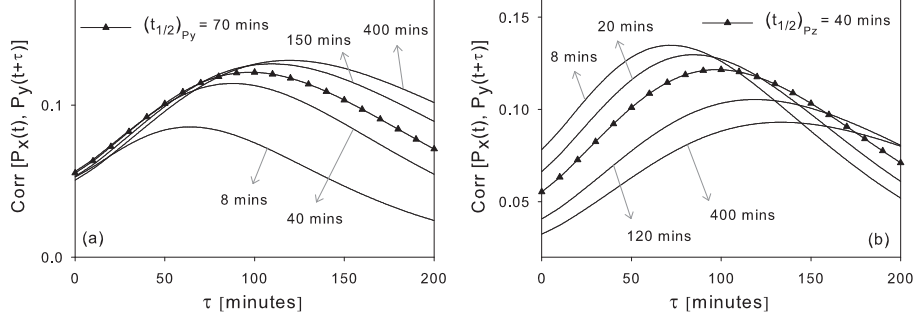
17

Figure 12: Cascaded Activation $X \rightarrow Z \rightarrow Y$. (a) For higher values of protein half-life $(t_{1/2})_{P_x}$, τ^* increases as $[63, 87, 97, 109, 118]$ minutes along with an increase in Corr*. (b) Similarly, for changes in $k_{P_z}^-$ or rather the half-life of P_z, τ^* increases as $[71, 84, 97, 118, 133]$ minutes along with a decrease in both Corr* and the stationary correlation (at $\tau = 0$).

5. Effect of parameters on stationary correlations in the case of $X \rightarrow Y$

The stationary covariance between molecular species is a function of system parameters that are the reaction rate constants. We show this in Appendix C in the case of a simple model of transcription. (Ozbudak et al., 2002) experimentally demonstrated the validity of the expression of protein variance of one such model by altering the transcription and translation rates. Variations in parameter values therefore provide us with some knowledge with regard to the system at hand. By varying each of the ten rate constants individually, over a vast range, we note the change in value of the stationary correlation Corr$[P_x(t), P_y(t)]$.

From Figure 13(a), we see that the correlations are smaller for lower values of the decay rates. 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 larger decay rates, due to the fewer number of molecules present, any fluctuations in them are highly correlated. The increase in the stationary correlation is greater in the range where the decay rates are smaller in value. For monotonically increasing $k_{M_x}^-$ and $k_{P_x}^-$, the stationary correlation decreases at a later stage. This could possibly point towards the network structure which in this case is $X \rightarrow Y$.

Similar reasoning suffices for the variation in the correlation for changes in the other parameters. Firstly, for lower values of k_{on} and for higher values of k_{off}, stationary correlations are high. For higher values of $K_D = k_{off}/k_{on}$, the transcription factor binds less to the DNA-complex thereby reducing the rate of
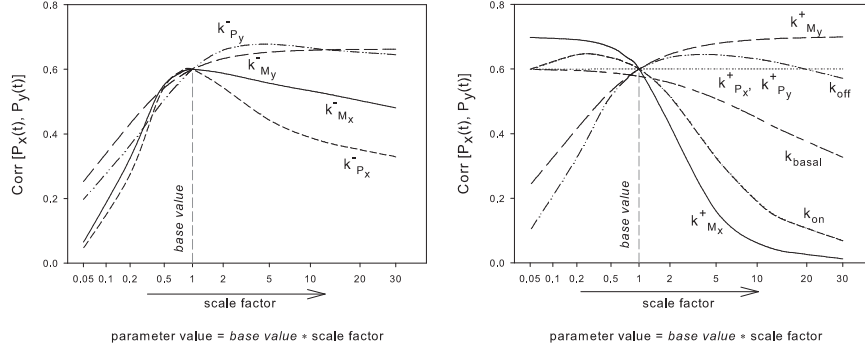
18

Figure 13: (a) Protein stationary correlation varies for changes in the decay rates whilst keeping the other parameters fixed at their base values. The decay rates are varied from $\{0.05 \times base\ value\}$ to $\{30 \times base\ value\}$, where base values are those of Table A.1. (b) The transcription, translation and binding/unbinding rates too have a corresponding influence on the stationary correlation.

production of M_y and in turn 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 $k_{M_y}^+$, as expected, protein correlations are high. On introducing the basal transcription rate k_{basal} as an additional parameter to the system, we notice that it has influence only over the magnitude of the covariances, whether stationary or dynamic, and not on the time-dealy of the peak τ^* itself. Finally, in these feedforward networks translation rates have no effect on the stationary correlations since they do not directly contribute to the transcription process and the proportionate increases in magnitudes of the covariances are normalized away. 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 fluctuations as observed in the correlations provide significant information regarding the rate constants of the reaction processes. The other complementary reason is that, such analyses act as analytical tools that aid in designing new synthetic regulatory networks whose fluctuation properties and performance in general, could be controlled.

6. Conclusions

In this paper, we set out to describe the temporal dependence of the propagation of fluctuations in small gene regulatory networks. In setting up the models, we have introduced the processes of promoter interactions, followed by tran-

19

scription and translation and investigated the contribution of these elements to the fluctuation properties. There are further time dependent effects that we have disregarded, such as the time for the first complete transcription, following potential pauses of RNA polymerases, entry of the transcripts into the cytoplasm to be translated by the ribosomes and re-entry of transcription factors into the nucleus, and potential effects due to post-transcriptional (such as due to miRNA) or post-translational (such as phosphorylation of the proteins for activity priming). Some of these processes occur at smaller time scales, and just as we could eliminate effects of binding and unbinding of proteins to form dimers on the overall characteristics of the correlation profile, we expect some of them to have a small effect on the location of the peak for instance. It is clear, however, that for many regulatory networks, such as those underlying the circadian clock system (Leloup and Goldbeter, 2004), phosphorylation of proteins are critically linked to the time of response. It should be possible to incorporate such effects either via the introduction of further components or by explicitly incorporating time delays and studying the effects of having additional exponentials in delay time in the spectral properties of the Jacobian. That is beyond the scope of this paper.

A key assumption behind the techniques explored in this paper is the possibility of being able to tag and monitor in a continuous fashion, the quantities of proteins or mRNA in a single cell over time. Fluorescent reporters have increased our ability to observe individual cells, which is a major shift from the microarray realm where bulk averages were the measured quantities. On the other hand, technologies such as flow cytometry measures the relative fluorescence intensities of cell populations enabling one to compute histograms of protein fluorescence levels and infer phenotypic distributions in clonal populations (Ozbudak et al., 2002; Elowitz et al., 2002; Raser and O'Shea, 2004) studied under varying parameter values. However, such measurements leave out the temporal aspect the distributions obtained miss crucial elements of 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., 2006), who tracked the fluorescently tagged proteins and concluded that the fluctuations varied slowly in time and that genes of the same pathway showed enhanced temporal correlations. Therefore such time-lapse measurements in single cells resulting in time-series of species could be of immense help in estimating the presence and nature of the regulatory activity between corresponding genes. The analytical framework

presented above would then be an ideal 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, the changes in physical properties of the protein molecules tagged with with heavier fluorescent particles would make learning parameter values from such data difficult. 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. 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. Tagged transcripts as utilized in (Raj et al., 2008) or tagged proteins in (Yu et al., 2006) and (Cai et al., 2006) are making single cell, single molecule observations possible, which makes the analytical techniques presented in this paper appropriate and useful.

Appendix A. Elementary Activator

The reaction set of an elementary activator $X \rightarrow Y$ is given in Table A.1. The regulatory sequence upstream of gene G_y is bound by the transcription factor P_x, which is an activator in the present case, resulting in complex C_y which is in effect the active or ON state of the gene. Transcription of M_y is initiated from this complex. The deterministic dynamics of this system are governed by a set of

$$\phi \xrightarrow{k_{M_x}^+} M_x \qquad\qquad 8.14 \times 10^{-3} \; nM \; min^{-1}$$

$$M_x \xrightarrow{k_{M_x}^-} \phi \qquad\qquad 0.0407 \; min^{-1}$$

$$M_x \xrightarrow{k_{P_x}^+} M_x + P_x \qquad 36.3 \; min^{-1}$$

$$P_x \xrightarrow{k_{P_x}^-} \phi \qquad\qquad 0.0184 \; min^{-1}$$

$$P_x + G_y \underset{k_{off}}{\overset{k_{on}}{\rightleftharpoons}} C_y \qquad 1 \; nM^{-1} \; min^{-1}, 200 \; min^{-1}$$

$$C_y \xrightarrow{k_{M_y}^+} C_y + M_y \qquad 0.52 \; min^{-1}$$

$$M_y \xrightarrow{k_{M_y}^-} \phi \qquad\qquad 0.077 \; min^{-1}$$

$$M_y \xrightarrow{k_{P_y}^+} M_y + P_y \qquad 143.2 \; min^{-1}$$

$$P_y \xrightarrow{k_{P_y}^-} \phi \qquad\qquad 0.0176 \; min^{-1}$$

Table A.1: Reaction set describing the process of activation between 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. nM stands for nano-molar concentration and *min* is minutes.

coupled ODEs that describe the time-evolution of the mean concentration levels of the species. The total amount of G_y present in the cell at time t is $C_y(t) + G_y(t) = G_y(t_0)$. Hence, $C_y(t) = \alpha - G_y(t)$ 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 $1 \; \mu m^3$, $G_y \approx 1 \; nM$. Therefore, at steady state, $\langle C_y \rangle = \frac{\alpha \langle P_x \rangle}{K_D + \langle P_x \rangle}$ and $\langle G_y \rangle = \frac{\alpha K_D}{K_D + \langle P_x \rangle}$. The redundancy in the original set of ODEs is eliminated by substituting for C_y.

The rate equations are,

$$\frac{dG_y}{dt} = -k_{on}P_xG_y + k_{off}(\alpha - G_y) \tag{A.1}$$

$$\frac{dM_x}{dt} = k_{M_x}^+ - k_{M_x}^- M_x \tag{A.2}$$

$$\frac{dM_y}{dt} = k_{M_y}^+(\alpha - G_y) - k_{M_y}^- M_y \tag{A.3}$$

$$\frac{dP_x}{dt} = k_{M_x}^+ M_x - k_{P_x}^- P_x - k_{on}P_xG_y + k_{off}(\alpha - G_y) \tag{A.4}$$

$$\frac{dP_y}{dt} = k_{P_y}^+ M_y - k_{P_y}^- P_y \tag{A.5}$$

From the above, $\langle M_y \rangle = \frac{k_{M_y}^+}{k_{M_y}^-}\frac{\alpha\langle P_x\rangle}{K_D+\langle P_x\rangle}$, which follows Michaelis-Menten (hyperbolic) type kinetics. $k_{M_y}^+$ is the transcription rate. Note that k_{basal} is excluded in the model as it does not represent the regulatory link between the two genes. The quantities required to solve for the dynamic covariances are the deterministic rates $\mathbf{R}$ and the stoichiometry ν which are obtained from the above rate equations.

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

$$k_{M_x}^- M_x, k_{M_y}^- M_y, k_{P_x}^- P_x, k_{P_y}^- P_y\Big)^T$$

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

The relation between $\mathbf{A}$, and the above quantities is given in equation (6), where $\mathbf{x}$ is the vector of variables $[G_y, M_x, M_y, P_x, P_y]$.

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

It now becomes essential to work within the range of the parameter values that is biologically plausible. For this purpose we adopt the {CHA4 → CHA1} regulatory link (Bornaes and et al., 1993). Information on TFs regulating genes in yeast is available through projects such as *YEASTRACT*[1] (Monteiro and et al., 2008). Cha4p protein is a TF activating the regulation of CHA1. In our model, CHA4 is the regulator gene X and CHA1 is the regulated gene Y, whose mean steady state protein and mRNA levels, and their respective decay rates 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 A.2. 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:

$$k_{(M,P)}^{-} = \frac{ln(2)}{(t_{1/2})_{(M,P)}} + \frac{ln(2)}{t_{double}} \tag{A.6}$$

The values of the other parameters, as given in Table A.1, are derived on solving for the deterministic rate equations at steady-state conditions. For example, at steady-state $k_{M_x}^{+} = k_{M_x}^{-} \langle M_x \rangle = 0.0407 \times 0.2 = 8.14 \times 10^{-3} \ nM \ min^{-1}$. The OFF and ON rates of the DNA are chosen such that their ratio $K_D = 200 \ nM$.

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 A.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.

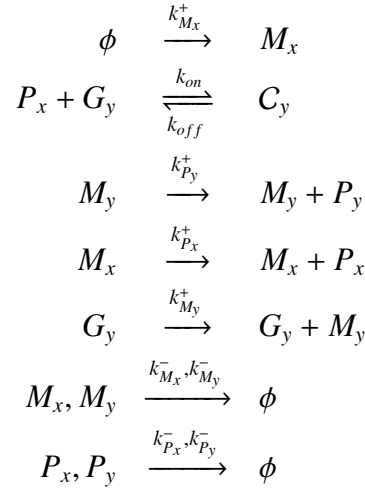## Appendix B. Reaction schemes and rate equations of other regulatory mechanisms

Appendix B.1. Elementary Repressor

The reaction scheme for this system is exactly the same as for the case of the two-gene *elementary activation* $X \rightarrow Y$, except that the trancription of G_y is now through a basal transcription rate of $k_{M_y}^{+}$. Gene G_y is switched to inactive state C_y

---

[1]http://www.yeastract.com

by the action of repressor P_x. M_y is transcribed when the promoter region is free of this repressor.

$$\phi \xrightarrow{k^+_{M_x}} M_x$$

$$P_x + G_y \underset{k_{off}}{\overset{k_{on}}{\rightleftharpoons}} C_y$$

$$M_y \xrightarrow{k^+_{P_y}} M_y + P_y$$

$$M_x \xrightarrow{k^+_{P_x}} M_x + P_x$$

$$G_y \xrightarrow{k^+_{M_y}} G_y + M_y$$

$$M_x, M_y \xrightarrow{k^-_{M_x}, k^-_{M_y}} \phi$$

$$P_x, P_y \xrightarrow{k^-_{P_x}, k^-_{P_y}} \phi$$

The deterministic rate equations describing the evolution of average concentration levels of species are:

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

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

$$\frac{dM_y}{dt} = k^+_{M_y}G_y - k^-_{M_y}M_y$$

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

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

The Jacobian matrix, the stoichiometric matrix and the vector of deterministic rates turn out to be same as in the case of an elementary activator system, with the main exception being the change in the sign of the transcription rates in the Jacobian matrix. This is denoted by an underbrace. This sign change in the off-diagonal element of **A** is what causes the correlations to go negative, due to the

corresponding change in the signs of the eigenvectors.

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

$$
\mathbf{v} = \begin{pmatrix}
-1 & +1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & +1 & 0 & 0 & -1 & 0 & 0 & 0 \\
0 & 0 & +1 & 0 & 0 & 0 & 0 & -1 & 0 & 0 \\
-1 & +1 & 0 & 0 & +1 & 0 & 0 & 0 & -1 & 0 \\
0 & 0 & 0 & 0 & 0 & +1 & 0 & 0 & 0 & -1
\end{pmatrix}
$$

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

$$
k_{M_x}^- M_x, k_{M_y}^- M_y, k_{P_x}^- P_x, k_{P_y}^- P_y \Big)^T
$$

Appendix B.2. Activation via Dimers

The reaction scheme is the same as that of elementary activation with the additional reactions of dimerization $2P_x \underset{k_b}{\overset{k_a}{\rightleftharpoons}} P_{x_2}$. Dimers P_{x_2} act as activators in the new system $P_{x_2} + G_y \underset{k_{off}}{\overset{k_{on}}{\rightleftharpoons}} C_y$. The deterministic rate equations are:

$$
\frac{dG_y}{dt} = k_{off}(\alpha - G_y) - k_{on}G_yP_{x_2}
$$

$$
\frac{dM_x}{dt} = k_{M_x}^+ - k_{M_x}^- M_x
$$

$$
\frac{dM_y}{dt} = k_{M_y}^+(\alpha - G_y) - k_{M_x}^- M_x
$$

$$
\frac{dP_x}{dt} = k_{P_x}^+ M_x - k_{P_x}^- P_x + 2k_bP_{x_2} - 2k_aP_x^2
$$

$$
\frac{dP_{x_2}}{dt} = k_{off}(\alpha - G_y) - k_{on}G_yP_{x_2} - k_bP_{x_2} + k_aP_x^2
$$

$$
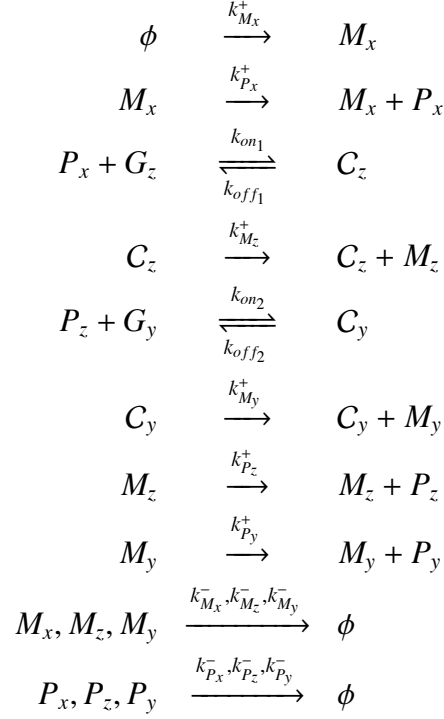\frac{dP_y}{dt} = k_{P_y}^+ M_y - k_{P_y}^- P_y
$$

26

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

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

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

Appendix B.3. Cascading Activation

The set of reactions describing the cascaded activation system, that employs dual activators, is as follows:

$$\phi \xrightarrow{k^+_{M_x}} M_x$$

$$M_x \xrightarrow{k^+_{P_x}} M_x + P_x$$

$$P_x + G_z \underset{k_{off_1}}{\overset{k_{on_1}}{\rightleftharpoons}} C_z$$

$$C_z \xrightarrow{k^+_{M_z}} C_z + M_z$$

$$P_z + G_y \underset{k_{off_2}}{\overset{k_{on_2}}{\rightleftharpoons}} C_y$$

$$C_y \xrightarrow{k^+_{M_y}} C_y + M_y$$

$$M_z \xrightarrow{k^+_{P_z}} M_z + P_z$$

$$M_y \xrightarrow{k^+_{P_y}} M_y + P_y$$

$$M_x, M_z, M_y \xrightarrow{k^-_{M_x}, k^-_{M_z}, k^-_{M_y}} \phi$$

$$P_x, P_z, P_y \xrightarrow{k^-_{P_x}, k^-_{P_z}, k^-_{P_y}} \phi$$

In the following set of rate equations, the redundant variables C_z and C_y are eliminated and consequently the terms α and β denote the initial concentrations $G_z(t_0)$

and $G_y(t_0)$ respectively:

$$\frac{dG_z}{dt} = -k_{on_1}G_zP_x + k_{off_1}(\alpha - G_z)$$

$$\frac{dG_y}{dt} = -k_{on_2}G_yP_z + k_{off_2}(\beta - G_y)$$

$$\frac{dM_x}{dt} = k_{M_x}^+ - k_{M_x}^-M_x$$

$$\frac{dM_z}{dt} = k_{M_z}^+(\alpha - G_z) - k_{M_z}^-M_z$$

$$\frac{dM_y}{dt} = k_{M_y}^+(\beta - G_y) - k_{M_y}^-M_y$$

$$\frac{dP_x}{dt} = k_{P_x}^+M_x - k_{P_x}^-P_x - k_{on_1}G_zP_x + k_{off_1}(\alpha - G_z)$$

$$\frac{dP_z}{dt} = k_{P_z}^+M_z - k_{P_z}^-P_z - k_{on_2}G_yP_z + k_{off_2}(\beta - G_y)$$

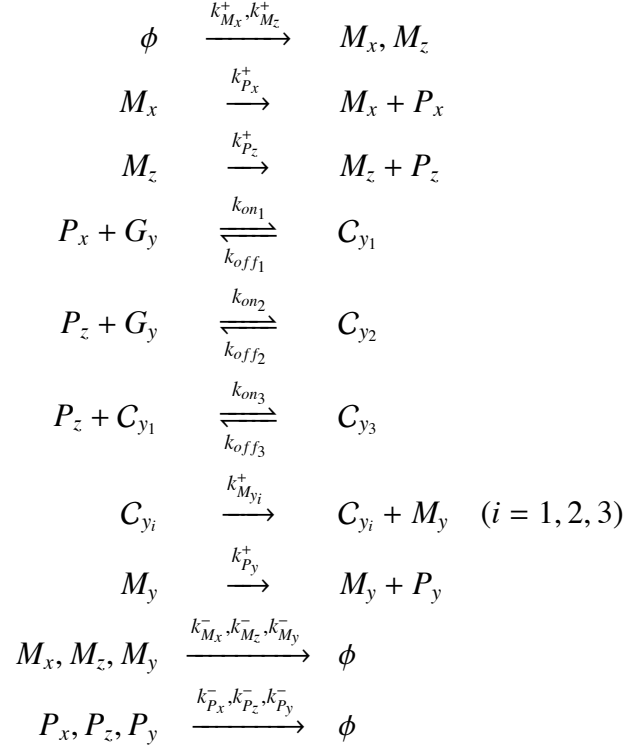$$\frac{dP_y}{dt} = k_{P_y}^+M_y - k_{P_y}^-P_y$$

$$\mathbf{A} = \begin{pmatrix} -(k_{off_1} + k_{on_1}P_x) & 0 & 0 & 0 & 0 & -k_{on_1}G_z & 0 & 0 \\ 0 & -(k_{off_2} + k_{on_2}P_z) & 0 & 0 & 0 & 0 & -k_{on_2}G_y & 0 \\ 0 & 0 & -k_{M_x}^- & 0 & 0 & 0 & 0 & 0 \\ -k_{M_z}^+ & 0 & 0 & -k_{M_z}^- & 0 & 0 & 0 & 0 \\ 0 & -k_{M_y}^+ & 0 & 0 & -k_{M_y}^- & 0 & 0 & 0 \\ -(k_{off_1} + k_{on_1}P_x) & 0 & k_{P_x}^+ & 0 & 0 & -(k_{P_x}^- + k_{on_1}G_z) & 0 & 0 \\ 0 & -(k_{off_2} + k_{on_2}P_z) & 0 & k_{P_z}^+ & 0 & 0 & -(k_{P_z}^- + k_{on_2}G_y) & 0 \\ 0 & 0 & 0 & 0 & k_{P_y}^+ & 0 & 0 & -k_{P_y}^- \end{pmatrix}$$

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

$$\mathbf{R} = [k_{on_1}G_zP_x, k_{off_1}(\alpha - G_z), k_{M_y}^+(\alpha - G_z), k_{on_2}G_yP_z, k_{off_2}(\beta - G_y), k_{M_y}^+(\beta - G_y), k_{M_x}^+,$$
$$k_{P_x}^+M_x, k_{P_z}^+M_z, k_{P_y}^+M_y, k_{M_x}^-M_x, k_{M_z}^-M_z, k_{M_y}^-M_y, k_{P_x}^-P_x, k_{P_z}^-P_z, k_{P_y}^-P_y]^T$$

Appendix B.4. Co-operative Activation

The following is the set of reactions describing the regulatory system, where two activators P_x and P_z act on the regulatory region upstream of G_y, either individually or through combined action:

$$\phi \xrightarrow{k^+_{M_x}, k^+_{M_z}} M_x, M_z$$

$$M_x \xrightarrow{k^+_{P_x}} M_x + P_x$$

$$M_z \xrightarrow{k^+_{P_z}} M_z + P_z$$

$$P_x + G_y \underset{k_{off_1}}{\overset{k_{on_1}}{\rightleftharpoons}} C_{y_1}$$

$$P_z + G_y \underset{k_{off_2}}{\overset{k_{on_2}}{\rightleftharpoons}} C_{y_2}$$

$$P_z + C_{y_1} \underset{k_{off_3}}{\overset{k_{on_3}}{\rightleftharpoons}} C_{y_3}$$

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

$$M_y \xrightarrow{k^+_{P_y}} M_y + P_y$$

$$M_x, M_z, M_y \xrightarrow{k^-_{M_x}, k^-_{M_z}, k^-_{M_y}} \phi$$

$$P_x, P_z, P_y \xrightarrow{k^-_{P_x}, k^-_{P_z}, k^-_{P_y}} \phi$$

Once again in the rate equations, G_y is chosen as the redundant variable, while the TF-DNA complexes C_{y_i} are retained in the model.

$$
\frac{dC_{y_1}}{dt} = k_{on_1} P_x (\alpha - C_{y_1} - C_{y_2} - C_{y_3}) - k_{off_1} C_{y_1} + k_{off_3} C_{y_3} - k_{on_3} C_{y_1} P_z
$$

$$
\frac{dC_{y_2}}{dt} = k_{on_2} P_z (\alpha - C_{y_1} - C_{y_2} - C_{y_3}) - k_{off_2} C_{y_2}
$$

$$
\frac{dC_{y_3}}{dt} = k_{on_3} C_{y_1} P_z - k_{off_3} C_{y_3}
$$

$$
\frac{dM_x}{dt} = k_{M_x}^+ - k_{M_x}^- M_x
$$

$$
\frac{dM_z}{dt} = k_{M_z}^+ - k_{M_z}^- M_z
$$

$$
\frac{dM_y}{dt} = k_{M_{y_1}}^+ C_{y_1} + k_{M_{y_2}}^+ C_{y_2} + k_{M_{y_3}}^+ C_{y_3} - k_{M_y}^- M_y
$$

$$
\frac{dP_x}{dt} = k_{off_1} C_{y_1} - k_{on_1} P_x (\alpha - C_{y_1} - C_{y_2} - C_{y_3}) + k_{P_x}^+ M_x - k_{P_x}^- P_x
$$

$$
\frac{dP_z}{dt} = k_{off_2} C_{y_2} - k_{on_2} P_z (\alpha - C_{y_1} - C_{y_2} - C_{y_3}) + k_{off_3} C_{y_3} - k_{on_3} C_{y_1} P_z + k_{P_z}^+ M_z - k_{P_z}^- P_z
$$

$$
\frac{dP_y}{dt} = k_{P_y}^+ M_y - k_{P_y}^- P_y
$$

As in previous cases, it is quite straightforward to derive the Jacobian matrix $\mathbf{A}$, the vector of deterministic rates $\mathbf{R}$ and the stoichimetric matrix $\mathbf{\nu}$ from the above set of deterministic dynamical equations.
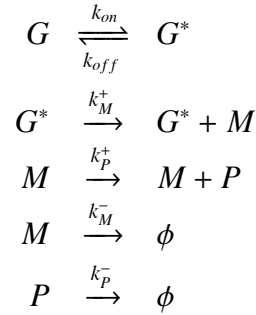
Appendix B.5. Feedforward Loops

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).

Figure B.14: Coherent and Incoherent FeedForward Loops. Gene activation is denoted by an arrowhead and repression by a dashhead.

Appendix C. Expression for τ^* in the case of 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. Here, 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^+. The reactions steps for such a system are:

$$G \underset{k_{off}}{\overset{k_{on}}{\rightleftharpoons}} G^*$$

$$G^* \overset{k_M^+}{\longrightarrow} G^* + M$$

$$M \overset{k_P^+}{\longrightarrow} M + P$$

$$M \overset{k_M^-}{\longrightarrow} \phi$$

$$P \overset{k_P^-}{\longrightarrow} \phi$$

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,

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

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:

$$\frac{d\langle G^*\rangle}{dt} = k_{on}(G(t_0) - \langle G^*\rangle) - k_{off}\langle G^*\rangle$$

$$\frac{d\langle M\rangle}{dt} = k_M^+\langle G^*\rangle - k_M^-\langle M\rangle$$

$$\frac{d\langle P\rangle}{dt} = k_P^+\langle M\rangle - k_P^-\langle P\rangle$$

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:

$$\mathbf{R} = [k_{on}(G(t_0) - \langle G^*\rangle), k_{off}\langle G^*\rangle, k_M^+\langle G^*\rangle, k_P^+\langle M\rangle, k_M^-\langle M\rangle, k_P^-\langle P\rangle]^T$$

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

The diffusion matrix given by $\nu\,\mathbf{diag}(\mathbf{R}(.))\,\nu^T$ is:

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

The next step is to solve for the stationary covariance matrix:

$$\mathbf{C} = \begin{pmatrix} C_{11} & C_{12} & C_{13} \\ C_{21} & C_{22} & C_{23} \\ C_{31} & C_{32} & C_{33} \end{pmatrix}$$

$$= \begin{pmatrix} \text{Var}[G^*(t), G^*(t)] & \text{Cov}[G^*(t), M(t)] & \text{Cov}[G^*(t), P(t)] \\ \text{Cov}[M(t), G^*(t)] & \text{Var}[M(t), M(t)] & \text{Cov}[M(t), P(t)] \\ \text{Cov}[P(t), G^*(t)] & \text{Cov}[P(t), M(t)] & \text{Var}[P(t), P(t)] \end{pmatrix}$$

where Var and Cov represent the stationary auto-covariance and covariances. Each element of the above matrix is evaluated by solving the Lyapunov equation systematically element-wise. For example, the first element of $\mathbf{AC} + \mathbf{CA}^T + \mathbf{BB}^T = \mathbf{0}$

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^*,

$$\text{Var}[G^*(t), G^*(t)] \quad = \quad \frac{(k_{on} - k_{off})\langle G^*\rangle + G(t_0)k_{off}}{2(k_{on} + k_{off})} \quad = \quad \langle G^*\rangle\Big[\frac{k_{off}}{k_{on} + k_{off}}\Big]$$

Continuing on similar lines, the covariances between all the variables are evaluated. Whiled doing so, we make approximations that the decay rates are much smaller when compared to the ON and OFF rates of the gene $(k_{on} + k_{off}) \gg k^-_{(M,P)}$ and that $k_{off} > k_{ON}$. Both these approximations are valid for biologically plausible values of rate constants (Bundschuh et al., 2003).

$$\text{Cov}[G^*(t), M(t)] \quad = \quad \langle M\rangle\frac{k^-_M k_{off}}{(k_{on} + k_{off} + k^-_M)(k_{on} + k_{off})} \approx \langle M\rangle\frac{k^-_M k_{off}}{(k_{on} + k_{off})^2}$$

$$\approx \quad \langle M\rangle\frac{k^-_M}{k_{off}}$$

$$\text{Var}[M(t), M(t)] \quad = \quad \langle M\rangle + \langle M\rangle\Big[\frac{k^+_M k_{off}}{(k_{on} + k_{off} + k^-_M)(k_{on} + k_{off})}\Big] \approx \langle M\rangle + \langle M\rangle\Big[\frac{k^+_M}{k_{off}}\Big]$$

$$\approx \quad \langle M\rangle$$

$$\text{Cov}[M(t), P(t)] \quad = \quad \langle M\rangle\frac{k^+_P}{k^-_M + k^-_P}\Big[1 + \frac{k^+_M k_{off}}{(k_{on} + k_{off})(k_{on} + k_{off} + k^-_M)}\Big(1 + \frac{k^-_M}{(k_{on} + k_{off} + k^-_P)}\Big)\Big]$$

$$\approx \quad \langle M\rangle\frac{k^+_P}{k^-_M + k^-_P}\Big[1 + \frac{k^+_M k_{off}}{(k_{on} + k_{off})^2}\Big(1 + \frac{k^-_M}{(k_{on} + k_{off})}\Big)\Big]$$

$$\approx \quad \langle M\rangle\frac{k^+_P}{k^-_M + k^-_P}\Big[1 + \frac{k^+_M}{k_{off}}\Big(1 + \frac{k^-_M}{k_{off}}\Big)\Big] \approx \langle M\rangle\frac{k^+_P}{k^-_M + k^-_P}\Big[1 + \frac{k^+_M}{k_{off}}\Big]$$

$$\approx \quad \langle M\rangle\frac{k^+_P}{k^-_M + k^-_P}$$

$$\text{Var}[P(t), P(t)] \quad = \quad \langle P\rangle + \frac{k^+_P}{k^-_P}\text{Cov}[M(t), P(t)]$$

$$\approx \quad \langle P\rangle + \langle P\rangle\frac{k^+_P}{k^-_M + k^-_P}$$

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 from 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, 2004) 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:

$$
\mathbf{V} \;=\; \begin{pmatrix} 0 & \dfrac{(k_{on}+k_{off}-k^-_M)(k_{on}+k_{off}-k^-_P)}{k^+_M k^+_P} & 0 \\[2ex] \dfrac{-(k^-_M-k^-_P)}{k^+_P} & \dfrac{-(k_{on}+k_{off}-k^-_P)}{k^+_P} & 0 \\[2ex] 1 & 1 & 1 \end{pmatrix}
$$

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:

$$
\mathbf{U}^T \;=\; \begin{pmatrix} \dfrac{-k^+_M k^+_P}{(k_{on}+k_{off}-k^-_M)(k^-_M-k^-_P)} & \dfrac{-k^+_P}{(k^-_M-k^-_P)} & 0 \\[2ex] \dfrac{k^+_M k^+_P}{(k_{on}+k_{off}-k^-_M)(k_{on}+k_{off}-k^-_P)} & 0 & 0 \\[2ex] \dfrac{k^+_M k^+_P}{(k_{on}+k_{off}-k^-_P)(k^-_M-k^-_P)} & \dfrac{k^+_P}{(k^-_M-k^-_P)} & 1 \end{pmatrix}
$$

The corresponding eigenvalues are:

$$
\lambda = [-k^-_M, \; -(k_{on} + k_{off}), \; -k^-_P]
$$

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

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

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,

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

On substituting for the **C** terms and doing approximations,

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

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, differentiating the above covariance term partially *w.r.t* τ and equating to 0, yields the $\tau = \tau^*$ at which the covariance reaches peak magnitude.

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

36

Since this time delay τ is shown to be a function of the decay rates, 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. For a perturbation of ΔM in the mRNAs, the response in the proteins is,

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

This response in the protein level also has the characteristic peak after a time-delay of (t_{resp}), which is evaluated by equating the time-evolution of $\langle P \rangle$ to zero,

$$\frac{\partial \langle P(t) \rangle}{\partial t} \;=\; 0$$

$$k_M^- e^{-k_M^- t_{resp}} - k_P^- e^{-k_P^- t_{resp}} \;=\; 0$$

$$t_{resp} \;=\; \frac{\ln\left[\frac{k_M^-}{k_P^-}\right]}{k_M^- - k_P^-}$$

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,

$$\lim_{x \to 1} t_{resp} \;=\; \lim_{x \to 1} \frac{1}{k_P^-}\frac{\ln(x)}{(x-1)}$$

$$=\; \lim_{x \to 1} \frac{1}{k_P^-}\frac{1/x}{1}$$

$$=\; \frac{1}{k_P^-}$$

Alon, U., Jun 2007. Network motifs: theory and experimental approaches. Nature Reviews Genetics 8, 450–461.

Arava, Y., Wang, Y., Storey, J. D., Liu, C. L., Brown, P. O., Herschlag, D., 2003. Genome-wide analysis of mrna translation profiles in saccharomycescerevisiae. PNAS 100 (7), 3889–3894.

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

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

Belle, A., Tanay, A., Bitincka, L., Shamir, R., O'Shea, E. K., 2006. Quantification of protein half-lives in the budding yeast proteome. PNAS 103 (35), 13004–13009.

Blake, W. J., Balázsi, G., Kohanski, M. A., Isaacs, F. J., Murphy, K. F., Kuang, Y., Cantor, C. R., Walt, D. R., Collins, J. J., December 2006. Phenotypic consequences of promoter-mediated transcriptional noise. Molecular Cell 24 (6), 853–865.

Bornaes, C., et al., M. W. I., 1993. 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.

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

Cai, L., Friedman, N., Xie, X. S., March 2006. Stochastic protein expression in individual cells at the single molecule level. Nature 440, 358–362.

Cox, C. D., McCollum, J. M., Allen, M. S., Dar, R. D., Simpson, M. L., Aug 2008. Using noise to probe and characterize gene circuits. PNAS 105 (31), 10809–10814.

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

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

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

Ghaemmaghami, S., et al., W.-K. H., 2003. Global analysis of protein expression in yeast. Nature 425, 737–741.

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

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

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

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

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

Lei, J., February 2009. Stochasticity in single gene expression with both intrinsic noise and fluctuation in kinetic parameters. J Theor Biol 256 (4), 485–92.

Leloup, J., Goldbeter, A., 2004. Modeling the mammalian circadian clock: sensitivity analysis and multiplicity of oscillatory mechanisms. J Theor Biol. 230 (4), 541–62.

Maamar, H., Raj, A., Dubnau, D., Jul 2007. Noise in gene expression determines cell fate in bacillus subtilis. Science 317, 526–529.

Mangan, S., Itzkovitz, S., Zaslaver, A., Alon, U., 2006. The incoherent feed-forward loop accelerates the response-time of the gal system of *Escherichia coli*. Journal of Molecular Biology 356 (5), 1073 – 1081.

Monteiro, P. T., et al., N. D. M., 2008. YEASTRACT-DISCOVERER: new tools to improve the analysis of transcriptional regulatory associations in Saccharomyces cerevisiae. Nucl. Acids Res. 36, D132–136.

Nachman, I., Regev, A., Ramanathan, S., November 2007. Dissecting timing variability in yeast meiosis. Cell 131 (3), 544–556.

Onsager, L., Feb 1931. Reciprocal relations in irreversible processes. i. Phys. Rev. 37 (4), 405–426.

Ozbudak, E. M., Thattai, M., Kurtser, I., Grossman, A. D., van Oudenaarden, A., May 2002. Regulation of noise in the expression of a single gene. Nature Genetics 31, 69–73.

Paulsson, J., 2004. Summing up the noise in gene networks. Nature 427 (415-18).

Pedraza, J. M., van Oudenaarden, A., 2005. Noise propagation in gene networks. Science 307, 1965–1969.

Raj, A., van den Bogaard, P., Rifkin, S. A., van Oudenaarden, A., , Tyagi, S., 2008. Imaging individual mrna molecules using multiple singly labeled probes. Nature Methods 9, 877–879.

Raj, A., van Oudenaarden, A., October 2008. Nature, nurture, or chance: Stochastic gene expression and its consequences. Cell Volume 135 (2), 216–226.

Raser, J. M., O'Shea, E. K., 2004. Control of stochasticity in eukaryotic gene expression. Science 304 (5678), 1811–1814.

Raser, J. M., O'Shea, E. K., September 2005. Noise in gene expression: Origins, consequences, and control. Science 309, 2010–2013.

Shahrezaei, V., Swain, P. S., November 2008. Analytical distributions for stochastic gene expression. PNAS 105 (45), 17256–17261.

Shen-Orr, S. S., Milo, R., Mangan, S., Alon, U., May 2002. Network motifs in the transcriptional regulation network of *Escherichia coli*. Nature Genetics 31, 64–68.

Sigal, A., Milo, R., Cohen, A., Geva-Zatorsky, N., Klein, Y., Liron, Y., Rosenfeld, N., Danon, T., Perzov, N., Alon, U., November 2006. Variability and memory of protein levels in human cells. Nature 444, 643–646.

Speck, T., Seifert, U., 2006. Restoring a fluctuation-dissipation theorem in a nonequilibrium steady state. Europhys. Lett. 74 (3), 391–396.

Tao, Y., Zheng, X., Sun, Y., 2007. Effect of feedback regulation on stochastic gene expression. Journal of Theoretical Biology 247, 827–836.

Thattai, M., van Oudenaarden, A., July 2001. Intrinsic noise in gene regulatory networks. PNAS 98 (15), 8614–8619.

Tomioka, R., Kimura, H., Kobayashi, T. J., Aihara, K., 2004. Multivariate analysis of noise in genetic regulatory networks. Journal of Theoretical Biology 229, 501–521.

van Kampen, N. G., 2007. Stochastic Processes in Physics and Chemistry, 3rd Edition. Elsevier, North-Holland Personal Library.

Wang, Y., Liu, C. L., Storey, J. D., Tibshirani, R. J., Herschlag, D., Brown, P. O., 2002. Precision and functional specificity in mrna decay. PNAS 99 (9), 5860–5865.

Wernet, M. F., Mazzoni, E. O., Celik, A., Duncan, D. M., Duncan, I., Desplan, C., March 2006. Stochastic spineless expression creates the retinal mosaic for colour vision. Nature 440, 174–180.

Yu, J., Xiao, J., Ren, X., Lao, K., Xie, S. X., 2006. Probing Gene Expression in Live Cells, One Protein Molecule at a Time. Science 311 (5767), 1600–1603.