Noise Characteristics of Feed Forward Loops

Bhaswar Ghosh, Rajesh Karmakar and Indrani Bose*

9th February 2008

Department of Physics
Bose Institute
93/1, A. P. C. Road
Kolkata - 700 009, India

*Author to be contacted for correspondence; e-mail: indrani@bosemain.boseinst.ac.in

Abstract

A prominent feature of gene transcription regulatory networks is the presence in large numbers of motifs, i.e, patterns of interconnection, in the networks. One such motif is the feed forward loop (FFL) consisting of three genes X, Y and Z. The protein product of x of X controls the synthesis of protein product y of Y. Proteins x and y jointly regulate the synthesis of z proteins from the gene Z. The FFLs, depending on the nature of the regulating interactions, can be of eight different types which can again be classified into two categories: coherent and incoherent. In this paper, we study the noise characteristics of FFLs using the Langevin formalism and the Monte Carlo simulation technique based on the Gillespie algorithm. We calculate the variances around the mean protein levels in the steady states of the FFLs and find that, in the case of coherent FFLs, the most abundant FFL, namely, the Type-1 coherent FFL, is the least noisy. This is however not so in the case of incoherent FFLs. The results suggest possible relationships between noise, functionality and abundance.

Keywords: feed forward loop, stochastic gene expression, noise, gene transcription regulatory network, Langevin formalism, Gillespie algorithm.

1. Introduction

Biological networks represent the complex webs of biomolecular interactions and reactions underlying cellular processes. Well-known examples of biological networks include metabolic reaction, protein-protein interaction and gene transcription regulatory networks (GTRNs) [1, 2]. The availability of large scale experimental data and powerful computational tools provide information on the structural and functional features of the complex

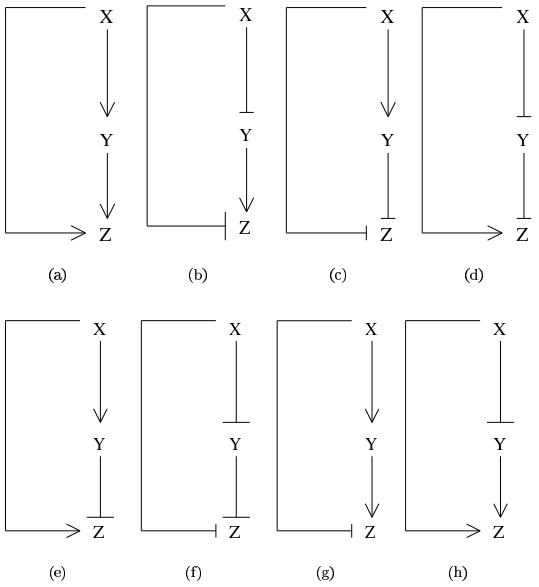

Figure 1: Eight types of FFLs: (a) Type-1, (b) Type-2, (c) Type-3, (d) Type-4 coherent FFLs, (e) Type-1, (f) Type-2, (g) Type-3, (h)Type-4 incoherent FFLs. The arrow sign denotes activation and the $\perp$ sign repression.

networks. In the case of a GTRN, the nodes of the network represent genes and two nodes are connected by a directed link if the protein product of one gene regulates the synthesis of proteins from the other gene. Existing databases on simple organisms like *E. coli* and *S. cerevisiae* show that the GTRNs of these organisms have common structural motifs like bi-fan, single input module (SIM) and feed forward loop (FFL) [3, 4, 5]. Such motifs are more abundant in the naturally occurring networks than in their randomized counterparts, highlighting the essential roles of motifs in network function.

The regulatory and other biochemical processes associated with a GTRN are probabilistic in nature giving rise to fluctuations in the levels of proteins synthesized by different genes. The magnitude of noise cannot be neglected when the number of biomolecules participating in the network processes is small. Recently, several theoretical [6, 7, 8, 9, 10] as well as experimental [11, 12, 13] studies have been carried out on the origins and consequences of stochasticity and the dependence of noise on some important parameters of gene expression (GE) like the transcription and translation rates. The effect of stochasticity may be both advantageous and disadvantageous. Stochasticity can give rise to phenotypic variations in an identical population of cells kept in the same environment. It thus plays a positive role in situations where phenotypic diversity is beneficial. In most cases, however, stochasticity acts to diminish fidelity in cellular processes. Noisy regulatory signals, for example, may not achieve the desired outcome introducing uncertainty in cellular behaviour.

Fraser et al. [14] have recently addressed the important issue of the relation of noise to the fitness of an organism. They estimate the noise in protein production for almost all the genes in *S. cerevisiae* and show that the amount of noise associated with protein

levels in the steady state has lower magnitude in the cases of essential genes and genes encoding subunits of multi-protein complexes. Fluctuations in the protein levels of these functionally important classes of genes are particularly detrimental to organismal fitness because of reduced functionality. The lower amounts of noise associated with the genes support the hypothesis that noise is an evolvable trait acted on by natural selection. In this paper, we consider a simple stochastic model of GE to determine the noise characteristics of a particular type of motif appearing in GTRNs, namely, the FFL [3, 4, 5]. A FFL is a three-node motif describing three genes X, Y and Z (figure 1). The protein x produced from gene X regulates protein synthesis from gene Y. Proteins x and y also jointly regulate the expression of gene Z. Inducer molecules S_x and S_y are in general required to activate or inhibit the function of protein molecules x and y. There are three transcriptional regulatory interactions in a FFL, each of which can have either positive (activation) or negative (repression) sign. The motif with three links can be in eight possible configurations which fall into two categories: coherent and incoherent (figure 1). In a coherent FFL, the sign of the direct regulation path from X to Z is the same as the overall sign of the indirect regulation path via Y. There are four such configurations. In the other four configurations, termed incoherent FFLs, the signs of the direct and indirect regulation paths are opposite. The two protein inputs x and y regulate the target gene Z through either an AND-gate or an OR-gate. In the first case, both x and y proteins are needed to regulate gene Z and in the second case, either x or y protein is sufficient for the regulation of Z. The functionality of the different types of FFLs has been determined using a simple mathematical analysis based on the deterministic rate equation approach [4]. The coherent FFL is found to serve as a sign-sensitive delay element. Consider the Type-1 coherent FFL with AND-gate regulation and a step-like pulse of x proteins as the input stimulus (signal). Expression of gene Z can only begin when the level of y proteins is sufficient to cross the activation threshold for Z. The response time is a measure of the speed of response and is given by the time taken for the z proteins to reach an amount which is half the steady state level. Sign-sensitive delay implies that the response time to step-like stimuli is asymmetric, i.e, the response time is delayed in one direction (pulse OFF to ON) and rapid in the other direction (ON to OFF). As a result, if the activation of the X gene is transient, the Z gene cannot be significantly activated, i.e, the input signal is not transduced through the FFL. The z proteins are synthesized only when the X gene is activated for a sufficiently long time interval. The Z gene switches off rapidly once the X gene is deactivated. In other words, the coherent FFL functions as a persistence detector, responding only to a persistent stimulus and filtering out fluctuations in the input signal. The role of the coherent FFL as sign-sensitive delay has been verified experimentally [15]. The incoherent FFLs function as sign-sensitive accelerators speeding up the response time in one direction (OFF to ON in the stimulus step) but not in the other direction (ON to OFF). Some incoherent FFLs act also as pulse generators. Amongst the coherent FFLs, the Type-1 FFL appears the maximum number of times in the GTRNs of *E. coli* and *S. cerevisiae.* Similarly, in the case of incoherent FFLs, the Type-1 FFL is the most abundant. We calculate the noise characteristics of the coherent and incoherent FFLs using the Langevin formalism [16] and the Monte Carlo simulation technique based on the Gillespie algorithm (GA) [17, 18].

We show that the most abundant coherent FFL, namely, the Type-1 FFL, is the least noisy. This is, however, not true in the case of the incoherent FFLs. The lower number of FFLs has been ascribed to their reduced functionality [4]. Noise is disadvantageous if it affects operational reliability. Our results on noise characteristics of FFLs suggest that noisy motifs are likely to be selected against during evolution if noise is detrimental to the function of the motifs.

2. Stochastic Model of GE

The simple stochastic model of GE has been studied earlier as a Markovian model for the gene induction process [19] and also to explore the possible origins of the genetic disorder, haploinsufficiency [10, 20]. In the minimal model, a gene can be in two possible states: inactive (G) and active (G^*). Due to stochasticity, the gene makes random transitions between the inactive and active states with k_a and k_d being the activation and deactivation rate constants. In the active state, protein production occurs with the rate constant β_p. Protein decay occurs with the rate constant γ_p. The protein decay rate has two components, one, the degradation rate and the other, the dilution rate of proteins due to cell growth and division. The reaction scheme RS-1 is shown in equation (1),

$$G \underset{k_d}{\overset{k_a}{\rightleftharpoons}} G^\star \xrightarrow{\beta_p} p \xrightarrow{\gamma_p} \Phi \tag{1}$$

Let $P(n_1, n_2, t)$ be the probability that at time t, n_1 genes are in the active state G^* and the number of protein molecules is n_2. The rate of change of the probability with respect to time is given by the Master Equation

$$\begin{aligned} \frac{\partial P(n_1,n_2,t)}{\partial t} = \; & k_a[(n_{tot}-n_1+1)P(n_1-1,n_2,t)-(n_{tot}-n_1)P(n_1,n_2,t)] \\ & +k_d[(n_1+1)P(n_1+1,n_2,t)-n_1P(n_1,n_2,t)] \\ & +\beta_p[n_1P(n_1,n_2-1,t)-n_1P(n_1,n_2,t)] \\ & +\gamma_p[(n_2+1)P(n_1,n_2+1,t)-n_2P(n_1,n_2,t)] \end{aligned} \tag{2}$$

where n_{tot} is the total number of genes.

For each rate constant, the gain term adds to the probability and the loss term subtracts from the same. The simplicity of the stochastic model enables one to calculate the mean protein level $< n_2 >$ and its variance $< \delta n_2^2 >=< n_2^2 > - < n_2 >^2$ in the steady state using the standard generating function approach. The results are:

$$< n_2 >= \frac{\beta_p}{\gamma_p} \frac{n_{tot}\, k_a}{k_a + k_d} \tag{3}$$

$$< \delta n_2^2 >=< n_2 > [1 + \frac{\beta_p\, k_d}{(k_a + k_d)(k_a + k_d + \gamma_p)}] \tag{4}$$

Also, the mean number of genes in the active state is given by

$$< n_1 >= \frac{n_{tot}\, k_a}{k_a + k_d} \tag{5}$$

The minimal model (equation (1)) describes constitutive GE. We now assume that the transition from the state G to the state G^* is brought about by activating regulatory molecules S. The reaction scheme RS-2 in the presence of such molecules is given by

$$\begin{array}{ccccccccc} & & k_1 & & k_a & & \beta_p & & \gamma_p & \\ G + S & \rightleftharpoons & & G\_S & \rightleftharpoons & G^\star & \longrightarrow & p & \longrightarrow & \Phi \\ & & k_2 & & k_d & & & & & \end{array} \tag{6}$$

where G_S represents the bound complex of G and S from which transition to the active state G^* occurs. The total number of genes n_{tot} is given by

$$n_{tot} = g + g_s + g^* \tag{7}$$

where g, g_s and g^* are the number of genes in the states G, G_S and G^* respectively. In the steady state, $\frac{dg}{dt} = 0$ and $\frac{dg^*}{dt} = 0$. From the first condition, one obtains

$$\frac{g\, s}{K_1} = g_s \tag{8}$$

where $K_1 = \frac{k_2}{k_1}$ is the equilibrium dissociation constant and s is the number of regulatory molecules. From the second condition, the expression for g^* in the steady state is given by

$$g^* = \frac{n_{tot} k_a \frac{s/K_1}{1+s/K_1}}{k_a \frac{s/K_1}{1+s/K_1} + k_d} \tag{9}$$

Expressions (5) and (9) for the number of genes in the active state G^* are equivalent on defining effective activation and deactivation rate constants

$$k_a^{'} = k_a \frac{s/K_1}{1 + s/K_1} \qquad k_d^{'} = k_d \tag{10}$$

The equivalence relations are useful as one can map the reaction scheme RS-2 onto the simpler scheme RS-1 while calculating mean protein levels and the associated variances. Regulatory molecules, in general, oligomerise to form an active complex S_n where n is the number of regulatory molecules contained in the complex. In this case, the effective rate constants $k_a^{'}$ and $k_d^{'}$ are given by

$$k_a^{'} = k_a \frac{(s/K)^n}{1 + (s/K)^n}, \qquad k_d^{'} = k_d \tag{11}$$

where $K^n = K_1 K_c$, K_c being the equilibrium dissociation constant for oligomerisation, i. e,

$$n\,S \quad \overset{K_c}{\rightleftharpoons} \quad S_n \tag{12}$$

When the regulatory molecules S act as repressors, the effective rate constants are given by

$$k_a^{'} = k_a \frac{1}{1+(s/K)^n}, \qquad k_d^{'} = k_d \tag{13}$$

In this case, repressor molecules on binding to genes prevent their activation to the state G^*.

We now apply the stochastic model of GE to determine the mean levels of proteins x, y and z and the variances thereof in the steady state of a FFL. The variances calculated are a measure of the intrinsic noise associated with GE as fluctuations in the number of regulatory molecules are ignored. Let β_i and γ_i $(i = x,\ y,\ z)$ be the rate constants for the synthesis and decay respectively of protein i. For proteins x, the mean protein level x_{av} and its variance $< \delta x^2 >$ in the steady state are obtained from equations (3) and (4) [10, 19] as (with $n_{tot} = 1$)

$$x_{av} =< x >= \frac{\beta_x}{\gamma_x}\ \frac{k_a}{k_a + k_d} \tag{14}$$

$$< \delta x^2 >=< x >\ \ [1 + \frac{\beta_x\, k_d}{(k_a + k_d)(k_a + k_d + \gamma_x)}] \tag{15}$$

where k_a and k_d are activation and deactivation rate constants of gene X. Protein molecules x regulate the activation of gene Y according to the reaction scheme RS-2. Mapping onto the simpler reaction scheme RS-1, one obtains in the steady state

$$y_{av} =< y >= \frac{\beta_y}{\gamma_y}\ \frac{k_a^{'}}{k_a^{'} + k_d^{'}} \tag{16}$$

$$< \delta y^2 >=< y >\ \ [1 + \frac{\beta_y\, k_d^{'}}{(k_a^{'} + k_d^{'})(k_a^{'} + k_d^{'} + \gamma_y)}] \tag{17}$$

The effective rate constants $k_a^{'}$ and $k_d^{'}$ have the forms given in equations (11) or (13) depending on whether the regulatory interaction is activating or repressing in nature. In the case of activation, assuming n to be 2,

$$k_a^{'} = k_{ay}\ \frac{(x/K_{xy})^2}{1+(x/K_{xy})^2}, \qquad k_d^{'} = k_{dy} \tag{18}$$

In (18), k_{ay} represents the limiting value of $k_a^{'}$ obtained when $\frac{x}{K_{xy}} \gg 1$. In the case of repression,

$$k_a^{'} = k_{ay}\ \frac{1}{1+(x/K_{xy})^2}, \qquad k_d^{'} = k_{dy} \tag{19}$$

Both the x and y proteins regulate the activation of the Z gene. The mapping of the associated reaction scheme onto the simpler reaction scheme RS-1 is still possible. The effective rate constants $k_a^{''}$ and $k_d^{''}$ have specific forms depending on the nature of the regulating interaction (activating/ repressing) and the type of logic gate (AND/ OR) in operation. The mean protein level in the steady state and its variance are

$$z_{av} =< z >= \frac{\beta_z}{\gamma_z} \ \frac{k_a^{''}}{k_a^{''} + k_d^{''}} \tag{20}$$

$$< \delta z^2 >=< z > \ [1 + \frac{\beta_z \ k_d^{''}}{(k_a^{''} + k_d^{''})(k_a^{''} + k_d^{''} + \gamma_z)}] \tag{21}$$

The activation and deactivation rate constants $k_a^{''}$ and $k_d^{''}$ are

$$k_a^{''} = k_{az} \ G(x, y, T_{xz}, T_{yz}), \qquad k_d^{''} = k_{dz} \tag{22}$$

where k_{az} is the limiting value of $k_a^{''}$. For the AND-gate,

$$G(x, y, T_{xz} \ , T_{yz}) = T_{xz} \ T_{yz} \tag{23}$$

For the OR-gate,

$$G(x, y, T_{xz} \ , T_{yz}) = \frac{(1 + (x/K_{xz})^2) \, T_{xz} + (1 + (y/K_{yz})^2) \, T_{yz}}{1 + (x/K_{xz})^2 + (y/K_{yz})^2} \tag{24}$$

This expression has been derived assuming that the regulatory molecules x and y compete to bind at the operator region of the gene Z, as in Ref. [4].
For activating regulatory interactions,

$$T_{xz} = \frac{(x/K_{xz})^2}{1 + (x/K_{xz})^2} \ , \qquad T_{yz} = \frac{(y/K_{yz})^2}{1 + (y/K_{yz})^2} \tag{25}$$

For repressing regulatory interactions,

$$T_{xz} = \frac{1}{1 + (x/K_{xz})^2} \ , \qquad T_{yz} = \frac{1}{1 + (y/K_{yz})^2} \tag{26}$$

The parameters K_{xy}, K_{yz} and K_{xz} appearing in (17), (25) and (26) are analogous to the parameter K in (11). In the steady state of the FFL, all three proteins x, y, z are in their steady state levels and the effective rate constants $k_a^{'}$, $k_a^{''}$ are calculated with the steady state values $x = x_{av}$ and $y = y_{av}$.

The FFL may be considered to be a two-step signaling cascade. The x and z proteins constitute respectively the input and output signals of the cascade. With stochasticity taken into account, it is desirable that cascades are able to transmit signals in a reliable manner. When fluctuations are considerable, there is a danger of the noise building up in successive steps of the cascade corrupting the final output signal. Thattai and Oudenaarden [16] have studied the noise characteristics of signaling cascades and have shown that under certain conditions the fluctuations in the output signal are bounded. Also, noise reduction is possible, i.e, the output signal is less noisy than the input signal.

3. Noise Characteristics of FFL

The variance around the mean protein level has two components: intrinsic and extrinsic. In the last section, the variance due to only the intrinsic part has been calculated. In this section, the fluctuations in the number of regulatory molecules, constituting extrinsic noise, are taken into account. The total variances in the steady state of the FFL are denoted as $< \delta x^2 >_{tot}$ (equals $< \delta x^2 >$ given in (15)), $< \delta y^2 >_{tot}$ and $< \delta z^2 >_{tot}$. The variances can be calculated using the method followed in [16]. We use Langevin equations to take stochasticity into account. The equation describing the production of protein x is given by

$$\dot{x} = \beta_x \, \frac{k_a}{k_a + k_d} - \gamma_x \, x + \eta_1(t) \tag{27}$$

where $\dot{x}$ represents a time derivative. Stochasticity is associated with the time-dependent noise term $\eta_1(t)$ in equation (27). The random variable $\eta_1(t)$ obeys white-noise statistics, i.e,

$$< \eta_1(t) >= 0, \qquad < \eta_1(t) \, \eta_1(t+\tau) >= q_1 \, \delta(\tau) \tag{28}$$

where $\delta(\tau)$ is the Dirac delta function and $< ... >$ denotes an ensemble average. The state dependences of $\eta_1(x,t)$ and $q(x)$ are ignored since we are interested in the steady state noise characteristics. In the absence of the noise term in equation (27), the mean protein level in the steady state ($\dot{x}(t) = 0$), as in equation (14), is recovered. We linearize equation (27) for fluctuations, assumed to be small, about the steady state to obtain

$$\delta\dot{x}(t) + \gamma_x \, \delta x = \eta_1(t) \tag{29}$$

Fourier transform of equation (29) yields

$$(i\,\omega + \gamma_x) \; \delta x(\omega) = \eta_1(\omega) \tag{30}$$

Next, taking ensemble average and applying condition (28), we get

$$< |\delta x(\omega)|^2 >= \frac{q_1}{\omega^2 + \gamma_x^2} \tag{31}$$

The steady state variance $< \delta x^2 >_{tot}$ is given by an inverse Fourier transform at $\tau = 0$, i.e,

$$< \delta x^2 >_{tot} = \frac{q_1}{2\,\gamma_x} \tag{32}$$

Since $< \delta x^2 >_{tot} = < \delta x^2 >$ (equation (15)), q_1 is known explicitly from equation (32). For protein y, the Langevin equation is given by

$$\dot{y} + \gamma_y \, y = \beta_y \, f_{xy}(x) + \eta_2(t) \tag{33}$$

with

$$< \eta_2(t) \, \eta_2(t+\tau) >= q_2 \, \delta(\tau) \tag{34}$$

In equation (33), the rate of creation of y proteins in terms of the x proteins is given by the first term on the r.h.s. The function $f_{xy}(x)$ is designated as the transfer function and is given by

$$f_{xy}(x) = \frac{k_a^{'}}{k_a^{'} + k_d^{'}} \tag{35}$$

where $k_a^{'}$ and $k_d^{'}$ are as defined in equations (18) and (19). Again, the mean protein level in the steady state, y_{av} (equation (16)) can be recovered from equation (33) by ignoring the noise term and putting $\dot{y} = 0$. Going through the same steps as before, the variance $< \delta y^2 >_{tot}$ is obtained as

$$< \delta y^2 >_{tot} = \frac{q_2}{2\,\gamma_y} + \frac{\beta_y^2\, c_x^2\, q_1}{2\,\gamma_x\,\gamma_y\,(\gamma_x + \gamma_y)} \tag{36}$$

The first term in equation (36) is the intrinsic noise term given by $< \delta y^2 >$ (equation (17)). The second term, describing extrinsic noise, arises due to the noise propagated from the input, i.e, due to the fluctuations in the number of x regulatory proteins. In the same equation, c_x is the derivative of the transfer function $f_{xy}(x)$, w.r.t x, calculated at the steady state value of x, i.e

$$c_x = \frac{\partial f_{xy}(x)}{\partial x} \Bigg|\; x = x_{av} \tag{37}$$

For the z proteins, the Langevin equation is

$$\dot{z} + \gamma_z\, z = \beta_z\, g_{xy}(x, y) + \eta_3(t) \tag{38}$$

with

$$< \eta_3(t)\, \eta_3(t + \tau) >= q_3\, \delta(\tau) \tag{39}$$

The transfer function $g_{xy}(x, y)$ is given by

$$g_{xy}(x, y) = \frac{k_a^{''}}{k_a^{''} + k_d^{''}} \tag{40}$$

where $k_a^{''}$ and $k_d^{''}$ have been defined in equations (22)-(26). The variance $< \delta z^2 >_{tot}$ is given by

$$\begin{aligned} < \delta z^2 >_{tot} = \; & \frac{q_3}{2\,\gamma_z} + \frac{q_2\,\beta_z^2\,d_y^2}{2\,\gamma_y\,\gamma_z\,(\gamma_y+\gamma_z)} + \frac{q_1\,\beta_z^2\,d_x^2}{2\,\gamma_x\,\gamma_z\,(\gamma_x+\gamma_z)} + \frac{q_1\,\beta_y^2\,\beta_z^2\,c_x^2\,d_y^2\,(\gamma_x+\gamma_y+\gamma_z)}{2\,\gamma_x\,\gamma_y\,\gamma_z\,(\gamma_x+\gamma_y)(\gamma_y+\gamma_z)(\gamma_x+\gamma_z)} \\ & + \frac{q_1\,\beta_y\,\beta_z^2\,\gamma_y\,c_x\,d_x\,d_y\,(\gamma_x+\gamma_y+\gamma_z)}{\gamma_x\,\gamma_y\,\gamma_z\,(\gamma_x+\gamma_y)(\gamma_y+\gamma_z)(\gamma_x+\gamma_z)} \end{aligned} \tag{41}$$

where

$$d_x = \frac{\partial g_{xy}(x,y)}{\partial x}\Bigg|_{x=x_{av},y=y_{av}} \quad , \quad d_y = \frac{\partial g_{xy}(x,y)}{\partial y}\Bigg|_{x=x_{av},y=y_{av}} \tag{42}$$

In equation (39), the first term $\frac{q_3}{2\gamma_z}$ is the intrinsic noise term given by $< \delta z^2 >$ (equation (21)). The other terms represent noise propagated from the earlier stages, i.e, occur due to fluctuations in the number of x and y regulatory molecules. These terms describe the extrinsic noise.

4. Results and Discussion

We now calculate the variances $< \delta x^2 >_{tot}$, $< \delta y^2 >_{tot}$ and $< \delta z^2 >_{tot}$ for the different FFLs. Our goal is to compare the variances for the same as well as different FFLs. For simplicity we assume that all γ_i's (i =x, y, z)=1 and $K_{xy} = K_{yz} = K_{xz} = 1$. The mean levels of proteins x, y and z in the steady state are kept the same in all the cases so that a meaningful comparison between the variances can be made. Figures 2 and 3 show the plots of $< \delta x^2 >_{tot}$ (line with long dashes), $< \delta y^2 >_{tot}$ (line with short dashes) and $< \delta z^2 >_{tot}$ (solid line) versus β_y for the coherent and incoherent FFLs respectively. The regulation of the Z gene by the x and y proteins is achieved via the AND gate. The plots have been obtained keeping the mean protein levels x_{av}, y_{av} and z_{av} fixed at $m = 5.0$. For this we put

$$\beta_x = \beta_z = 10,\ k_a = k_d = 20,\ k_a^{''} = k_d^{''},\ k_{az} = 20,\ k_a^{'} = \frac{m\, k_d^{'}}{\beta_y - m},\ k_d^{'} = 20, \tag{43}$$

For the coherent Type-1 FFL with AND-gate regulation, the values of k_{ay} and $k_a^{''}$ are fixed from the relations

$$k_a^{'} = k_{ay}\, \frac{m^2}{1+m^2} \tag{44}$$

and

$$k_a^{''} = k_{az}\, \frac{m^4}{(1+m^2)^2} \tag{45}$$

Equivalent relations hold true for the other types of FFLs. An examination of figure 2 shows that the Type-1 coherent FFL is the least noisy amongst all the coherent FFLs. The number of times the Type-1,Type-2, Type-3 and Type-4 coherent FFLs appear in the GTRNs of *E. coli* (*S. cerevisiae*) are 28, 2, 4, 1 (26, 5, 0, 0) [4]. The most abundant coherent FFL, namely the Type-1 FFL, is the least noisy. This is not true for the incoherent FFLs. The number of times the Type-1, Type-2, Type-3, Type-4 incoherent FFLs appear in the GTRNs of *E. coli* (*S. cerevisiae*) are 5, 0, 1, 1 (21, 3, 1, 0). The most abundant incoherent FFL, namely, the Type-1 FFL, is more noisy than, say, the Type-4 incoherent FFL, which is practically absent in the GTRNs.

The reasons as to why some FFLs occur more often than the others in GTRNs, are not well understood. Generally speaking, reduced functionality of a motif may be a possible

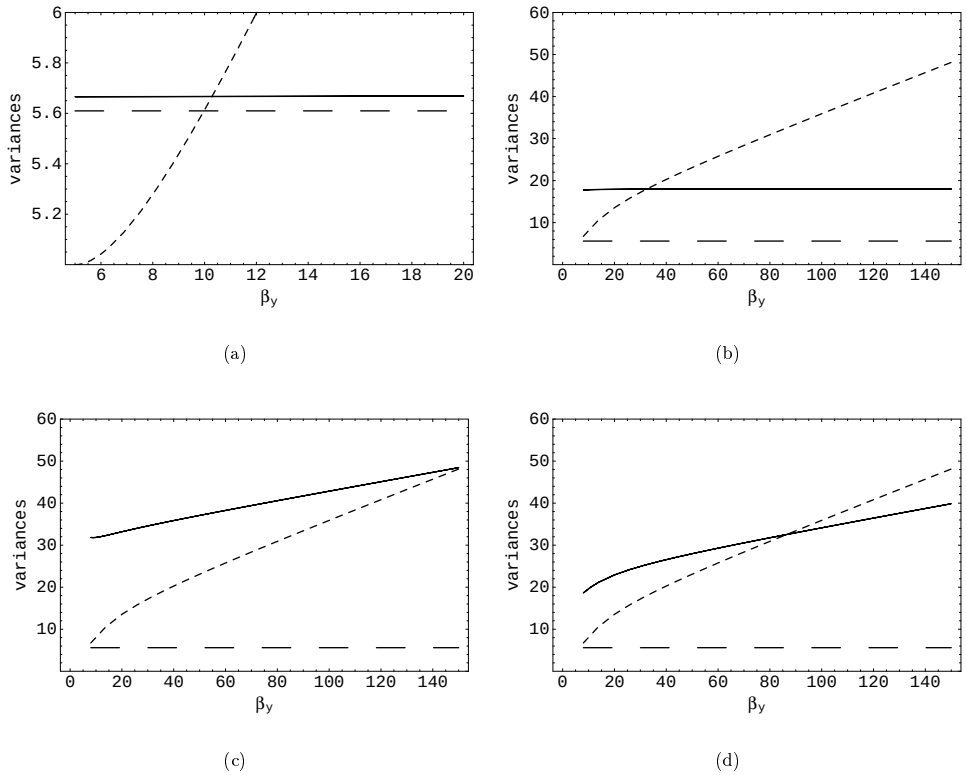

Figure 2: Variances $< \delta x^2 >_{tot}$ (line with long dashes), $< \delta y^2 >_{tot}$ (line with short dashes), $< \delta z^2 >_{tot}$ (solid line) versus β_y for (a) Type-1, (b) Type-2, (c) Type-3 and (d) Type-4 coherent FFLs controlled by AND-gate. The mean protein level is fixed at m=5. The other parameter values are mentioned in the text

reason for its lower abundance, i.e, being selected against during evolution. As suggested by Mangan and Alon [4], for AND-gate FFLs, Types-3 and 4 have reduced functionality compared to Types- 1 and 2, as the former respond to at most one input stimulus (S_x) whereas the latter respond to both the input stimuli S_x and S_y. Also, Type-1 coherent FFL gains advantage from increased cooperativity leading to a sharper response in the presence of stimuli. For low x concentrations, the effective Hill coefficient (a measure of cooperativity) is 6 (for $n = 2$ in equation (12)) whereas the same, for the other FFLs, is 2. We now discuss the relationship between noise, function and abundance. For the sake of clarity, we focus attention on the Type-1 and Type-4 coherent FFLs. Figure 4 shows plots for the total variances around the mean protein level $m = 5$ when the input noise $< \delta x^2 >_{tot}$ is higher than that in the cases of figures 2 and 3. The parameter values changed from equation (43) are $k_a = k_d = 5$, $k_d^{'} = 30$ and $k_{az} = 30$. In the case of the Type-1 coherent FFL, one finds the existence of a parameter region in which the variance decreases in the successive stages of the FFL so that the output noise is less than the input noise. Such a parameter region is absent in the case of the Type-4 coherent FFL. Another

notable feature of the plots in figures 2 and 4 is that $< \delta y^2 >_{tot}$ and $< \delta z^2 >_{tot}$ in the case of the Type-1 FFL have almost linear dependences on β_y whereas the same quantities are more nonlinear in the case of the Type-4 FFL. For the Type-1 FFL, the dominant contribution to $< \delta z^2 >_{tot}$ is from the internal noise associated with the expression of the Z gene. Fluctuations in the x and y protein levels have little effect on the total noise. In the case of the Type-4 FFL, the extrinsic contribution to noise is greater than that in the case of the Type-1 FFL. In short, figures 2 and 4 show that the Type-1 FFL acts as a better filter of noise. As mentioned in the Introduction, one possible function of coherent FFLs is as a persistent detector or equivalently as a filter which attenuates the input noise. The Type-1 coherent FFL being less noisy than the Type-4 coherent FFL, functions better as a noise filter. The reduced functionality of the Type-4 coherent FFL explains its lower abundance from an evolutionary point of view. Similar reasoning holds true for Type-2 and Type-3 coherent FFLs. Thus for coherent FFLs, noise is disadvantageous as it erodes the function of a FFL as a persistent detector. For incoherent FFLs, functioning as sign-sensitive accelerators, noise appears to have no direct relationship with abundance, i.e, noise is not detrimental to the functioning of the FFLs.

Our analysis of the noise characteristics of FFLs is based on the Langevin formalism which is approximate in nature. To establish the validity of the results, we have calculated the variances using Monte Carlo simulation based on the GA [17, 18]. The GA provides a numerical solution of the Master Equation leading to an accurate description of the time course of evolution of a stochastic system. A brief description of the GA is as follows. Consider N chemical species participating in M chemical reactions. Let $X(i)$, $i = 1, 2, 3, ..., N$ denotes the number of molecules of the ith chemical species. Given the values of $X(i)$, $i = 1, 2, 3, ...N$ at time t, the GA is designed to answer two questions: (1) when will the next reaction occur? and (2) what type of reaction will it be? Let the next reaction occur at time $t + \tau$. Knowing the type of reaction, one can adjust the numbers of participating molecules in accordance with the specific reaction scheme. Thus, with repeated applications of the GA, one can keep track of how the numbers, $X(i)$'s, change as a function of time due to the occurrence of M different types of chemical reactions. Each reaction μ ($\mu = 1, 2, 3, ..., M$) has a stochastic rate constants C_μ associated with it. The rate constant has the interpretation that $C_\mu dt$ is the probability that a particular combination of reacting molecules participates in the μth reaction in the infinitesimal time interval $(t, t + dt)$. If h_μ is the number of distinct molecular combinations for the μth reaction, then $a_\mu dt = h_\mu C_\mu dt$ is the probability that the μth reaction occurs in the infinitesimal time interval (t, t+dt). The implementation of the GA algorithm is described in detail in Refs. [17, 18]. We use the algorithm to determine the evolution of the number of z proteins of a FFL as a function of time. Figures 5(a) and 6(a) show the results for the coherent Type-1 and Type-4 FFL respectively. The solid line, in each case, represents the mean trajectory obtained from a solution of the deterministic equations. The reactions considered are those associated with a FFL. Expression of each gene X , Y and Z is according to the reaction scheme RS-2 (equation (6)). For the X gene, there is no regulatory molecule S. The x proteins dimerize (equation (12) with $n = 2$) and the dimers regulate expression of the Y gene. The y proteins also dimerize to regulate expression of the gene Z. Considering AND-gate

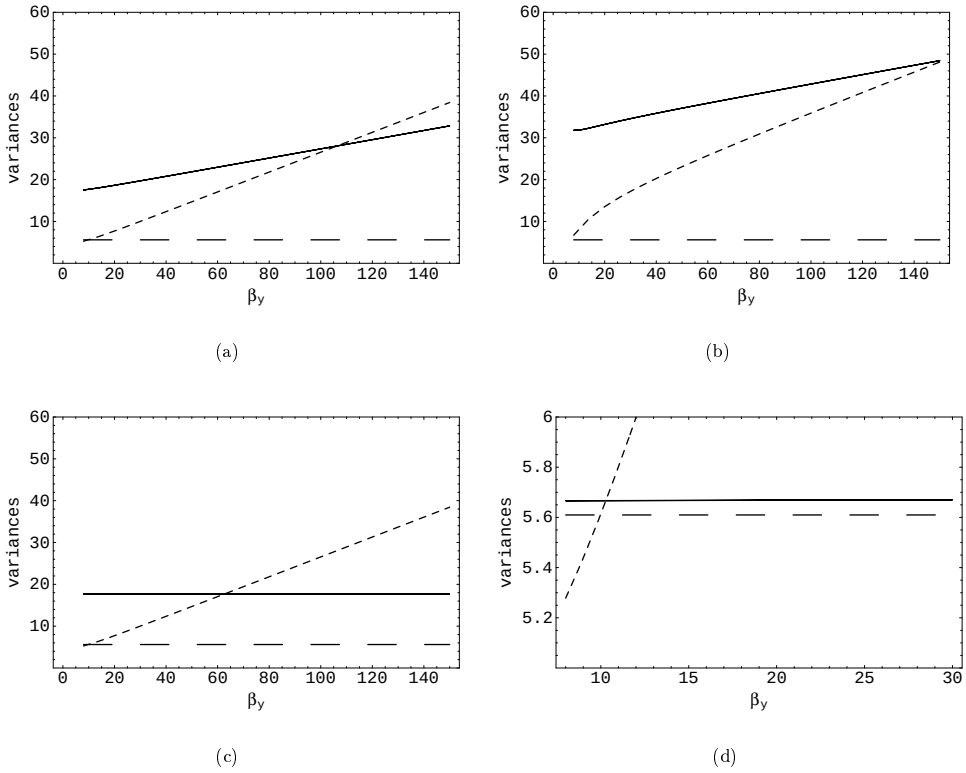

Figure 3: Variances $< \delta x^2 >_{tot}$ (line with long dashes), $< \delta y^2 >_{tot}$ (line with short dashes), $< \delta z^2 >_{tot}$ (solid line) versus β_y for (a) Type-1, (b) Type-2, (c) Type-3 and (d) Type-4 incoherent FFLs controlled by AND-gate. The mean protein level is fixed at m=5. The other parameter values are mentioned in the text.

regulation of the Z gene expression, both the x and y protein dimers bind simultaneously at the operator region for activation of the gene. Other possibilities like the operator region unoccupied or occupied by a single dimer are considered but the gene remains in the inactive state in these cases. The stochastic rate constants C_μ's are equal to the rate constants k_μ's since in the deterministic approach the numbers and not the concentrations of the different molecules are considered. Figure 5(b) and 6(b) show the histograms describing the distribution of protein levels, $N(z)$ versus z, for the coherent Type-1 and Type-4 FFLs respectively. The histograms have been obtained by accumulating data over 5000 trial runs. The distribution is broader in the case of the Type-4 coherent FFL indicating that it is more noisy than the Type-1 FFL. The variances for Type-1 and Type-4 distributions are 110.612 and 329.990 respectively. The simulation results support the results obtained by using the Langevin formalism that the Type-4 coherent FFL is more noisy than the Type-1 coherent FFL.

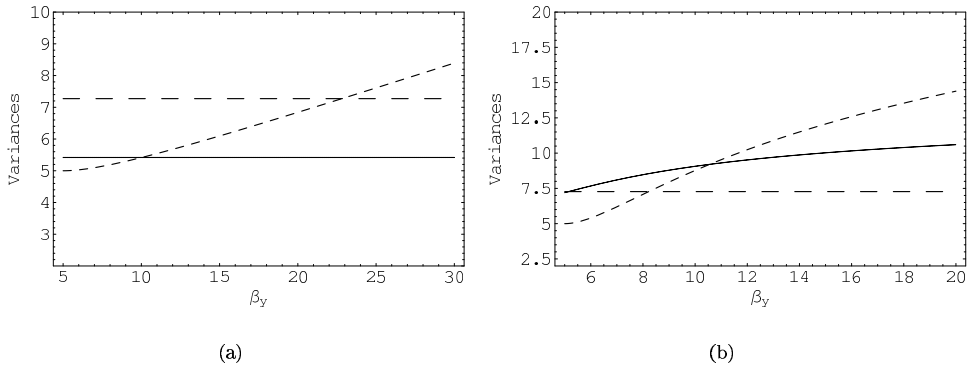

Figure 4: Variances $< \delta x^2 >_{tot}$ (line with long dashes), $< \delta y^2 >_{tot}$ (line with short dashes), $< \delta z^2 >_{tot}$ (solid line) versus β_y for (a) Type-1 coherent FFL and (b) Type-4 coherent FFL controlled by AND-gate. The mean protein level is fixed at m=5. The input noise is greater than that in the case of figure 2.

5. Conclusion and Outlook

In this paper, we have studied the noise characteristics of coherent and incoherent FFLs using the Langevin formalism as well as a numerical simulation technique based on the Gillespie algorithm. Noise is undesirable if it affects operational reliability. Coherent FFLs function as noise filters and the performance of the Type-1 FFL is found to be the best since the propagation of noise associated with the input signal is the least in this case. The coherent Type-1 FFL is the most abundant of FFL motifs appearing in the GTRNs of simple organisms. The functional superiority of the Type-1 FFL, amongst the four coherent FFLs, is the main reason why the particular motif is favoured by natural selection. Mangan and Alon [4] have speculated that increased effective cooperativity of the Type-1 FFL might be responsible for its evolutionary advantage. Thattai and van Oudenaarden [16] have shown that increased cooperativity leads to noise reduction. This possibly explains why the Type-1 coherent FFL has less output noise than the other coherent FFLs. For the incoherent FFLs, no clear conclusion regarding the role of noise can be arrived at as concrete results are lacking. Noise may be advantageous to function in certain cases. Stochastic resonance is a phenomena in which noise in threshold systems facilitates detection of subthreshold signals [21]. In stochastic focusing, fluctuations (noise) sharpen the response to an input signal, i.e, make a graded response mechanism work more like a threshold one [22]. Further studies are needed to ascertain whether noise aids the function of incoherent FFLs in some manner similar to stochastic focusing. If this is true, then the most abundant motif need not be the least noisy. Regulatory cascades of which the FFL is a special case can exhibit interesting kinetic phenomena which include even transient ones like pulse generation [23, 24]. It will be of considerable interest to determine the effect of noise on such phenomena.

Fraser et al. [14] have addressed the question of whether noise associated with GE

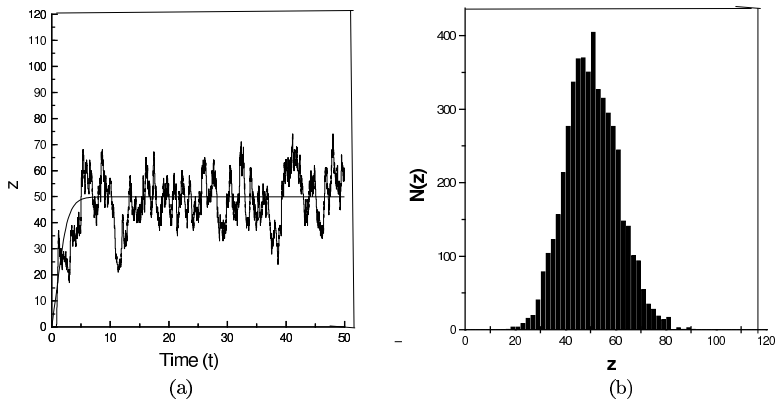

Figure 5: (a) The number of proteins z(t) as a function of time t for the Type-1 coherent FFL. The time trajectory is obtained using the GA. The solid line determines the mean curve. (b) Histogram describing the distribution of protein levels (N(z) versus z) in the steady state.

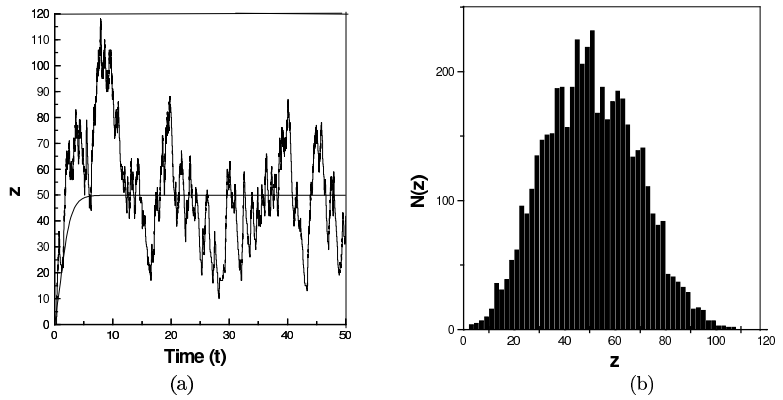

Figure 6: (a) The number of proteins z(t) as a function of time t for the Type-4 coherent FFL. The time trajectory is obtained using the GA. The solid line determines the mean curve. (b) Histogram describing the distribution of protein levels (N(z) versus z) in the steady state.

has any significant effect on the fitness of an organism. They have estimated the noise in protein production of almost all the *S. cerevisiae* genes using an experimentally verified model of stochastic GE. Their major finding is that noise is minimized in the cases of genes for which it is likely to be most harmful. These genes include essential genes, i.e, genes whose deletion is lethal to the organism and genes which synthesize the subunits of multi-protein complexes. Both types of genes are expected to be sensitive to noise. For essential genes, fluctuations in protein levels may have considerable effect on functional viability if the levels fall below the threshold required for normal cellular activity. Similarly, in the case of a multy-protein complex, fluctuations in the amounts of protein subunits may hinder the appropriate assembly of the entire complex. The observations of Fraser et al. are in agreement with our results on coherent FFLs. Since noise has a deleterious effect on the function of a coherent FFL as a persistence detector, it is minimized in the case of the best performing Type-1 FFL.

Acknowledgements

R.K. is supported by the Council of Scientific and Industrial Research, India under Sanction No. 9/15 (239)/2002 - EMR-1
B.G. is supported by the Council of Scientific and Industrial Research, India under Sanction No. 9/15 (282)/2003 - EMR-1

References

[1] Alm E and Arkin A P 2003 *Curr. Opin. Struct. Biol.* 13 193-202

[2] Barabási A L and Oltvai Z N 2004 *Nat. Rev. Genet.* 5 101-03

[3] Shen-Orr S, Milo R, Mangan S and Alon U 2002 *Nature Genet.* 31 64-68

[4] Mangan S and Alon U 2003 *Proc. Natl. Acad. Sci.* 100 11980-85

[5] Milo R et al 2002 *Science* 298 824-27

[6] Rao C V, Wolf D M and Arkin A P 2002 *Nature* 420 231-37

[7] Thattai M and van Oudenaarden A 2001 *Proc. Natl. Acad. Sci.* 98 8614-19

[8] Swain P S, Elowitz M B and Siggia E D 2002 *Proc. Natl. Acad. Sci.* 99 12795-800

[9] Kepler T B and Elston T C 2001 *Biophys. J.* 81 3116-36

[10] Bose I and Karmakar R 2004 *in Biology of Genetic Dominance* edited by Veitia R A (USA :Landes Bioscience), Chapter 6

[11] Ozbudak E M, Thattai M, Kurtser I, Grossman A D and van Oudenaarden A 2002 *Nature Genet.* 31 69-73

[12] Elowitz M B, Levine A J, Siggia E D and Swain P S 2002 *Science* 297 1183-86

[13] Blake W J, Kaern M, Cantor C R and Collins J J 2003 *Nature* 422 633-37

[14] Fraser H B et al 2004 *PloS Biology* 2 1-13

[15] Mangan S, Zaslaver A and Alon U 2003 *J. Mol. Biol.* 334 197-204

[16] Thattai M and van Oudernaarden A 2002 *Biophys. J.* 82 2943-50

[17] Gillespie D T 1977 *J. Phys. Chem.* 81 2340-61

[18] Gillespie D T 1976 *J. Comput. Phys.* 22 403-34

[19] Peccoud J and Ycart B 1995 *Theor. Pop. Biol.* 98 222-34

[20] Cook D L, Gerber A N and Tapscott S J 1998 *Proc. Natl. Acad.* Sci 95 15641-46

[21] Wiesenfeld K and Moss F 1995 Nature 373 33-6

[22] Paulsson J, Berg O G and Ehrenberg M 2000 *Proc. Natl. Acad. Sci.* 97 7148-53

[23] Bolouri H and Davidson E H 2003 *Proc. Natl. Acad. Sci.* 100 9371-76

[24] Basu S, Mehreja R, Thiberge S, Chen M- T and Weiss R 2004 *Proc. Natl. Acad. Sci.* 101 6355-60