
Biomolecular Feedback Systems

Domitilla Del Vecchio
MIT

Richard M. Murray
Caltech

Version 1.0a, January 19, 2014
© California Institute of Technology
All rights reserved.

This is the electronic edition of *Biomolecular Feedback Systems* and is available from <http://www.cds.caltech.edu/~murray/BFS>. Hardcover editions may be purchased from Princeton University Press, <http://press.princeton.edu/titles/10285.html>.

This manuscript is for personal use only and may not be reproduced, in whole or in part, without written consent from the publisher (see <http://press.princeton.edu/permissions.html>).

Chapter 7

Design Tradeoffs

In this chapter we describe some of the design tradeoffs arising from the interaction between synthetic circuits and the host organism. We specifically focus on two issues. The first issue is concerned with the effects of competition for shared cellular resources on circuit's behavior. In particular, circuits (endogenous and exogenous) share a number of cellular resources, such as RNA polymerase, ribosomes, ATP, enzymes, and nucleotides. The insertion or induction of synthetic circuits in the cellular environment changes the demand for these resources, with possibly undesired repercussions on the functioning of the circuits. Independent circuits may become coupled when they share common resources that are not in overabundance. This fact leads to constraints among the concentrations of proteins in synthetic circuits, which should be accounted for in the design phase. The second issue we consider is the effect of biological noise on the design of devices requiring high gains. Specifically, we illustrate possible design tradeoffs between retroactivity attenuation and noise amplification that emerge due to the intrinsic noise of biomolecular reactions.

7.1 Competition for Shared Cellular Resources

Exogenous circuits, just like endogenous ones, use cellular resources—such as ribosomes, RNA polymerase (RNAP), enzymes and ATP—that are shared among all the circuitry of the cell. From a signals and systems point of view, these interactions can be depicted as in Figure 7.1. The cell's endogenous circuitry produces resources as output and exogenous synthetic circuits take these resources as inputs. As a consequence, as seen in Chapter 6, there is retroactivity from the exogenous circuits to the cellular resources. This retroactivity creates indirect coupling between the exogenous circuits and can lead to undesired crosstalk. In this chapter, we study the effect of the retroactivity from the synthetic circuits to shared resources in the cellular environment by focusing on the effect on availability of RNA polymerase and ribosomes, for simplicity. We then study the consequence of this retroactivity, illustrating how the behavior of individual circuits becomes coupled. These effects are significant for any resource whose availability is not in substantial excess compared to the demand by exogenous circuits.

In order to illustrate the problem, we consider the simple system shown in Figure 7.2, in which two modules, a constitutively expressed gene (Module 1) and a gene activated by a transcriptional activator A (Module 2), are present in the cel-

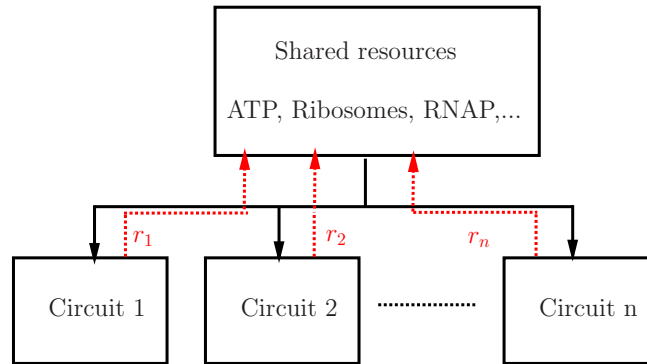


Figure 7.1: The cellular environment provides resources to synthetic circuits, such as RNA polymerase, ribosomes, ATP, nucleotides, proteases, etc. These resources can be viewed as an “output” of the cell’s endogenous circuitry and an input to the exogenous circuits. Circuit i takes these resources as input and, as a consequence, it causes a retroactivity r_i to its input. Hence, the endogenous circuitry has a retroactivity to the output s that encompasses all the retroactivities applied by the exogenous circuits.

lular environment. In theory, Module 2 should respond to changes in the activator A concentration, while Module 1, having a constitutively active promoter, should display a constant expression level that is independent of the activator A concentration. Experimental results, however, indicate that this is not the case: Module 1’s output protein concentration P_1 also responds to changes in the activator A concentration. In particular, as the activator A concentration is increased, the concentration of protein P_1 can substantially decrease. This fact can be qualitatively explained by the following reasoning. When A is added, RNA polymerase can bind to DNA promoter D_2 and start transcription, so that the free available RNA polymerase decreases as some is bound to the promoter D_2 . Transcription of Module 2 generates mRNA and hence ribosomes will have more ribosome binding sites to which they can bind, so that less ribosomes will be free and available for other reactions. It follows that the addition of activator A leads to an overall decrease of the free RNA polymerase and ribosomes that can take part in the transcription and translation reactions of Module 1. The net effect is that less of P_1 protein will be produced.

The extent of this effect will depend on the overall availability of the shared resources, on the biochemical parameters, and on whether the resources are regulated. For example, it is known that ribosomes are internally regulated by a combination of feedback interactions [59]. This, of course, may help compensate for changes in the demand of these resources, though experiments demonstrate that the coupling effects are indeed noticeable [95].

In this chapter, we illustrate how this effect can be mathematically explained by explicitly accounting for the usage of RNA polymerase and ribosomes in the

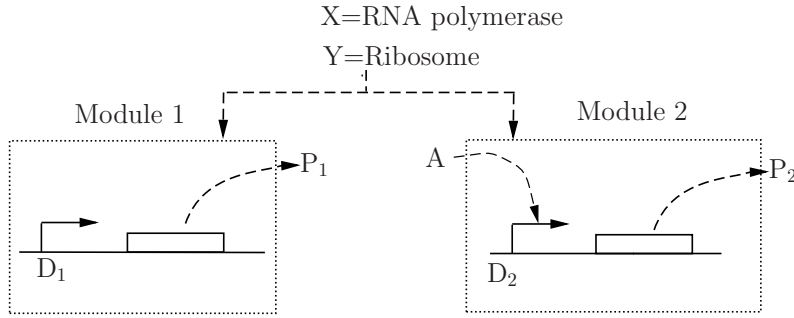


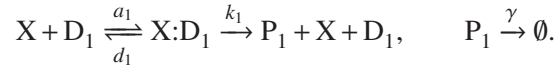
Figure 7.2: Module 1 has a constitutively active promoter that controls the expression of protein P_1 , while Module 2 has a promoter activated by activator A , which controls the expression of protein P_2 . The two modules do not share any transcription factors, so they are not “connected”. Both of them use RNA polymerase (X) and ribosomes (Y) for the transcription and translation processes.

transcription and translation models of the circuits. To simplify the mathematical analysis and to gather analytical understanding of the key parameters at the basis of this phenomenon, we first focus on the usage of RNA polymerase, neglecting the usage of ribosomes. We then provide a computational model that accounts for both RNA polymerase and ribosome utilization and illustrate quantitative simulation results.

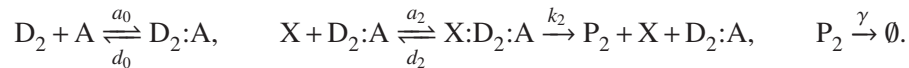
Analytical study

To illustrate the essence of the problem, we assume that gene expression is a one-step process, in which the RNA polymerase binds to the promoter region of a gene resulting in a transcriptionally active complex that, in turn, produces the protein. That is, we will be using the lumped reactions (2.12), in which on the right-hand side of the reaction we have the protein instead of the mRNA.

By virtue of this simplification, we can write the reactions describing Module 1 as:



The reactions describing Module 2 can be written similarly, recalling that in the presence of an activator the reactions should be modified according to equation (2.21). Taking this into account, the reactions of Module 2 are given by



We let $D_{\text{tot},1}$ and $D_{\text{tot},2}$ denote the total concentration of DNA for Module 1 and Module 2, respectively, and we let $K_0 = d_0/a_0$, $K_1 = d_1/a_1$, and $K_2 = d_2/a_2$. By

approximating the complexes concentrations by their quasi-steady state values, we obtain the expressions

$$[X:D_1] = D_{\text{tot},1} \frac{X/K_1}{1 + X/K_1}, \quad [X:D_2:A] = D_{\text{tot},2} \frac{(A/K_0)(X/K_2)}{1 + (A/K_0)(1 + X/K_2)}. \quad (7.1)$$

As a consequence, the differential equation model for the system is given by

$$\begin{aligned} \frac{dP_1}{dt} &= k_1 D_{\text{tot},1} \frac{X/K_1}{1 + X/K_1} - \gamma P_1, \\ \frac{dP_2}{dt} &= k_2 D_{\text{tot},2} \frac{(A/K_0)(X/K_2)}{1 + (A/K_0)(1 + X/K_2)} - \gamma P_2, \end{aligned}$$

so that the steady state values of P_1 and P_2 are given by

$$P_1 = \frac{k_1 D_{\text{tot},1}}{\gamma} \frac{X/K_1}{1 + X/K_1}, \quad P_2 = \frac{k_2 D_{\text{tot},2}}{\gamma} \frac{(A/K_0)(X/K_2)}{1 + (A/K_0)(1 + X/K_2)}.$$

These two values are indirectly coupled through the conservation law of RNA polymerase. Specifically, we let X_{tot} denote the total concentration of RNA polymerase. This value is mainly determined by the cell growth rate and for a given growth rate it is about constant [15]. Then, we have that $X_{\text{tot}} = X + [X:D_1] + [X:D_2:A]$, which, considering the expressions of the quasi-steady state values of the complexes concentrations in equation (7.1), leads to

$$X_{\text{tot}} = X + D_{\text{tot},1} \frac{X/K_1}{1 + X/K_1} + D_{\text{tot},2} \frac{(A/K_0)(X/K_2)}{1 + (A/K_0)(1 + X/K_2)}. \quad (7.2)$$

We next study how the steady state value of X is affected by the activator concentration A and how this effect is reflected in a dependency of P_1 on A . To perform this study, it is useful to write $\alpha := (A/K_0)$ and note that for α sufficiently small (sufficiently small amounts of activator A), we have that

$$\frac{\alpha(X/K_2)}{1 + \alpha(1 + X/K_2)} \approx \alpha(X/K_2).$$

Also, to simplify the derivations, we assume that the binding of X to D_1 is sufficiently weak, that is, $X \ll K_1$. In light of this, we can rewrite the conservation law (7.2) as

$$X_{\text{tot}} = X + D_{\text{tot},1} \frac{X}{K_1} + D_{\text{tot},2} \alpha \frac{X}{K_2}.$$

This equation can be explicitly solved for X to yield

$$X = \frac{X_{\text{tot}}}{1 + (D_{\text{tot},1}/K_1) + \alpha(D_{\text{tot},2}/K_2)}.$$

This expression depends on α , and hence on the activator concentration A . Specifically, as the activator is increased, the value of the free X concentration monotonically decreases. As a consequence, the equilibrium value P_1 will also depend on A according to

$$P_1 = \frac{k_1 D_{\text{tot},1}}{\gamma} \frac{X_{\text{tot}}/K_1}{1 + (D_{\text{tot},1}/K_1) + \alpha(D_{\text{tot},2}/K_2)},$$

so that P_1 monotonically decreases as A is increased. That is, Module 1 responds to changes in the activator of Module 2. From these expressions, we can also deduce that if $D_{\text{tot},1}/K_1 \gg \alpha D_{\text{tot},2}/K_2$, that is, the demand for RNA polymerase in Module 1 is much larger than that of Module 2, then changes in the activator concentration will lead to small changes in the free amount of RNA polymerase and in P_1 .

This analysis illustrates that forcing an increase in the expression of any protein causes an overall decrease in available resources, which leads to a decrease in the expression of other proteins. As a consequence, there is a tradeoff between the amount of protein produced by one circuit and the amount of proteins produced by different circuits. In addition to a design tradeoff, this analysis illustrates that “unconnected” circuits can affect each other because they share common resources. This can, in principle, lead to a dramatic departure of a circuit’s behavior from its nominal one.

As an exercise, the reader can verify that similar results hold in the case in which Module 2 has a repressible promoter instead of one that can be activated (see Exercise 7.2).

The model that we have presented here contains many simplifications. In addition to the mathematical approximations performed and to the fact that we did not account for ribosomes, the model neglects the transcription of endogenous genes. In fact, RNA polymerase is also used for transcription of chromosomal genes. While the qualitative behavior of the coupling between Module 1 and Module 2 is not going to be affected by including endogenous transcription, the extent of this coupling may be substantially impacted. In the next section, we illustrate how the presence of endogenous genes may affect the extent to which the availability of RNA polymerase decreases upon addition of exogenous genes.

7.2 Estimates of RNA Polymerase Perturbations by Exogenous Plasmids

In the previous section, we illustrated the mechanism by which the change in the availability of a shared resource, due to the addition of synthetic circuits, can cause crosstalk between unconnected circuits. The extent of this crosstalk depends on the amount by which the shared resource changes. This amount, in turn, depends on the specific values of the dissociation constants, the total resource amounts, and the fraction of resource that is used already by natural circuits. Here, we consider

how the addition of an external plasmid affects the availability of RNA polymerase, considering a simplified model of the interaction of RNA polymerase with the exogenous and natural DNA.

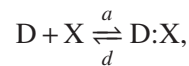
In *E. coli*, the amount of RNA polymerase and its partitioning mainly depends on the growth rate of the cell [15]: with 0.6 doublings/hour there are only 1500 molecules/cell, while with 2.5 doublings/hour this number is 11400. The fraction of active RNA polymerase molecules also increases with the growth rate. For illustration purposes, we assume here that the growth rate is the highest considered in [15], so that 1 molecule/cell corresponds to approximately 1nM concentration. In this case, a reasonable partitioning of the total pool of RNA polymerase of concentration $X_{\text{tot}} = 12 \mu\text{M}$ is the following [55]:

specifically DNA-bound (at promoter) X_s : 30% (4000 molecules/cell, that is, $X_s = 4 \mu\text{M}$),

non-specifically DNA-bound X_n : 60% (7000 molecules/cell, that is, $X_n = 7 \mu\text{M}$),

free X: 10% (1000 molecules/cell, that is, $X = 1 \mu\text{M}$).

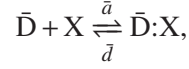
By [16], the number of initiations per promoter can be as high as 30/minute in the case of constitutive promoters, and 1-3/minute for regulated promoters. Here, we choose an effective value of 5 initiations/minute per promoter, so that on average, 5 molecules of RNA polymerase can be simultaneously transcribing each gene, as transcribing a gene takes approximately a minute [4]. There are about 1000 genes expressed in exponential growth phase [46], hence we approximate the number of promoter binding sites for X to 5000, or $D_{\text{tot}} = 5 \mu\text{M}$. The binding reaction for specific binding is of the form



in which D represents DNA promoter binding sites DNA^P in total concentration D_{tot} . Consequently, we have $D_{\text{tot}} = D + [D:X]$. At the equilibrium, we have $[D:X] = X_s = 4 \mu\text{M}$ and $D = D_{\text{tot}} - [D:X] = D_{\text{tot}} - X_s = 1 \mu\text{M}$. With dissociation constant $K_d = d/a$ the equilibrium is given by $0 = DX - K_d[D:X]$, hence we have that $K_d = DX/[D:X] = 0.25 \mu\text{M}$, which can be interpreted as an “effective” dissociation constant. This is in the range 1 nM – 1 μM suggested by [38] for specific binding of RNA polymerase to DNA. Therefore, we are going to model the specific binding of RNA polymerase to the chromosome of *E. coli* in exponential growth phase as one site with concentration D_{tot} and effective dissociation constant K_d .

Furthermore, we have to take into account the rather significant amount of RNA polymerase bound to the DNA other than at the promoter region ($X_n = 7 \mu\text{M}$). To do so, we follow a similar path as in the case of specific binding. In particular, we

model the non-specific binding of RNA polymerase to DNA as

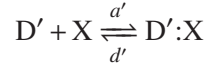


in which \bar{D} represents DNA binding sites with concentration \bar{D}_{tot} and effective dissociation constant $\bar{K}_d = \bar{d}/\bar{a}$. At the equilibrium, we have that the concentration of RNA polymerase non-specifically bound to DNA is given by

$$X_n = [\bar{D}:X] = \frac{\bar{D}_{\text{tot}}X}{X + \bar{K}_d}.$$

As the dissociation constant \bar{K}_d of non-specific binding of RNA polymerase to DNA is in the range 1 – 1000 μM [38], we have $X \ll \bar{K}_d$, yielding $X_n = [\bar{D}:X] \approx X\bar{D}_{\text{tot}}/\bar{K}_d$. Consequently, we obtain $\bar{D}_{\text{tot}}/\bar{K}_d = X_n/X = 7$.

Now, we can consider the addition of synthetic plasmids. Specifically, we consider high-copy number plasmids (copy number 100 – 300) with one copy of a gene under the control of a constitutive promoter. We abstract it by a binding site for RNA polymerase D' to which X can bind according to the following reaction:



where D' is the RNA polymerase-free binding site and $D':X$ is the site bound to RNA polymerase. Consequently, we have $D'_{\text{tot}} = D' + [D':X]$, where $D'_{\text{tot}} = 1 \mu\text{M}$, considering 200 copies of plasmid per cell and 5 RNA polymerase molecules per gene. The dissociation constant corresponding to the above reaction is given by $K' = d'/a'$. At the steady state we have

$$[D':X] = D'_{\text{tot}} \frac{X}{K'_d + X},$$

together with the conservation law for RNA polymerase given by

$$X + [D:X] + [\bar{D}:X] + [D':X] = X_{\text{tot}}. \quad (7.3)$$

Here, we did not account for RNA polymerase molecules paused or queuing on the chromosome; moreover, we also neglected the resistance genes on the plasmid and all additional sites (specific or not) to which RNA polymerase can also bind. Hence, we are underestimating the effect of load presented by the plasmid. Solving equation (7.3) for the free RNA polymerase amount X gives the following results. These results depend on the ratio between the effective dissociation constant K_d of RNA polymerase binding with the natural DNA promoters and the dissociation constant K'_d of binding with the plasmid promoter:

$K'_d = 0.1K_d$ (RNA polymerase binds stronger to the plasmid promoter) results in $X = 0.89 \mu\text{M}$, that is, the concentration of free RNA polymerase decreases by about 11%;

$K'_d = K_d$ (binding is the same) results in $X = 0.91 \mu\text{M}$, consequently, the concentration of free RNA polymerase decreases by about 9%;

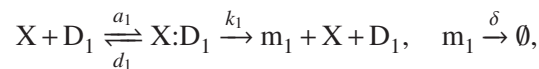
$K'_d = 10K_d$ (RNA polymerase binds stronger to the chromosome) results in $X = 0.97 \mu\text{M}$, which means that the concentration of free RNA polymerase decreases by about 3%.

Note that the decrease in the concentration of free RNA polymerase is greatly reduced by the significant amount of RNA polymerase being non-specifically bound to the DNA. For instance, in the second case when $K'_d = K_d$, the RNA polymerase molecules sequestered by the synthetic plasmid can be partitioned as follows: about 10% is taken from the pool of free RNA polymerase molecules X , another 10% comes from the RNA polymerase molecules specifically bound, and the overwhelming majority (80%) decreases the concentration of RNA polymerase non-specifically bound to DNA. That is, this weak binding of RNA polymerase to DNA acts as a buffer against changes in the concentration of free RNA polymerase.

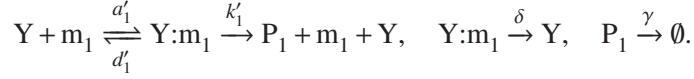
We conclude that if the promoter on the synthetic plasmid has a dissociation constant for RNA polymerase that is in the range of the dissociation constant of specific binding, the perturbation on the available free RNA polymerase is about 9%. This perturbation, even if fairly small, may in practice result in large effects on the protein concentration. This is because it may cause a large perturbation in the concentration of free ribosomes. In fact, one added copy of an exogenous plasmid will lead to transcription of several mRNA molecules, which will demand ribosomes for translation. Hence, a small increase in the demand for RNA polymerase may be associated with a dramatically larger increase in the demand for ribosomes. This is illustrated in the next section through a computational model including ribosome sharing.

Computational model and numerical study

In this section, we introduce a model of the system in Figure 7.2, in which we consider both the RNA polymerase and the ribosome usage. We let the concentration of RNA polymerase be denoted by X and the concentration of ribosomes be denoted by Y . We let m_1 and P_1 denote the concentrations of the mRNA and protein in Module 1 and let m_2 and P_2 denote the concentrations of the mRNA and protein in Module 2. The reactions of the transcription process in Module 1 are given by (see Section 2.2):



while the translation reactions are given by

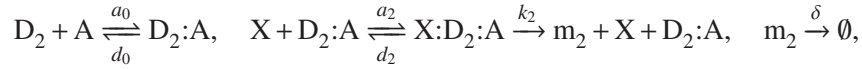


The resulting system of differential equations is given by

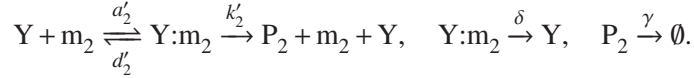
$$\begin{aligned} \frac{d}{dt} [X:D_1] &= a_1 X D_1 - (d_1 + k_1) [X:D_1], \\ \frac{dm_1}{dt} &= k_1 [X:D_1] - a'_1 Y m_1 + d'_1 [Y:m_1] - \delta m_1 + k'_1 [Y:m_1], \\ \frac{d}{dt} [Y:m_1] &= a'_1 Y m_1 - (d'_1 + k'_1) [Y:m_1], \\ \frac{dP_1}{dt} &= k'_1 [Y:m_1] - \gamma P_1, \end{aligned} \quad (7.4)$$

in which $D_1 = D_{\text{tot},1} - [X:D_1]$ from the conservation law of DNA in Module 1.

The reactions of the transcription process in Module 2 are given by (see Section 2.3)



while the translation reactions are given by



The resulting system of differential equations is given by

$$\begin{aligned} \frac{d}{dt} [D_2:A] &= a_0 D_2 A - d_0 [D_2:A] - a_2 X [D_2:A] + (d_2 + k_2) [X:D_2:A], \\ \frac{d}{dt} [X:D_2:A] &= a_2 X [D_2:A] - (d_2 + k_2) [X:D_2:A], \\ \frac{dm_2}{dt} &= k_2 [X:D_2:A] - a'_2 Y m_2 + d'_2 [Y:m_2] - \delta m_2 + k'_2 [Y:m_2], \\ \frac{d}{dt} [Y:m_2] &= a'_2 Y m_2 - (d'_2 + k'_2) [Y:m_2] - \delta [Y:m_2], \\ \frac{dP_2}{dt} &= k'_2 [Y:m_2] - \gamma P_2, \end{aligned} \quad (7.5)$$

in which we have that $D_2 = D_{\text{tot},2} - [D_2:A] - [X:D_2:A]$ by the conservation law of DNA in Module 2.

The two modules are coupled by the conservation laws for RNA polymerase and ribosomes given by

$$X_{\text{tot}} = X + [X:D_1] + [X:D_2:A], \quad Y_{\text{tot}} = Y + [Y:m_1] + [Y:m_2],$$

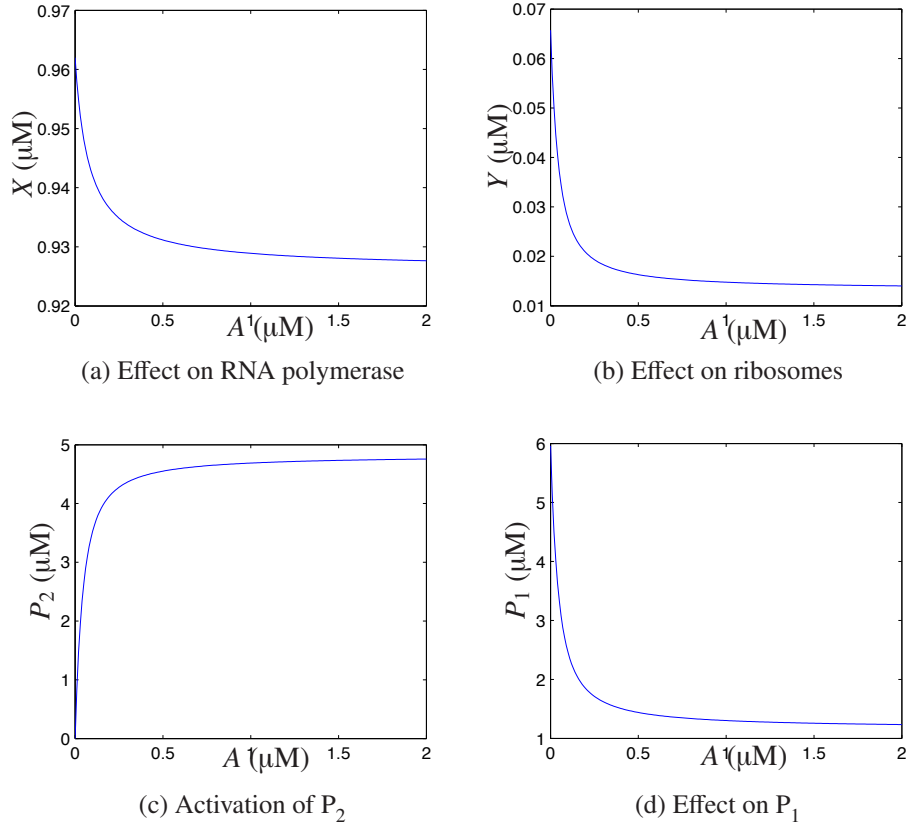


Figure 7.3: Simulation results for the ordinary differential equation model (7.4)–(7.5). For this model, the parameter values were taken from <http://bionumbers.hms.harvard.edu> as follows. For the concentrations, we have set $X_{\text{tot}} = 1 \mu\text{M}$, $Y_{\text{tot}} = 10 \mu\text{M}$, and $D_{\text{tot},1} = D_{\text{tot},2} = 0.2 \mu\text{M}$. The values of the association and dissociation rate constants were chosen such that the corresponding dissociation constants were in the range of dissociation constants for specific binding. Specifically, we have $a_0 = 10 \mu\text{M}^{-1}\text{min}^{-1}$, $d_0 = 1 \text{min}^{-1}$, $a_2 = 10 \mu\text{M}^{-1}\text{min}^{-1}$, $d_2 = 1 \text{min}^{-1}$, $a'_2 = 100 \mu\text{M}^{-1}\text{min}^{-1}$, $d'_2 = 1 \text{min}^{-1}$, $a_1 = 10 \mu\text{M}^{-1}\text{min}^{-1}$, $d_1 = 1 \text{min}^{-1}$, $a'_1 = 10 \mu\text{M}^{-1}\text{min}^{-1}$, and $d'_1 = 1 \text{min}^{-1}$. The transcription and translation rate constants were chosen to give a few thousands of protein copies per cell and calculated considering the elongation speeds, the average length of a gene, and the average number of RNA polymerase per gene and of ribosomes per transcript. The resulting values chosen are given by $k_1 = k_2 = 40 \text{min}^{-1}$ and $k'_1 = k'_2 = 0.006 \text{min}^{-1}$. Finally, the decay rates are given by $\gamma = 0.01 \text{min}^{-1}$ corresponding to a protein half life of about 70 minutes and $\delta = 0.1 \text{min}^{-1}$ corresponding to a mRNA half life of about 7 minutes.

which we employ in systems (7.4)–(7.5) by writing

$$X = X_{\text{tot}} - [X:D_1] - [X:D_2:A], \quad Y = Y_{\text{tot}} - [Y:m_1] - [Y:m_2].$$

Simulation results are shown in Figure 7.3a–7.3d, in which we consider cells growing at high rate. In the simulations, we have chosen $X_{\text{tot}} = 1 \mu\text{M}$ to account for the fact that the total amount of RNA polymerase in wild type cells at fast division rate is given by about $10 \mu\text{M}$ of which only $1 \mu\text{M}$ is free, while the rest is bound to the endogenous DNA. Since in the simulations we did not account for endogenous DNA, we assumed that only $1 \mu\text{M}$ is available in total to the two exogenous modules. A similar reasoning was employed to set $Y_{\text{tot}} = 10 \mu\text{M}$. Specifically, in exponential growth, we have about $34 \mu\text{M}$ of total ribosomes concentration, but only about 30% of this is free, resulting in about $10 \mu\text{M}$ concentration of ribosomes available to the exogenous modules (<http://bionumbers.hms.harvard.edu>).

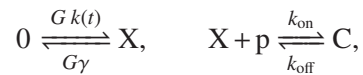
Figure 7.3a illustrates that as the activator concentration A increases, there is no substantial perturbation on the free amount of RNA polymerase. However, because the resulting perturbation on the free amount of ribosomes (Figure 7.3a) is significant, the resulting decrease of P_1 is substantial.

7.3 Stochastic Effects: Design Tradeoffs in Systems with Large Gains

As we have seen in Chapter 6, a biomolecular system can be rendered insensitive to retroactivity by implementing a large input amplification gain in a negative feedback loop. However, relying on high gains, this type of design may have undesired effects in the presence of noise, as seen in a different context in Section 5.2. Here, we employ the Langevin equation introduced in Chapter 4 to analyze this problem.

Consider a system, such as the transcriptional component of Figure 6.4, in which a protein X is produced, degraded, and is an input to a downstream system, such as a transcriptional component. Here, we assume that both the production and the degradation of protein X can be tuned through a gain G , something that can be realized through the designs illustrated in Chapter 6. Hence, the production rate of X is given by a time-varying function $Gk(t)$ while the degradation rate is given by $G\gamma$.

The system can be simply modeled by the chemical equations



in which we assume that the binding sites p are in total constant amount denoted p_{tot} , so that $p + C = p_{\text{tot}}$.

We have shown in Section 6.5 that increasing the gain G is beneficial for attenuating the effects of retroactivity on the upstream component applied by the connected downstream system. However, as shown in Figure 7.4, increasing the gain G impacts the frequency content of the noise in a single realization. In particular, as G increases, the realization shows perturbations (with respect to the mean value) with higher frequency content.

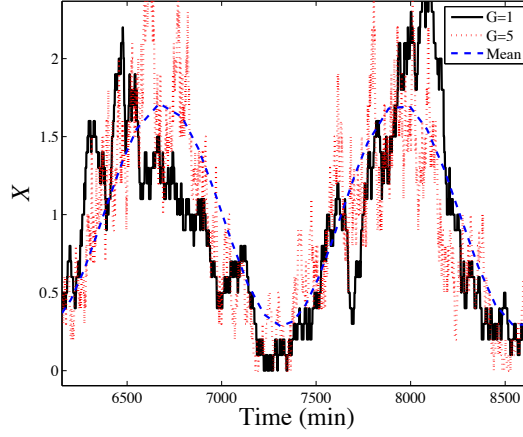


Figure 7.4: Stochastic simulations illustrating that increasing the value of G produces perturbations of higher frequency. Two realizations are shown with different values of G without load. The parameters used in the simulations are $\gamma = 0.01 \text{ min}^{-1}$ and the frequency of the input is $\omega = 0.005 \text{ rad/min}$ with input signal given by $k(t) = \gamma(1 + 0.8 \sin(\omega t)) \text{ nM/min}$. The mean of the signal is given as reference.

To study this problem, we employ the Langevin equation (Section 4.1). For our system, we obtain (assuming unit volume for simplifying the mathematical derivations):

$$\begin{aligned} \frac{dX}{dt} &= Gk(t) - G\gamma X - k_{\text{on}}(p_{\text{tot}} - C)X + k_{\text{off}}C + \sqrt{Gk(t)} \Gamma_1(t) - \sqrt{G\gamma X} \Gamma_2(t) \\ &\quad - \sqrt{k_{\text{on}}(p_{\text{tot}} - C)\bar{X}} \Gamma_3(t) + \sqrt{k_{\text{off}}C} \Gamma_4(t), \\ \frac{dC}{dt} &= k_{\text{on}}(p_{\text{tot}} - C)X - k_{\text{off}}C + \sqrt{k_{\text{on}}(p_{\text{tot}} - C)\bar{X}} \Gamma_3(t) - \sqrt{k_{\text{off}}C} \Gamma_4(t). \end{aligned} \quad (7.6)$$

The above system can be viewed as a nonlinear system with five inputs, $k(t)$ and $\Gamma_i(t)$ for $i = 1, 2, 3, 4$. Let $k(t) = \bar{k}$, and $\Gamma_1(t) = \Gamma_2(t) = \Gamma_3(t) = \Gamma_4(t) = 0$ be constant inputs and let \bar{X} and \bar{C} be the corresponding equilibrium points. Then for small amplitude signals $\tilde{k}(t) = k(t) - \bar{k}$ the linearization of the system (7.6) leads, with abuse of notation, to

$$\begin{aligned} \frac{dX}{dt} &= G\tilde{k}(t) - G\gamma X - k_{\text{on}}(p_{\text{tot}} - \bar{C})X + k_{\text{on}}\bar{X}\bar{C} + k_{\text{off}}C \\ &\quad + \sqrt{G\bar{k}} \Gamma_1(t) - \sqrt{G\gamma\bar{X}} \Gamma_2(t) + \sqrt{k_{\text{off}}\bar{C}} \Gamma_4(t) - \sqrt{k_{\text{on}}(p_{\text{tot}} - \bar{C})\bar{X}} \Gamma_3(t), \\ \frac{dC}{dt} &= k_{\text{on}}(p_{\text{tot}} - \bar{C})X - k_{\text{on}}\bar{X}\bar{C} - k_{\text{off}}C - \sqrt{k_{\text{off}}\bar{C}} \Gamma_4(t) + \sqrt{k_{\text{on}}(p_{\text{tot}} - \bar{C})\bar{X}} \Gamma_3(t). \end{aligned}$$

We can further simplify the above expressions by noting that $\gamma\bar{X} = \bar{k}$ and $k_{\text{on}}(p_{\text{tot}} - \bar{C})\bar{X} = k_{\text{off}}\bar{C}$. Also, since Γ_j are independent identical Gaussian white pro-

cesses, we can write $\Gamma_1(t) - \Gamma_2(t) = \sqrt{2}N_1(t)$ and $\Gamma_3(t) - \Gamma_4(t) = \sqrt{2}N_2(t)$, in which $N_1(t)$ and $N_2(t)$ are independent Gaussian white noise processes identical to $\Gamma_j(t)$. This simplification leads to the system

$$\begin{aligned}\frac{dX}{dt} &= G\tilde{k}(t) - G\gamma X - k_{\text{on}}(p_{\text{tot}} - \bar{C})X + k_{\text{on}}\bar{X}C + k_{\text{off}}C + \sqrt{2G\bar{k}}N_1(t) - \sqrt{2k_{\text{off}}\bar{C}}N_2(t), \\ \frac{dC}{dt} &= k_{\text{on}}(p_{\text{tot}} - \bar{C})X - k_{\text{on}}\bar{X}C - k_{\text{off}}C + \sqrt{2k_{\text{off}}\bar{C}}N_2(t).\end{aligned}\quad (7.7)$$

This is a system with three inputs: the deterministic input $\tilde{k}(t)$ and two independent white noise sources $N_1(t)$ and $N_2(t)$. One can interpret N_1 as the source of the fluctuations caused by the production and degradation reactions while N_2 is the source of fluctuations caused by binding and unbinding reactions. Since the system is linear, we can analyze the different contributions of each noise source separately and independent from the signal $\tilde{k}(t)$.

We can simplify this system by taking advantage once more of the separation of time scales between protein production and degradation and the reversible binding reactions, defining a small parameter $\epsilon = \gamma/k_{\text{off}}$ and letting $K_d = k_{\text{off}}/k_{\text{on}}$. By applying singular perturbation theory, we can set $\epsilon = 0$ and obtain the reduced system on the slow time scale as performed in Section 6.3. In this system, the transfer function from N_1 to X is given by

$$H_{XN_1}(s) = \frac{\sqrt{2G\bar{k}}}{s(1 + \bar{R}) + G\gamma}, \quad \bar{R} = \frac{p_{\text{tot}}/K_d}{((\bar{k}/\gamma)/K_d + 1)^2}. \quad (7.8)$$

The zero frequency gain of this transfer function is equal to

$$H_{XN_1}(0) = \frac{\sqrt{2\bar{k}}}{\sqrt{G\gamma}}.$$

Thus, as G increases, the zero frequency gain decreases. But for large enough frequencies ω , $j\omega(1 + \bar{R}) + G\gamma \approx j\omega(1 + \bar{R})$, and the amplitude is approximately given by

$$|H_{XN_1}(j\omega)| \approx \frac{\sqrt{2\bar{k}G}}{\omega(1 + \bar{R})},$$

which is a monotonically increasing function of G . This effect is illustrated in Figure 7.5. The frequency at which the amplitude of $|H_{XN_1}(j\omega)|$ computed with $G = 1$ intersects the amplitude $|H_{XN_1}(j\omega)|$ computed with $G > 1$ is given by the expression

$$\omega_e = \frac{\gamma\sqrt{G}}{(1 + \bar{R})}.$$

Thus, when increasing the gain from 1 to $G > 1$, we reduce the noise at frequencies lower than ω_e but we increase it at frequencies larger than ω_e .

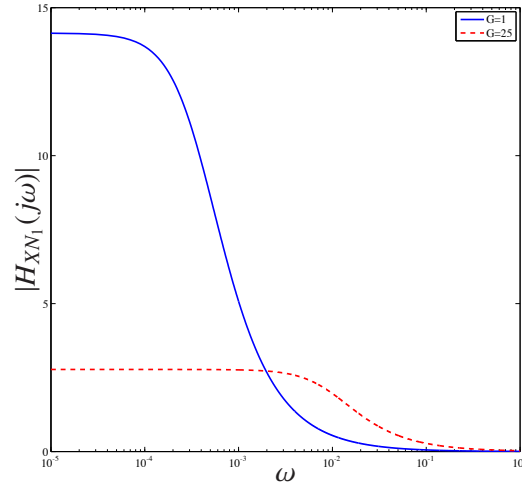


Figure 7.5: Magnitude of the transfer function $H_{XN_1}(s)$ as a function of the input frequency ω . The parameters used in this plot are $\gamma = 0.01 \text{ min}^{-1}$, $K_d = 1 \text{ nM}$, $k_{\text{off}} = 50 \text{ min}^{-1}$, $\omega = 0.005 \text{ rad/min}$, $p_{\text{tot}} = 100 \text{ nM}$. When G increases, the contribution from N_1 decreases at low frequency but it spreads to a higher range of the frequency.

While retroactivity contributes to filtering noise in the upstream system as it decreases the bandwidth of the noise transfer function (expression (7.8)), high gains contribute to increasing noise at frequencies higher than ω_e . In particular, when increasing the gain from 1 to $G > 1$ we reduce the noise in the frequency ranges below ω_e , but the noise at frequencies above ω_e increases. If we were able to indefinitely increase G , we could send G to infinity attenuating the deterministic effects of retroactivity while amplifying noise only at very high, hence not relevant, frequencies.

In practice, however, the value of G is limited. For example, in the insulation device based on phosphorylation, G is limited by the amounts of substrate and phosphatase that we can have in the system. Hence, a design tradeoff needs to be considered when designing insulation devices: placing the largest possible G attenuates retroactivity but it may increase noise in a possibly relevant frequency range.

In this chapter, we have presented some of the tradeoffs that need to be accounted for when designing biomolecular circuits in living cells and focused on the problem of competition for shared resources and on noise. Problems of resource sharing, noise, and retroactivity are encompassed in a more general problem faced when engineering biological circuits, which is referred to as “context-dependence”. That is, the functionality of a module depends on its context. Context-dependence is due to a number of different factors. These include unknown regulatory interactions

between the module and its surrounding systems; various effects that the module has on the cell network, such as metabolic burden [12] and effects on cell growth [82]; and the dependence of the module's parameters on the specific biophysical properties of the cell and its environment, including temperature and the presence of nutrients. Future biological circuit design techniques will have to address all these additional problems in order to ensure that circuits perform robustly once interacting in the cellular environment.

Exercises

7.1 Assume that both Module 1 and Module 2 considered in Section 7.1 can be activated. Extend the analytical derivations of the text to this case.

7.2 A similar derivation to what was performed in Section 7.1 can be carried if R were a repressor of Module 2. Using a one-step reaction model for gene expression, write down the reaction equations for this case and the ordinary differential equations describing the rate of change of P_1 and P_2 . Then, determine how the free concentration of RNA polymerase is affected by changes in R and how P_1 is affected by changes in R .

7.3 Consider again the case of a repressor as considered in the previous exercise. Now, consider a two-step reaction model for transcription and build a simulation model with parameter values as indicated in the text and determine the extent of coupling between Module 1 and Module 2 when the repressor is increased.

7.4 Consider the system (7.7) and calculate the transfer function from the noise source N_2 to X .

7.5 Consider the insulation device based on phosphorylation illustrated in Section 6.5. Perform stochastic simulations to investigate the tradeoff between retroactivity attenuation and noise amplification when key parameters are changed. In particular, you can perform one study in which the time-scale of the cycle changes and a different study in which the total amounts of substrate and phosphatase are changed.

