# Biomolecular Feedback Systems

Domitilla Del Vecchio
MIT

Richard M. Murray
Caltech

Classroom Copy v0.6c, July 11, 2012

# Chapter 8
## Design Tradeoffs

In this chapter, we describe a number of design tradeoffs due to the fact that the synthetic circuits interact with the host organism. We specifically focus on two issues: effects of retroactivity from synthetic circuits on the host organism and effects of biological noise on the design of insulation devices. In particular, circuits use a number of cellular resources that are shared among all circuits in the cell. Hence, they increase the loading on these resources, with possibly undesired repercussions on the functioning of the circuits themselves. Specifically, independent circuits are actually coupled through sharing common resources. We analyze the effects of this general phenomenon by illustrating it on the RNA polymerase usage. The same reasoning can be applied to any shared resource that is not in substantial excess with respect to the amounts of circuit copies placed in the cell. We also illustrate possible mechanisms to avoid this problem by employing several of the robustness tools of Chapter 3. Further, we illustrate the possible tradeoffs between retroactivity attenuation and noise amplification, due to noisy cellular environments.

### 8.1 Metabolic Burden

All biomolecular circuits use cellular resources, such as ribosomes, RNA polymerase, and ATP, that are shared among all the circuitry of the cell, whether this circuitry is synthetic or natural. As a consequence, the introduction of synthetic circuits in the cell environment is potentially perturbing the availability of these resources, leading to undesired and often unpredictable side effects on cell metabolism. In this chapter, we study the effect of the retroactivity or "back-action" from the synthetic circuits to shared resources in the cellular environment by focusing on the demand for RNA polymerase, for simplicity. The effects that we highlight are significant for any resource whose availability is not in substantial excess compared to the added demand by synthetic circuits. We will then study possible mechanisms that can be engineered to attenuate the side effects of retroactivity on shared resources, focusing on RNA polymerase as an example and employing some of the adaptation techniques outlined in Chapter 3 and Chapter 6.

In order to illustrate the problem, we consider the example system shown in Figure 8.2, in which two modules, an inducible promoter (module A) and a constitutive promoter (module B), are both present in the cellular environment. In theory, module A should respond to changes in the inducer concentration while module B,

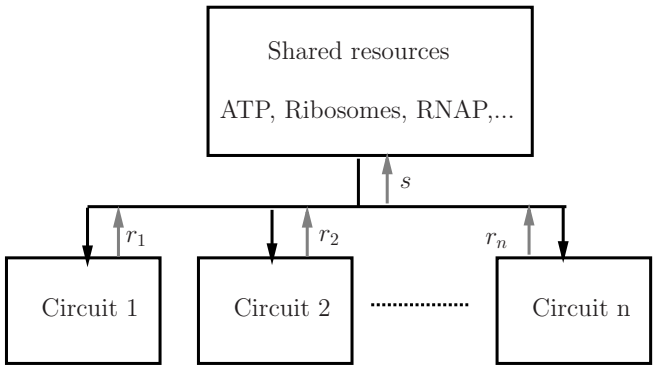



Figure 8.1: The cellular environment provides resources to synthetic circuits, such as RNA polymerase, ribosomes, ATP, proteases, etc. Circuit $i$ uses these resources and as a consequence, it has a retroactivity to the input $r_i$. The system generating shared resources has thus a retroactivity to the output $s$ that encompasses all the retroactivities applied by the circuits.

featuring a constitutive promoter, should display a constant expression level that is independent of the inducer amount. Experimental results, however, indicate that this is not the case since module B also responds to changes in inducer concentration. We illustrate how this effect can be justified mathematically by accounting for competition of shared resources needed in gene expression. To simplify the analysis, we focus on one such shared resource, the RNA polymerase.

Experimental observations indicate that increased amounts of inducer lead to decreased expression of the constitutive promoter in module B. In the case of a positive inducer, this can be qualitatively explained as follows. When the inducer amount is increased, an increased amount of active activator will be present leading to increased levels of transcription in module A. These increased levels of transcription will increase the demand for RNA polymerase and, as a consequence, smaller amounts of RNA polymerase will be free to promote transcription in mod-

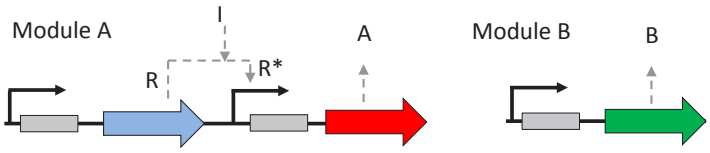



Figure 8.2: Module A has an inducible promoter that can be activated (or repressed) by transcription factor R. Such a transcription factor, when an activator, is activated by inducer I. When R is a repressor, it is repressed by the inducer I. The output of Module A is protein A. Module B has a constitutive promoter producing protein B.

ule B. Hence, module B will experience a decreased transcription rate. Similarly, the presence of larger amounts of transcript in module A will apply a larger loading on the concentration of ribosomes for the translation process. As a result, smaller amounts of ribosomes will be free to promote translation in module B. The net result is that lower expression will be observed in module B when the inducer of module A is increased. A similar reasoning can be performed in the case of a negative inducer.

The extent of this effect will depend on the availability of resources and whether they are regulated. It is known that RNA polymerase and ribosomes are regulated by the cell through negative feedback [50, 59]. This may help compensating for changes in the demand of such resources.

To mathematically demonstrate this phenomenon, we first perform a simple analytical study assuming that gene expression is a one-step process. We then perform a numerical study employing a mechanistic two-step model for gene expression.

### Analytical study using a simple model with a positive inducer

To illustrate the essence of the problem under study, 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, which, in turn, produces the corresponding protein at some constant rate. We first analyze module A, assuming module B is not present, and module B, assuming module A is not present. Then, we consider the case in which both of them are present and compare the levels of output proteins to the cases in which only one module is present.

#### *Only module A is present*

Let X denote the RNA polymerase, R the inactive activator, I the inducer, R* the active activator, that is, R bound to the activator I, p the amount of unbound promoter of module A, and A the output protein of module A. The reactions describing the system are given by (see Section 2.3):

$$\mathrm{R}+\mathrm{I}\underset{k_-}{\overset{k_+}{\rightleftharpoons}}\mathrm{R}^*,\quad \mathrm{R}^*+\mathrm{p}\underset{k_-}{\overset{k_+}{\rightleftharpoons}}\mathrm{C},\quad \mathrm{C}+\mathrm{X}\underset{k'_-}{\overset{k'_+}{\rightleftharpoons}}\mathrm{C}^*,\quad \mathrm{C}^*\xrightarrow{k}\mathrm{A}+\mathrm{C}+\mathrm{X},\quad \mathrm{A}\xrightarrow{\delta}\emptyset, \tag{8.1}$$

in which C is the complex promoter-activator and C* is the transcriptionally active complex promoter:activator:RNA polymerase. In addition, we assume that the total amount of X is conserved and denote such a total amount by $X_{\mathrm{tot}}$. Further, we assume that the total amount of promoter p is conserved and denote such a total amount by $p_{\mathrm{tot}}$. Let $R_{\mathrm{tot}} := R + R^*$ denote the total amount of transcription factor. We are interested in determining the steady state levels of $X$ and of $A$ as a function

of the inducer amounts $I$. The steady state values satisfy

$$R^* = \frac{R_{\text{tot}} I}{K_{\text{d}} + I} \qquad \text{with} \qquad K_{\text{d}} = k_-/k_+, \tag{8.2}$$

$$C = \frac{R^* p}{K_{\text{d}}} \qquad \text{with} \qquad K_{\text{d}} = k_-/k_+, \tag{8.3}$$

$$C^* = \frac{CX}{K'_{\text{d}}} \qquad \text{with} \qquad K'_{\text{d}} = k'_-/k'_+. \tag{8.4}$$

Combining these along with the conservation law $C + C^* + p = p_{\text{tot}}$ leads to

$$p = \frac{p_{\text{tot}}}{R^*/K_{\text{d}} + R^* X/(K'_{\text{d}} K_{\text{d}}) + 1},$$

in which, to simplify the derivations, we assume that

$$\frac{R^*}{K_{\text{d}}}\left(1 + \frac{X}{K'_{\text{d}}}\right) \ll 1,$$

which, in turn, is satisfied if the amount of activator $I$ is sufficiently small or if the total amount of protein $R$ is small. As a consequence, we assume in the remainder of this section that $p \approx p_{\text{tot}}$. Employing the conservation law for $X$, that is, $X_{\text{tot}} = X + C^*$, we finally obtain that

$$X = \frac{X_{\text{tot}}}{1 + \frac{p_{\text{tot}} R^*}{K'_{\text{d}} K_{\text{d}}}} = \frac{X_{\text{tot}}}{1 + p_{\text{tot}} R_{\text{tot}} I/(K'_{\text{d}} K_{\text{d}}(K_{\text{d}} + I))},$$

as a consequence, as the positive inducer concentration $I$ is increased, the amount of free RNA polymerase ($X$) decreases (see Figure 8.3). Also, since $Y = (k/\delta)C^*$, we have that

$$A = \frac{k}{\delta}\left(\frac{\frac{p_{\text{tot}} X_{\text{tot}} R_{\text{tot}} I}{K_{\text{d}} + I}}{K'_{\text{d}} + K_{\text{d}} + p_{\text{tot}} \frac{R_{\text{tot}} X_{\text{tot}} I}{K_{\text{d}} + I}}\right),$$

which increases with $I$ as expected.

### *Only module B is present*

When only module B is present, since its promoter is constitutive, it will display a constant expression level for any fixed $X_{\text{tot}}$. Denoting $q$ the amounts of unbound promoter in Module B, we have the reactions

$$\text{q} + \text{X} \underset{k'_-}{\overset{k'_+}{\rightleftharpoons}} C, \qquad C \xrightarrow{k} \text{B} + C, \qquad \text{B} \xrightarrow{\delta} \emptyset, \tag{8.5}$$

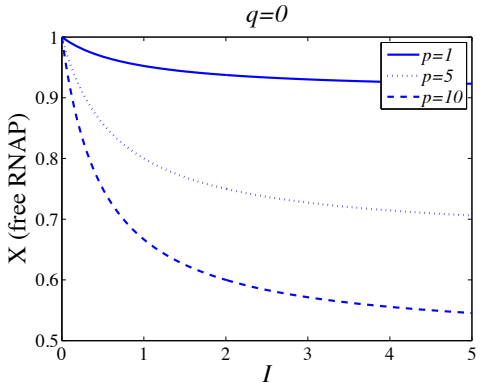



Figure 8.3: Plots showing that when the inducer level $I$ is changed, the amount of free RNA polymerase $X$ is also changed. The larger the amounts of promoter $p$, the larger the effect of the inducer on the free available RNA polymerase. For the simulation, we chose all the dissociation constants equal to one, $R_{\text{tot}} = 0.1$, $X_{\text{tot}} = 1$, and $k = 0.01\text{nMmin}^{-1}$, and $\delta = 0.01\text{min}^{-1}$. All the concentrations are in nM.

with conservation law for $X$ given by $X_{\text{tot}} = X + C$. The steady state values satisfy

$$C = \frac{Xq}{K_{\text{d}}'}, \qquad K_{\text{d}}' = k_-'/k_+', \qquad B = \frac{kC}{\delta}.$$

These relations along with the conservation law for $X$ lead to

$$X = \frac{X_{\text{tot}}}{1 + \frac{q}{K_{\text{d}}'}} \text{ and } B = \frac{k}{\delta}\left(\frac{X_{\text{tot}} q}{K_{\text{d}}' + q}\right),$$

which increases with $X_{\text{tot}}$ and $q$ as expected. Note that here, for simplifying the derivations, we have not used the conservation law $q_{\text{tot}} = q + C$. The reader can verify that the same result would hold accounting for the conservation law (see Exercises).

### *Both modules A and B are present*

When both modules are present, the set of reactions describing the system is just the union of the set of reactions describing the two modules, that is, equations (8.1) and equations (8.5). The steady state values also still satisfy the same relations as before. The only difference is the conservation law for $X$, which is now given by

$$X + C^* + C = X_{\text{tot}}.$$

Employing this conservation law along with the steady state relations gives

$$X = \frac{X_{\text{tot}}}{1 + (R^* p_{\text{tot}})/(K_{\text{d}}' K_{\text{d}}) + q/K_{\text{d}}'},$$

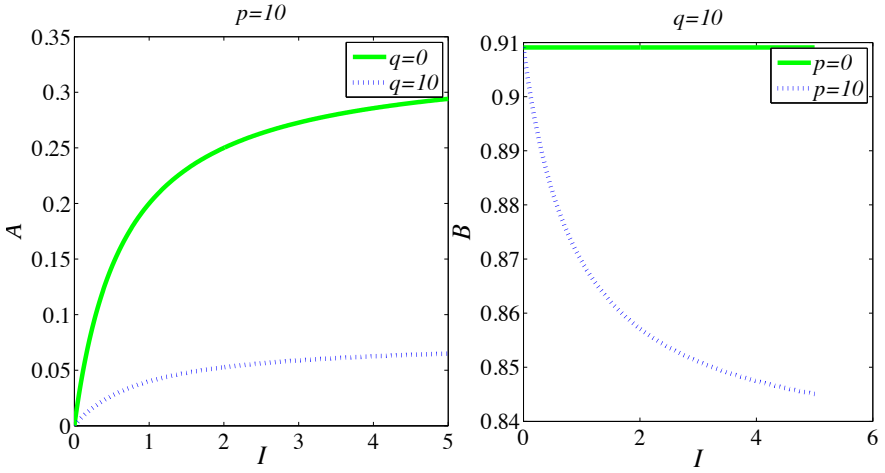



Figure 8.4: (Left) Effect on the expression of A when module B is added to the system: the expression level of A changes, but it maintains its response to the inducer. (Right) Effect on the expression of B when Module A is added to the system. When Module A is absent, the expression of B does not respond to inducer changes. By contrast, when Module A is present, the expression of B responds to inducer changes. For the simulation, we chose all the dissociation constants equal to one, $R_{\text{tot}} = 0.1$, $X_{\text{tot}} = 1$, and $k = 0.01\text{nMmin}^{-1}$, and $\delta = 0.01\text{min}^{-1}$. All the concentrations are in nM.

$$A = \frac{k}{\delta}\left(\frac{\frac{p_{\text{tot}}X_{\text{tot}}R_{\text{tot}}I}{K_{\text{d}}+I}}{K'_{\text{d}} + K_{\text{d}} + p_{\text{tot}}\frac{R_{\text{tot}}X_{\text{tot}}A}{K_{\text{d}}+I} + K'_{\text{d}}K_{\text{d}}\frac{q}{K_{\text{d}}'}}\right) \text{ and } B = \frac{k}{\delta}\left(\frac{X_{\text{tot}}q}{K_{\text{d}}'\left(1+\frac{p_{\text{tot}}}{K'_{\text{d}}K_{\text{d}}}\frac{R_{\text{tot}}I}{K_{\text{d}}+I}\right)+q}\right).$$

From this expression, it is clear that

(1) due to the presence of module B, the amounts of output protein Y of module A is lower for any given value of the inducer $I$;

(2) module B also responds to the inducer of module A. Specifically, the amounts of output protein $Z$ decreases when the amounts of inducer $I$ is increased.

These conclusions are summarized in Figure 8.4, which shows the steady state values of $B$ and $A$ when the inducer amount $I$ is changed as compared to the case in which the modules were not both present in the system.

As an exercise, the reader can verify that a similar result would hold for the case of a negative inducer (see Exercises).

### Estimate of the effects of adding external plasmids on the availability of RNAP

In the previous section, we illustrated qualitatively the mechanism by which the change in the availability of a shared resouce, due to the addition of synthetic cir-

cuits, can cause unexpected 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 amonts, and the fraction of resource that is used already by natural circuits. In fact, as we will see in the following sections, if the resource has a very large number of clients already, i.e., a very large fan-out, its changes due to the addition of more clients will be smaller. Hence, it is important to account for these in the calculation as follows.

In *E. coli*, the amount of RNA polymerase and its partitioning mainly depends on the growth rate of the cell [14]: with 0.6 doublings/hour there are only 1500 molecules/ cell, but 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 lowest considered in [14]. Therefore, a reasonable estimate is that the total number of RNA polymerase is about 1000. Since the fraction of immature core enzyme at low growth rate is only a few percent [15], we assume that the total number of functional RNA polymerase is 1000 per cell, that is, $X_{\text{tot}} = 1000\text{nM}$. Based on the data presented in [15], a reasonable partitioning of RNA polymerase is the following:

active core enzyme: 15% (150 molecules/cell or $X_{\text{a}} = 150\text{nM}$),

promoter-bound holoenzyme: 15% (150 molecules/cell or $X_{\text{p}} = 150\text{nM}$),

free holoenzyme: 5% (50 molecules/cell or $X_{\text{f}} = 50\text{nM}$),

inactive DNA-bound core: 65% (650 molecules/cell $X_{\text{i}} = 650\text{nM}$).

There are about 1000 genes expressed in exponential growth phase [48], hence the number of binding sites for X is about 1000, or $p_{\text{tot}} = 1000\text{nM}$, and we assume that all the 150 promoter-bound holoenzymes are bound to these promoters. The binding reaction is of the form

$$\text{p} + \text{X}_{\text{f}} \underset{d}{\overset{a}{\rightleftharpoons}} \text{C}_1$$

where p is the RNA polymerase-free promoter and $\text{C}_1$ is the RNA polymerase:promoter complex. Consequently, we have $p_{\text{tot}} = p + C_1$. Since only one RNA polymerase can bind to any given promoter, at the equilibrium we have $C_1 = X_{\text{p}} = 150\text{nM}$ and $p = p_{\text{tot}} - C_1 = p_{\text{tot}} - X_{\text{p}} = 850\text{nM}$. With dissociation constant $K_{\text{d}} = \frac{d}{a}$ the equilibrium is given by $0 = X_{\text{f}} p - K_{\text{d}} C_1$, hence we have that

$$K_{\text{d}} = \frac{p}{C_1} X_{\text{f}} \approx 300\text{nM},$$

which is interpreted as an "effective" dissociation constant. This is in the range $1 - 1000\text{nM}$ suggested by [40] for specific binding of RNA polymerase to DNA.

Therefore, we are going to model the binding of RNA polymerase to the promoters of the chromosome of *E. coli* in exponential phase as one promoter with concentration $p_{\text{tot}}$ and effective dissociation constant $K_{\text{d}}$.

Furthermore, we have to take into account the rather significant amount of RNA polymerase bound to the DNA other than at the promoter region ($X_{\text{a}} + X_{\text{i}} = 800\text{nM}$). To do so, we assume that the fraction $m = X_{\text{a}} + X_{\text{i}} + X_{\text{p}}/X_{\text{p}}$ is approximately constant at the equilibrium.

Now, we can consider the addition of synthetic plasmids. Specifically, we consider the plasmid pSB1AK3 (copy number $100 - 300$) with one copy of a gene under the control of a constitutive promoter. The binding of RNA polymerase to the constitutive promoter is modeled by

$$\text{q} + \text{X}_{\text{f}} \underset{d'}{\overset{a'}{\rightleftharpoons}} \text{C}_2$$

where q is the RNA polymerase-free promoter and $\text{C}_2$ is the RNA polymerase: promoter complex. Consequently, we have $q_{\text{tot}} = q + C_2$. The dissociation constant is given by $K'_{\text{d}} = \frac{d'}{a'}$. The total concentration of promoters $q_{\text{tot}}$ can be determined by considering the copy number of the plasmid, which is $100 - 300$ plasmids/cell, so that we set $q_{\text{tot}} \approx 200\text{nM}$. At the equilibrium, we have

$$C_2 = q_{\text{tot}} \frac{X_{\text{f}}}{K'_{\text{d}} + X_{\text{f}}}.$$

We also have

$$C_1 = p_{\text{tot}} \frac{X_{\text{f}}}{K_{\text{d}} + X_{\text{f}}}.$$

The conservation law for RNA polymerase must be now considered in order to determine the equilibrium concentrations:

$$X_{\text{f}} + m\, C_1 + C_2 = X_{\text{tot}}. \tag{8.6}$$

Here, we did not account for RNA polymerase molecules paused, queuing and actively transcribing on the plasmid, moreover, we also neglected the resistance genes on the pSB1AK3 plasmid. Hence, we are underestimating the effect of load presented by the plasmid.

Solving equation (8.6) for the free RNA polymerase amount $X_{\text{f}}$ gives the following results. These results depend on the ratio between the effective dissociation constant $K_{\text{d}}$ and the dissociation constant $K'_{\text{d}}$ of RNA polymerase from the plasmid promoter:

> $K'_{\text{d}} = 0.1 K_{\text{d}}$ (RNA polymerase binds better to the plasmid promoter) results in $X_{\text{f}} = 21\text{nM}$, $C_1 = 69\text{nM}$ and $C_2 = 85\text{nM}$. Hence, the concentration of free RNA polymerase decreases by about 60%;

$K'_{\rm d} = K_{\rm d}$ (binding is the same) results in $X_{\rm f} = 41$nM, $C_1 = 126$nM and $C_2 = 25$nM. Hence, the concentration of free RNA polymerase decreases by about 20%;

$K'_{\rm d} = 10K_{\rm d}$ (RNA polymerase binds better to the chromosome) results in $X_{\rm f} = 49$nM, $C_1 = 147$nM and $C_1^B = 3$nM. Hence, the concentration of free RNA polymerase decreases by about 2%.

We conclude that if the promoter on the synthetic plasmids has a dissociation constant for RNA polymerase that is smaller than the effective one calculated above, the perturbation on the available free RNA polymerase can be significant.

### Numerical study using a mechanistic model with a positive inducer

In this section, we introduce a mechanistic model of the system in Figure 8.2, in which we consider both the RNA polymerase and the ribosome usage, no approximating assumption are made, and biochemical parameters are chosen from the literature. Specifically, for inducer I we consider AHL, transcription factor R is LuxR, the output of module A is RFP, and the output of module B is GFP. We denote the concentration or RNA polymerase by $X_{\rm rnap}$ and the concentration of ribosomes by $X_{\rm rb}$. We denote by $m_A$ and $A$ the concentrations of the mRNA of RFP and of RFP protein, respectively, while we denote by $m_B$ and $B$ the concentrations of the mRNA of GFP and of GFP protein, respectively. Denoting by $R^*$ the concentration of the complex of LuxR with AHL (equal to $LuxR_{\rm tot}I/(K_{\rm d}+I)$ with $LuxR_{\rm tot}$ the total amount of LuxR), we have (see Section 2.3) the following reactions for module A transcription

$$\mathrm{R}^* + \mathrm{p}_1 \underset{k_{x2}}{\overset{k_{x1}}{\rightleftharpoons}} \mathrm{C}_1, \quad \mathrm{C}_1 + \mathrm{X}_{\rm rnap} \underset{k_{x4}}{\overset{k_{x3}}{\rightleftharpoons}} \mathrm{TC}_1, \quad \mathrm{TC}_1 \xrightarrow{k_1} \mathrm{m}_\mathrm{A} + \mathrm{X}_{\rm rnap} + \mathrm{C}_1, \quad \mathrm{m}_\mathrm{A} \xrightarrow{\delta_1} \emptyset$$

and for module A translation

$$\mathrm{m}_\mathrm{A} + \mathrm{X}_{\rm rb} \underset{k_{r2}}{\overset{k_{r1}}{\rightleftharpoons}} \mathrm{RC}_1, \quad \mathrm{RC}_1 \xrightarrow{k_{r3}} \mathrm{A} + \mathrm{X}_{\rm rb} + \mathrm{m}_\mathrm{A}, \quad \mathrm{A} \xrightarrow{\delta_2} \emptyset,$$

in which $\mathrm{C}_1$ is the complex of active transcription factor with the promoter controlling A, $\mathrm{TC}_1$ is the complex of $\mathrm{C}_1$ with $\mathrm{X}_{rnap}$, $\delta_1$ is the decay rate of mRNA, $\delta_2$ is the decay rate of protein, $\mathrm{RC}_1$ is the complex of $\mathrm{X}_{rb}$ with the mRNA ribosome binding site, $k_1$ is the rate of transcription, and $k_{r3}$ is the rate of translation. The

resulting system of differential equations is given by

$$\begin{aligned}
\frac{d\,C_1}{dt} &= k_{x1}\,R^*\,p_1 - k_{x2}\,C_1 - k_{x3}\,X_{\text{rnap}}\,C_1 + (k_{x4}+k_1)TC_1 \\
\frac{d\,TC_1}{dt} &= k_{x3}\,X_{\text{rnap}}\,C_1 - (k_{x4}+k_1)\,TC_1 \\
\frac{d\,m_A}{dt} &= k_1\,TC_1 - k_{r1}\,X_{\text{rb}}\,m_A + k_{r2}\,RC_1 - \delta_1\,m_A + k_{r3}\,RC_1 \\
\frac{d\,RC_1}{dt} &= k_{r1}\,X_{\text{rb}}\,m_A - (k_{r2}+k_{r3})\,RC_1 \\
\frac{d\,A}{dt} &= k_{r3}\,RC_1 - \delta_2\,A,
\end{aligned} \tag{8.7}$$

in which, we have that $p_1 = p_{1,tot} - C_1 - TC_1$ by the conservation law of DNA in module A.

For module B, we have the following reactions for transcription

$$\mathrm{X_{rnap} + p_2 \underset{k_{x7}}{\overset{k_{x6}}{\rightleftharpoons}} TC_2}, \quad \mathrm{TC_2 \xrightarrow{k_2} m_B + X_{rnap} + p_2}, \quad \mathrm{m_B \xrightarrow{\delta_1} \emptyset}$$

and the following reactions for translation

$$\mathrm{m_B + X_{rb} \underset{k_{r5}}{\overset{k_{r4}}{\rightleftharpoons}} RC_2}, \quad \mathrm{RC_2 \xrightarrow{k_{r6}} B + X_{rb} + m_B}, \quad \mathrm{B \xrightarrow{\delta_2} \emptyset},$$

in which $\mathrm{TC_2}$ is the transcriptionally active complex of promoter with RNA polymerase, $k_2$ is the transcription rate, $\mathrm{RC_2}$ is the complex of ribosome binding site with the ribosome, and $k_{r6}$ is the translation rate. The resulting system of differential equations is given by

$$\begin{aligned}
\frac{d\,TC_2}{dt} &= k_{x6}\,X_{\text{rnap}}\,p_2 - (k_{x7}+k_2)\,TC_2 \\
\frac{d\,m_B}{dt} &= k_2\,TC_2 - k_{r4}\,X_{\text{rb}}\,m_B + k_{r5}\,RC_2 - \delta_1\,m_B + k_{r6}\,RC_2 \\
\frac{d\,RC_2}{dt} &= k_{r4}\,X_{\text{rb}}\,m_B - (k_{r5}+k_{r6})\,RC_2 \\
\frac{d\,B}{dt} &= k_{r6}\,RC_2 - \delta_2\,B,
\end{aligned} \tag{8.8}$$

in which $p_2 = p_{2,tot} - TC_2$ from the conservation law of DNA in module B.

We consider two cases: (case 1) either Module A or Module B is present in the cellular environment and (case 2) Module A and Module B are both present in the cellular environment. In either case, the differential equations for the two modules are the same. The difference between the two cases is in the conservation law for the shared resources $\mathrm{X_{rnap}}$ and $\mathrm{X_{rb}}$. Specifically, in case 1 we have that

$$\text{Module A:} \qquad X_{\text{rnap,tot}} = X_{\text{rnap}} + TC_1, \quad X_{\text{rb,tot}} = X_{\text{rb}} + RC_1$$

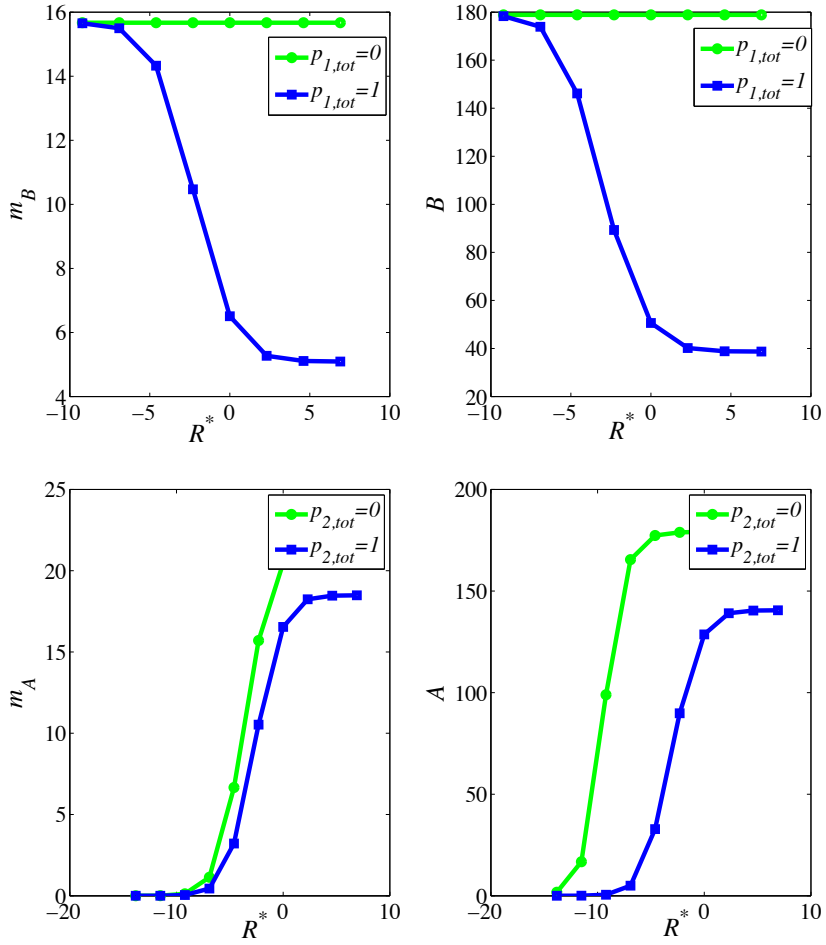



Figure 8.5: (Up: Module B) Effect on the mRNA and protein steady state response to the stimulus $R^*$ in the presence of module A ($p_{1,tot} \neq 0$). In the presence of module A also module B responds to the stimulus of module A. (Down: Module A) Effect on the mRNA and protein steady state response to the stimulus $R^*$ in the presence of module B ($p_{2,tot} \neq 0$). In the presence of module B, there is an increase of the apparent $Km$ of the steady state characteristic (right-side plot). The values of the parameters for the numerical simulation of the mechanistic model are given by $k_{x1} = 1$ ([9]), $k_{x2} = 1$ ([9]), $k_{x3} = 100$, $k_{x4} = 1$, $k_1 = 1$ ([9]), $k_{x6} = 2000$, $k_{x7} = 1$, $k_2 = 1$, $k_{r1} = 100$, $k_{r2} = 1$, $k_{r3} = 9$, $k_{r4} = 100$, $k_{r5} = 1$, $k_{r6} = 9$, $\delta_1 = 0.04$ ([9]), $\delta_2 = 0.05$. RNA polymerase and ribosomes total concentrations have been assumed to be equal to one. Concentration units are in nM.

and

$$\text{Module B:} \qquad X_{\text{rnap,tot}} = X_{\text{rnap}} + TC_2, \quad X_{\text{rb,tot}} = X_{\text{rb}} + RC_2,$$

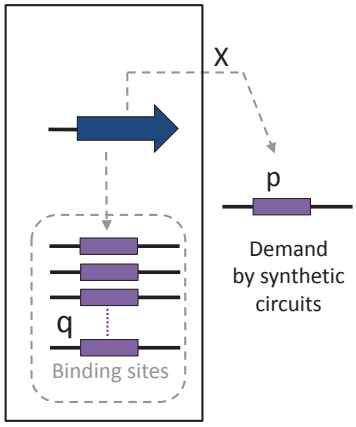



Figure 8.6: Effectively, binding sites q introduce a reservoir for X, so that more X is freed up from sites q when the demand increases.

in case 2, we have that

$$X_{\text{rnap,tot}} = X_{\text{rnap}} + TC_1 + TC_2, \quad X_{\text{rb,tot}} = X_{\text{rb}} + RC_1 + RC_2,$$

which leads to a coupling between the model of Module A and that of Module B.

The results are shown in Figure 8.5. The presence of module A, causes module B to also respond to the inducer of module A. The presence of module B also affects the response of module A to its inducer by decreasing the steady state values of the output and by increasing the value of half maximal induction.

### Engineering adaptation to changing demands of cellular resources

We have seen that competition for shared resources leads unwanted crosstalk between unconnected circuits. In order to prevent this, there are two main techniques that can be employed. The first approach is to make the amount of free X robust to changes in the circuits that use it. That is, one would like to maintain a roughly constant X when circuits are added or removed from the cell environment. The second approach is to allow potentially large excursions of X when circuits are added or removed from the cell environment, but engineer circuits so that their function is unaltered by changes in X, that is, its function adapts to changes in X.

*Engineering X robustness to changing demands by large fan-out*

In wild type *E. coli* cells, only 2% of the total amount of functional RNA polymerase is unbound (free) and only about 20% of the total amount of ribosomes is unbound [86, 14]. This suggests that in a natural system the RNA polymerases and ribosomes have a large number of sites, also called fan-out, to which they bind so that only a small fraction is free to be employed in synthetic circuits. Here, we illustrate that such a large number of sites contribute to the robustness of the concentration of these resources to changes in the demand.

Assume the sites to which X binds are denoted q and assume that we add some more sites, denoted p, belonging, for example to synthetic circuits. The introduction of sites p will increase the demand of X and will tend to decrease the amount of free X. However, such a decrease can be compensated by having the X bound to sites q unbind and increase the amount of free X. In this sense, sites q can be thought of a reservoir of X, from which X is released when needed. If this reservoir is much larger than the perturbation $p$, we should expect that $X$ will stay about constant after the addition of sites p.

To mathematically justify this reasoning, assume that X is in total amount $X_{\text{tot}}$ and let $p \ll q$ (Figure 8.6). Let $C_0$ denote the concentration of the complex of X with sites q and $C_1$ the concentration of the complex of X with sites p. The quasi-steady state approximation of these binding reactions gives $C_0 = (q/K_{\text{d}})X$ and $C_1 = (p/K_{\text{d}})X$, in which $K_{\text{d}}$ is the dissociation constant of X with the sites. The conservation law for X gives the free amount of X as

$$X = \frac{X_{\text{tot}}}{1 + (p/K_{\text{d}}) + (q/K_{\text{d}})},$$

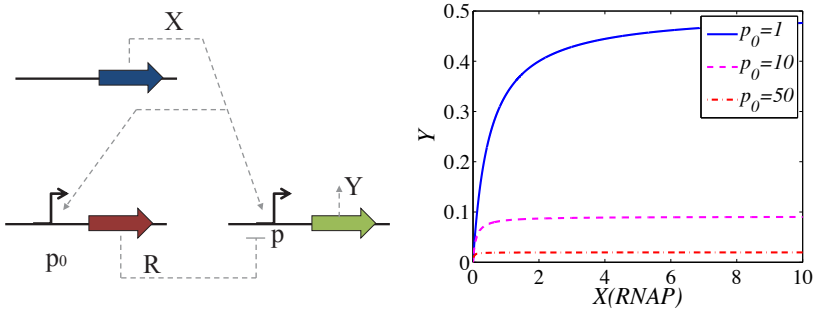
due to the addition of binding sites p. The resulting perturbation is given by

$$\Delta X = X_{\text{tot}} \frac{p/K_{\text{d}}}{(1 + (q/K_{\text{d}}))(1 + (p/K_{\text{d}}) + (q/K_{\text{d}}))},$$

from which, it is clear that as $q$ increases, the perturbation $\Delta X$ goes to zero. Since $X$ also goes to zero as $q$ increases, it is more meaningful to determine the percentage variation of X, which is given by

$$\frac{\Delta X}{X} = \frac{p/K_{\text{d}}}{(1 + (q/K_{\text{d}}) + (p/K_{\text{d}}))},$$

which can be made arbitrarily small by increasing $q$. Hence, sufficiently large values of $q$ lead to low sensitivity of the change in X when additional circuits are added or removed from the cell. As a consequence, the induced perturbation on the circuits in the cell can be reduced by increasing $q$.




(a) (b)

Figure 8.7: The output protein expression $Y$ does not sensibly depend on the amounts of available RNA polymerase (X) for sufficiently high values of $p_0$.

### *Engineering adaptation in circuits to changes of X*

We have seen in Section 3.2 that incoherent feedforward loops can engineer adaptation to changes in their input. Here, we show how this mechanism can be employed in order to make the expression level of a protein in a synthetic circuit independent of the availability of X (RNA polymerase).

Let Y be a protein that is constitutively expressed by a promoter p in total amounts $p_{\text{tot}}$. Its expression level is going to be proportional to $X(p_{\text{tot}}/K_{\text{d}})$, so that if there is a perturbation in the free amount of X, there is going to be a proportional perturbation in the amount of Y. In order to make the expression level of Y independent of changes in X, we add to the circuit expressing Y an auxiliary circuit that constitutively expresses a repressor protein R, which competes with X for the promoter sites p, causing an effective repression of Y (Figure 8.7(a)).

The idea of this design is as follows. When the availability of X decreases, the steady state value of Y should also decrease. At the same time, the amounts of R also decreases, resulting in a consequent decrease of the repression of Y, so that the steady state value of Y should increase. If these two effects are well balanced, one should expect that no substantial change of $Y$ is observed. This is mathematically studied by considering the reactions involved in the system and their associated ODE.

Specifically, let $p_0$ denote the amounts of promoter expressing protein R, let $C'$ be the concentration of the complex of protein R with promoter p, and let $C$ be the concentration of the complex of X with promoter p. Since X and R bind to p competitively, we have that $p_{\text{tot}} = p + C + C'$. As a consequence, at the steady state, we have that

$$C = \frac{p_{\text{tot}}(X/K_{\text{d}})}{(X/K_{\text{d}}) + (R/K'_{\text{d}}) + 1},$$

in which $R = Kp_0(X/K_{\text{d}})$ with $K$ proportional to the strength of promoter $p_0$ and

$K'_{\rm d}$ the dissociation constant of R with p. Since the steady state value of $Y$ is proportional to the amount of complex C, we have that

$$Y \propto \frac{p_{\rm tot}(X/K_{\rm d})}{X(1/K_{\rm d}+(Kp_0)/(K'_{\rm d}K_{\rm d}))+1}.$$

As $p_0$ becomes larger, we have that approximately $Y \propto (p_{\rm tot}K'_{\rm d})/(Kp_0)$, which is not dependent on $X$ and, as a consequence, is not affected by changes in $X$. That is, the circuit's output $Y$ adapts to changes in its input $X$. This is also shown in Figure 8.7 (b), in which the steady state value of $Y$ becomes more and more insensitive to changes in X as $p_0$ is increased. Of course, increasing $p_0$ decreases also the steady state value of $Y$, so the amounts of promoters $p$ and $p_0$ should be chosen comparably large in such a way that a desired value of $Y$ is not too low.

## 8.2 Stochastic Effects: Design Tradeoffs between Retroactivity and Noise

[1]As we have seen in Chapter 7, a biomolecular system can be rendered insensitive to retroactivity by implementing a large input amplification gain in a negative feedback loop. However, relying on a large negative feedback, this type of design may have undesired effects as seen in a different context in Section 6.2. Also, it is not clear so far what the effect of retroactivity is on the noise content of the upstream system. Here, we employ the Langevin equation seen in Chapter 4 to answer these questions.

Consider a transcriptional system that takes a transcription factor U as an input and produces a transcription factor Z as output. The transcription rate of the gene $z$, which expresses the protein Z, is given by a time varying function $Gk(t)$ that depends on the transcription factor U. This dependency is not modeled, since it is not central to our discussion. The parameter $G$ models the input amplification gain. The degradation rate of protein Z is also assumed to be tunable and thus identified by $G\delta$. The variable gain parameter $G$ will be adjusted to improve the insulation properties.

The transcription factor Z is also an input to the downstream load through the reversible binding of Z to promoter sites p. Neglecting the Z messenger RNA dynamics, the system can be modeled by the chemical equations

$$0 \underset{G\,k(t)}{\overset{G\delta}{\rightleftharpoons}} \mathrm{Z}, \qquad \mathrm{Z}+\mathrm{p} \underset{k_{\rm off}}{\overset{k_{\rm on}}{\rightleftharpoons}} \mathrm{C}.$$

We assume that $k(t)$ and $\delta$ are of the same order and denote $K_{\rm d} = k_{\rm off}/k_{\rm on}$. We also assume that the production and decay processes are slower than binding and

[1]This section is extracted from Jayanthi and Del Vecchio CDC 2009.

unbinding reactions, that is, $k_{\text{off}} \gg G\delta$, $k_{\text{on}} \gg G\delta$ as performed before. Let the total concentration of promoter be $p_{\text{tot}}$. The deterministic ordinary differential equation model is given by

$$\begin{aligned} \frac{dZ}{dt} &= Gk(t) - G\delta Z + k_{\text{off}} Z - k_{\text{on}}(p_{\text{tot}} - C)Z, \\ \frac{dC}{dt} &= -k_{\text{off}} C + k_{\text{on}}(p_{\text{tot}} - C)Z. \end{aligned} \tag{8.9}$$

To identify by what amounts $G$ should be increased to compensate the retroactivity effect, we perform a linearized analysis of (8.9) about $k(t) = k$, and the corresponding equilibrium $Z = k/\delta$ and $C = Zp_{\text{tot}}/(Z + K_{\text{d}})$. By performing the linearized analysis as in Section 7.3, letting $z = Z - Z$ and $\tilde{k} = k - k$, we obtain

$$\frac{dz}{dt} = \frac{G}{1 + R_l}(\tilde{k}(t) - \delta z), \qquad R_l = \frac{K_{\text{d}} p_{\text{tot}}}{(k/\delta + K_{\text{d}})^2}. \tag{8.10}$$

Thus, we should choose $G \approx 1 + R_l$ to compensate for retroactivity from the load. In real systems, however, there are practical limitations on how much the gain can be increased so that retroactivity may not be completely rejected.

### Dynamic effects of retroactivity

We have shown that increasing the gain $G$ is beneficial for rejecting retroactivity to the upstream component. However, as shown in Figure 8.8, increasing the gain $G$ impacts the frequency content of the noise in a single realization. For low values of $G$, the error signal between a realization and the mean is of lower frequency when compared to a higher gain.

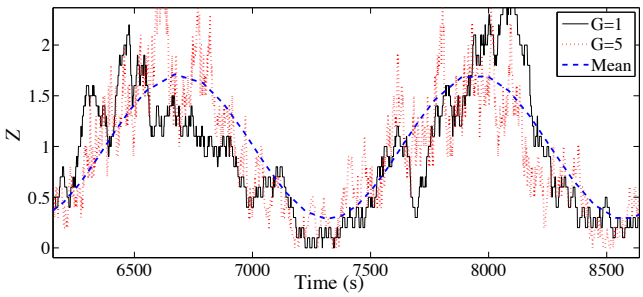



Figure 8.8: Increasing the value of $G$ produces a disturbance signal of higher frequency. Two realizations are shown with different values for $G$ without load. The parameters used in the simulations are $\delta = 0.01\text{nM}^{-1}\text{s}^{-1}$, $K_{\text{d}} = 20\text{nM}$, $k_{\text{off}} = 50\text{nM}^{-1}\text{s}^{-1}$, $\omega = 0.005\text{rad/s}$ and $\Omega = 10\text{nM}^{-1}$. The input signal used is $k(t) = \delta(1 + 0.8\sin(\omega t))\text{s}^{-1}$. The mean of the signal is given as reference.

To study this problem, we employ the Langevin equation

$$\frac{dX_i}{dt} = \sum_{j=1}^{M} \xi_{ij} a_j(\mathbf{X}(t)) + \sum_{j=1}^{M} \xi_{ij} a_j^{1/2}(\mathbf{X}(t)) N_j(t),$$

in which $N_j(t)$ are independent Gaussian white noise processes. For our system, we obtain

$$\begin{aligned}
\frac{dZ}{dt} =& Gk(t) - G\delta Z - k_{\text{on}}(p_{\text{tot}} - C)Z + k_{\text{off}}C + \sqrt{Gk(t)}\, N_1(t) - \sqrt{G\delta Z}\, N_2(t) \\
&- \sqrt{k_{\text{on}}(p_{\text{tot}} - C)Z}\, N_3(t) + \sqrt{k_{\text{off}}C}\, N_4(t), \\
\frac{dC}{dt} =& k_{\text{on}}(p_{\text{tot}} - C)Z - k_{\text{off}}C + \sqrt{k_{\text{on}}(p_{\text{tot}} - C)Z}\, N_3(t) - \sqrt{k_{\text{off}}C}\, N_4(t).
\end{aligned} \tag{8.11}$$

The above system can be viewed as a non-linear system with five inputs, $k(t)$ and $N_i(t)$ for $i = 1, 2, 3, 4$. Let $k(t) = k$, $N_1(t) = N_2(t) = N_3(t) = N_4(t) = 0$ be constant inputs and let $Z$ and $C$ be the corresponding equilibrium points. Then for small amplitude signals $\tilde{k}(t)$ the linearization of the system (8.11) leads, with abuse of notation, to

$$\begin{aligned}
\frac{dZ}{dt} =& G\tilde{k}(t) - G\delta Z - k_{\text{on}}(p_{\text{tot}} - C)Z + k_{\text{on}}ZC + k_{\text{off}}C \\
&+ \sqrt{Gk}\, N_1(t) - \sqrt{\delta Z}\, N_2(t) - \sqrt{k_{\text{off}}C}\, N_3(t) + \sqrt{k_{\text{on}}(p_{\text{tot}} - C)Z}\, N_4(t) \\
\frac{dC}{dt} =& k_{\text{on}}(p_{\text{tot}} - C)Z - k_{\text{on}}ZC - k_{\text{off}}C + \sqrt{k_{\text{off}}C}\, N_3(t) - \sqrt{k_{\text{on}}(p_{\text{tot}} - C)Z}\, N_4(t).
\end{aligned} \tag{8.12}$$

We can further simplify the above expressions by noting that $\delta Z = Gk$ and $k_{\text{on}}(p_{\text{tot}} - C)Z = k_{\text{off}}C$. Also, since $N_j$ are independent identical Gaussian white noises, we can write $N_1(t) - N_2(t) = \sqrt{2}\Gamma_1(t)$ and $N_3(t) - N_4(t) = \sqrt{2}\Gamma_2(t)$, in which $\Gamma_1(t)$ and $\Gamma_2(t)$ are independent Gaussian white noises identical to $N_j(t)$. This simplification leads to the system

$$\begin{aligned}
\frac{dZ}{dt} =& G\tilde{k}(t) - G\delta Z - k_{\text{on}}(p_{\text{tot}} - C)Z + k_{\text{on}}ZC + k_{\text{off}}C + \sqrt{2Gk}\Gamma_1(t) - \sqrt{2k_{\text{off}}C}\Gamma_2(t), \\
\frac{dC}{dt} =& k_{\text{on}}(p_{\text{tot}} - C)Z - k_{\text{on}}ZC - k_{\text{off}}C + \sqrt{2k_{\text{off}}C}\Gamma_2(t).
\end{aligned} \tag{8.13}$$

This is a system with three inputs: the deterministic input $\tilde{k}(t)$ and two independent white noise sources $\Gamma_1(t)$ and $\Gamma_2(t)$. One can interpret $\Gamma_1$ as the source of the fluctuations caused by the production and degradation reactions while $\Gamma_2$ 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 $\tilde{k}(t)$.

The transfer function from $\Gamma_1$ to $Z$ is (after setting $\delta/k_{\text{off}} = \epsilon = 0$)

$$H_{Z\Gamma_1}(s) = \frac{\sqrt{2Gk}}{s(1+R_l)+G\delta}. \tag{8.14}$$

The zero frequency gain of this transfer function is equal to $H_{Z\Gamma_1}(0) = \sqrt{2k}/\sqrt{G}\delta$. Thus, as $G$ increases, the zero frequency gain decreases. But for large enough frequencies $\omega$, $j\omega(1+R_l)+G\delta \approx j\omega(1+R_l)$, and the amplitude $|H_{Z\Gamma_1}(j\omega)| \approx \sqrt{2kG}/\omega(1+R_l)$ becomes a monotone function of $G$. This effect is illustrated in the upper plot of Figure 8.9. The frequency at which the amplitude of $|H_{Z\Gamma_1}(j\omega)|$ computed with $G = 1$ intersects the amplitude $|H_{Z\Gamma_1}(j\omega)|$ computed with $G > 1$ is given by the expression

$$\omega_e = \frac{\delta\sqrt{G}}{(1+R_l)}.$$

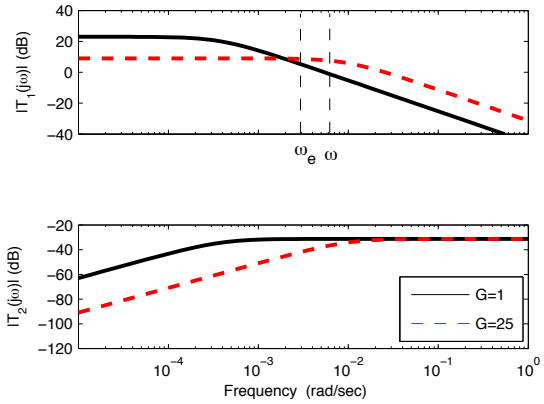



Figure 8.9: Magnitude of the transfer functions $H_{Z\Gamma_1}(s)$ and $H_{Z\Gamma_2}(s)$. The parameters used in this plot are $\delta = 0.01\text{nM}^{-1}\text{s}^{-1}$, $K_{\text{d}} = 1\text{nM}$, $k_{\text{off}} = 50\text{nM}^{-1}\text{s}^{-1}$, $\omega = 0.005\text{rad/s}$, $p_{\text{tot}} = 100\text{nM}$. When $G$ increases from 1 to $1+R_l = 25$, contribution from $\Gamma_1$ decreases but it now spreads to a higher range of the spectrum. Note that there was an increase on the noise at the frequency of interest $\omega$. Increasing $G$ reduces the contribution from $\Gamma_2$ in the low frequency range, leaving the high frequency range unaffected. Note also that the amplitude of $H_{Z\Gamma_2}$ is significantly smaller than that of $H_{Z\Gamma_1}$.

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

The transfer function from the second white noise source $\Gamma_2$ to $Z$ is given by

$$H_{Z\Gamma_1}(s) = \left[\sqrt{\epsilon}\sqrt{2\delta C}s\right] / \left[\epsilon s^2 + (\epsilon G\delta + \delta(p_{\text{tot}} - C) + \delta Z + \delta K_{\text{d}})s + G\delta(\delta Z + \delta K_{\text{d}})\right].$$

This transfer function has one zero at $s = 0$ and two poles at

$$s_{\pm} = \frac{\delta}{2\epsilon}\Big[ -\epsilon G - (p_{\text{tot}} - C) - Z + K_{\text{d}} \pm \sqrt{(\epsilon G + (p_{\text{tot}} - C) + Z + K_{\text{d}})^2 - 4\epsilon G(Z + K_{\text{d}})}\Big]. \tag{8.15}$$

When $\epsilon \to 0$, $s_- \to -\infty$ and $s_+ \to -G\delta/(1 + R_l)$. Thus, the contribution of $\Gamma_2(t)$ to Z is relevant only on the high frequency range due to the high-pass nature of the transfer function. Furthermore, increasing the gain $G$ increases the cutoff frequency given by the pole $s_+$. It is also important to note that $H_{Z\Gamma_2}(s)$ is scaled by $\sqrt{\epsilon}$, making the noise on the low-frequency caused by $H_{Z\Gamma_2}(t)$ negligible when compared to that caused by $H_{Z\Gamma_1}(t)$. The Bode plot of the transfer function $H_{Z\Gamma_2}(s)$ is shown in the lower plot of Figure 8.9.

While retroactivity contributes to filtering noise in the upstream system as it decreases the bandwidth of the noise transfer function, high gains contribute to increasing noise at frequencies higher than $\omega_e$. In particular, when increasing the gain from 1 to $G$ we reduce the noise in the frequency ranges below $\omega_e = \delta\sqrt{G}/(R_l + 1)$, but the noise at frequencies above $\omega_e$ 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 thar 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 increases noise in a possibly relevant frequency range.

## Exercises

**8.1** Consider the reactions in equation (8.5). Consider the conservation law for the sites q, that is, $q_{\text{tot}} = q + C$. Determine how the final expression for X would modify in this case.

**8.2** In the case of a negative inducer, a similar derivation can be carried if R were a repressor and R* was the inactive form of the repressor when bound to the negative inducer, denoted I. The reactions in this case are given by

$$\text{R} + \text{I} \underset{k_-}{\overset{k_+}{\rightleftharpoons}} \text{R}^*,\ \text{R} + \text{p} \underset{k_-}{\overset{k_+}{\rightleftharpoons}} \text{C},\ \text{p} + \text{X} \underset{k'_-}{\overset{k'_+}{\rightleftharpoons}} \text{C}^*,\ \text{C}^* \xrightarrow{k} \text{A} + \text{C}^*,\ \text{A} \xrightarrow{\delta} \phi, \tag{8.16}$$

in which now C is the complex of the promoter bound to the repressor, to which the RNA polymerase X cannot bind to start transcription, while $C^*$ is the complex of X with the free promoter, which is transcriptionally active. Determine the expressions for the steady state values of $X$, $A$, and $B$.

**8.3** Consider a transcriptional component expressing protein X and assume that we connect this transcriptional component to a downstream transcriptional component by having X bind to promoter sites p in the downstream system. Neglecting the mRNA dynamics gives the system of equations

$$\begin{aligned} \dot{X} &= k - \delta X - k_{\text{on}} X(p_T - C) + k_{\text{off}} C \\ \dot{C} &= k_{\text{on}} X(p_T - C) - k_{\text{off}} C, \end{aligned}$$

as we have seen in class, in which $p_T$ is the total amount of downstream system promoter binding sites and $k$ is the constant production rate. We want to show here that the steady state response of $X$ adapts to the introduction of binding sites p. To do so, we would like to show that there is a "hidden" integral feedback in this system. Address this by the following two steps:

(a) Let $u := p_T$ and find a good choice of $x$ and $y$ such that the above system takes the standard form for integral feedback seen in class:

$$\begin{aligned} \dot{x} &= f(y) \\ \dot{y} &= g(x, y, u), \end{aligned}$$

and show that the steady state value of $X$ does not depend on $p_T$.

(b) Show that $(\dot{x}, \dot{y}) \to 0$ as $t \to \infty$, so that you know that upon a constant change in $u$, $y$ returns to its original value after a transient.