Towards modularity in biological networks while avoiding retroactivity

Hari Sivakumar[1] and João P. Hespanha[1]

Abstract— **We approach the solution to the problem "are biological networks modular" using a systems theory approach. We propose a method to divide a particular family of gene regulatory networks into "modules" that are functionally isolated from each other, so that the behavior of a composite network of two or more modules can be predicted from the input-output characteristics of the individual modules. This method provides a platform for the creation of new foundational modules using which networks can be decomposed. We present our work for the deterministic case, while providing a well-known example from the literature to validate our approach.**

I. Introduction

Biological systems are inherently complex, consisting of several entities that interact in a nonlinear fashion [1]. A useful way to reduce the complexity of such systems would be to decompose them into several functionally isolated subsystems, or *modules*, which was proposed by Hartwell et al in 1999 [2]. An ideal module, as defined by Sauro, is one where the behavior of a composite network of two or more modules can be predicted from the input-output characteristics of the individual modules [3]. Our goal in this paper is to show that it is possible to divide a complex biological network into functional modules, which can then be interconnected to predict the global behavior of the network. We illustrate our points using a family of gene regulatory networks that can be divided into *single-input single-output (SISO) modules* as examples.

Multiple methods have been used to characterize gene regulatory networks, including graph-based topological models, boolean network-based models and hybrid models [4]. However, these methods focus on describing the network as a whole. It has been argued that breaking up complex biological networks into functionally isolated modules could be useful when analyzing complex bio-molecular networks, and when designing biological circuits that can be synthesized and added to cells to alter their behavior [5]. It could also be useful when attempting to tune parameters in a complex network to obtain desired properties. Moreover, dividing a complex network into modules has been shown to be useful in system identification [6] and model reduction [7].

We propose a method for dividing a complex gene regulatory network into input-output modules, in particular "activator" and "repressor" modules, which is described in more detail in Section II. At a high level, each of these modules consists of a set of chemical reactions involving a specific set of species. The inputs and outputs of these modules are defined to be "production rates" of particular chemical species, which usually depend on the concentration of proteins that are internal to a module. A typical gene regulatory network consists of many such interconnected modules. Interconnection between two modules occurs when a species is common to both modules, usually due to the concentration of one species governing the rate of expression of the other species. The modules are defined in such a way that changing parameters, or even adding and removing chemical reactions within a module, does not affect the input-output relationships of any of the other modules in the network. In Section III, we show that these modules can be interconnected in such a way that one could predict global system behavior from the input-output dynamics of each module. In Section IV, we demonstrate how this technique is useful when analyzing an actual gene regulatory network that can be divided into SISO modules.

Because gene regulatory networks are typically nonlinear systems, we characterize a module by two attributes: (i) a (typically nonlinear) *input-output static characteristic function* $g(\cdot)$ that specifies how a constant value u^* for the module's input maps to the corresponding equilibrium value $y^* = g(u^*)$ for the module's output and (ii) a u^*-*dependent transfer function* of the linearized system around the equilibrium state corresponding to the constant input u^*. Given a system of interconnected modules, the steady-state values at each input/output interconnection can be determined by combining the equations describing the input-output static characteristic functions of each module. The overall network transfer function can also be computed using standard transfer function calculus, which permits the use of a wealth of linear systems tools to study the local stability of equilibria. The input-dependent transfer functions enable a small analysis to determine the input-output behavior of the system around pre-specified equilibria. In addition, when combined with the input-output static characteristic functions, it is possible to obtain information about many properties about the non-linear system, such as whether the system has asymptotically stable equilibrium points [8] or whether its outputs converge monotonically to a steady state without oscillating, using monotone systems theory [9], [10]. Small signal analysis has also been shown to be useful when identifying parameters and pathways in complex biological networks [6], [11], [12].

Del Vecchio et. al. recently developed a concept known as *retroactivity*, which describes the change in input-output dynamics of a biological module upon its interconnection

This material is based upon work supported by the National Science Foundation under Grants No. ECCS-0835847 and EF-1137835.

[1]H. Sivakumar and J. P. Hespanha are with the Department of Electrical and Computer Engineering, University of California Santa Barbara, Santa Barbara, CA 93106, USA hari at ece.ucsb.edu,hespanha at ece.ucsb.edu

with other modules [5]. A discussion about the connection between their work and ours is described in Section V.

The use of transfer function calculus to analyze biological systems has been proposed before in a PLoS paper by Shin and Bleris [13]. In contrast to their paper, our method provides a "bottom-up" approach to map chemical reactions to self-contained modules, which is useful when extrapolating this analysis to more complicated biological networks. Their approximation of a hill equation, which describes the binding-unbinding dynamics of a protein with a promoter, by a linear function, could also be inaccurate in parameter regions where retroactivity is high.

II. Gene Transcriptional Modules

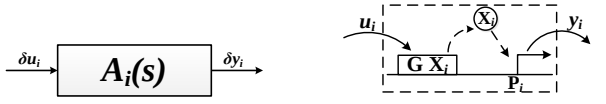
A. Activator Module

Consider a simple gene transcriptional network in isolation, where a protein X_i acts as an activator for the expression of the gene that produces the protein X_j. The system consists of the reversible reaction where X_i binds to a promoter P_i to form a promoter complex X_i:P_i. The complex X_i:P_i then governs the rate at which protein X_j is produced. For simplicity of presentation, mRNA dynamics are omitted, which is a valid assumption when the protein's life-time is much longer than the mRNA's lifetime [14]. However, we do include the mRNA dynamics in the example in Section IV for consistency with related work. This network can be described by the following set of chemical equations:

$$\phi \xrightarrow{u_i} \mathrm{X}_i \qquad \mathrm{X}_i \xrightarrow{\gamma_i} \phi$$

$$\mathrm{X}_i + \mathrm{P}_i \underset{k_{\mathrm{off}i}}{\overset{k_{\mathrm{on}i}}{\rightleftharpoons}} \mathrm{X}_i{:}\mathrm{P}_i \qquad \mathrm{X}_i{:}\mathrm{P}_i \xrightarrow{\alpha_i} \mathrm{X}_i{:}\mathrm{P}_i + \mathrm{X}_j.$$

This network can now be thought of as an activator *module*. The input to this module is the *rate of production* u_i of the activator protein X_i, which is assumed to arise from an extrinsic source, and the output from the module is the *rate of production* $y_i \coloneqq \alpha_i X_i{:}P_i$ of protein X_j. The symbols $k_{\mathrm{off}i}$ and $k_{\mathrm{on}i}$ are the rate constants for the backward and forward equations of the binding reaction, respectively. The symbol α_i is the rate constant for the translation of protein X_j, and γ_i is the rate constant for the decay of protein X_i. The symbol $P_{\mathrm{tot}i}$ is used to denote the total concentration of bound and unbound promoters, i.e. $P_{\mathrm{tot}i} = X_i{:}P_i + P_i$. Molecular variables in an upright text format (X_i, X_j, X_i:P_i, P_i, etc.) refer to the molecules themselves, while those in an italic format (X_i, X_j, $X_i{:}P_i$, P_i, etc.) refer to the *concentrations* of those molecules. A biological realization of this module is shown in Figure 1(b). The symbol G X_i is used to represent the gene that encodes protein X_i.



(a) (b)

Fig. 1. (a) Activator module (b) Biological realization of activator module

Using the standard law of mass action kinetics, which states that the rate of reaction is proportional to the product of the concentrations of the reactants, this network can be described by the following system of equations:

$$\dot{\mathbf{x_i}} = f_i(\mathbf{x_i}, u_i) \qquad y_i = \begin{bmatrix} 0 & \alpha_i \end{bmatrix} \mathbf{x_i},$$

where

$$\mathbf{x_i} \coloneqq \begin{bmatrix} X_i \\ X_i{:}P_i \end{bmatrix} \qquad f_i(\mathbf{x_i}, u_i) \coloneqq$$
$$\begin{bmatrix} u_i - \gamma_i X_i - k_{\mathrm{on}i} X_i (P_{\mathrm{tot}i} - X_i{:}P_i) + k_{\mathrm{off}i} X_i{:}P_i \\ k_{\mathrm{on}i} X_i (P_{\mathrm{tot}i} - X_i{:}P_i) - k_{\mathrm{off}i} X_i{:}P_i \end{bmatrix}. \quad (1)$$

This equation's equilibrium point for a generic input u_i^* is

$$X_i^* = \frac{1}{\gamma_i} u_i^*, \qquad X_i{:}P_i^* = \frac{k_{\mathrm{on}i} P_{\mathrm{tot}i} u_i^*}{\gamma_i k_{\mathrm{off}i} + k_{\mathrm{on}i} u_i^*}. \quad (2)$$

The input-output static characteristic function is defined to be the equilibrium point of the output from the module as a function of the value of a constant input to the module. For the activator module, this can easily be computed to be

$$y_i^* = \frac{\alpha_i k_{\mathrm{on}i} P_{\mathrm{tot}i} u_i^*}{\gamma_i k_{\mathrm{off}i} + k_{\mathrm{on}i} u_i^*} \coloneqq g_i(u_i^*).$$

Linearizing about this equilibrium, we obtain an approximation of the dynamics that is valid when the state perturbations $\delta X_i \coloneqq X_i - X_i^*$ and $\delta X_i{:}P_i = X_i{:}P_i - X_i{:}P_i^*$, and the input perturbation $\delta u_i \coloneqq u_i - u_i^*$, are sufficiently small:

$$\begin{bmatrix} \delta\dot{X}_i \\ \delta\dot{X}_i{:}P_i \end{bmatrix} = \begin{bmatrix} -\gamma_i - k_{\mathrm{on}i}(P_{\mathrm{tot}i} - X_i{:}P_i^*) & k_{\mathrm{off}i} + k_{\mathrm{on}i} X_i^* \\ k_{\mathrm{on}i}(P_{\mathrm{tot}i} - X_i{:}P_i^*) & -(k_{\mathrm{off}i} + k_{\mathrm{on}i} X_i^*) \end{bmatrix}$$
$$\begin{bmatrix} \delta X_i \\ \delta X_i{:}P_i \end{bmatrix} + \begin{bmatrix} 1 \\ 0 \end{bmatrix} \delta u_i$$
$$\delta y_i = \begin{bmatrix} 0 & \alpha_i \end{bmatrix} \begin{bmatrix} \delta X_i \\ \delta X_i{:}P_i \end{bmatrix},$$

where the output perturbation δy_i is the difference between production rate y_i of the activated protein X_j and its equilibrium value y_i^*. It can then be shown that the transfer function of this module from the input perturbation δu_i to the output perturbation δy_i is

$$A_i(s) = \frac{\alpha_i k_{\mathrm{on}i} (P_{\mathrm{tot}i} - \frac{k_{\mathrm{on}i} P_{\mathrm{tot}i} u_i^*}{\gamma_i k_{\mathrm{off}i} + k_{\mathrm{on}i} u_i^*})}{s^2 + \theta_i s + \psi_i},$$

where

$$\theta_i \coloneqq k_{\mathrm{off}i} + k_{\mathrm{on}i}(P_{\mathrm{tot}i} + \frac{1}{\gamma_i} u_i^* - \frac{k_{\mathrm{on}i} P_{\mathrm{tot}i} u_i^*}{\gamma_i k_{\mathrm{off}i} + k_{\mathrm{on}i} u_i^*}) + \gamma_i$$
$$\psi_i \coloneqq \gamma_i (k_{\mathrm{off}i} + \frac{k_{\mathrm{on}i}}{\gamma_i} u_i^*).$$

A systems theory representation of this linearized module is shown in Figure 1(a).

It is worth emphasizing that there are two properties of this module that are central to creating biological modules that are composable.

(i) The activator module incorporates the decay dynamics of protein X_i within the module itself.
(ii) The input and output of the activator module are both *production rates*, and not concentrations of proteins.

B. Repressor Module

In a similar fashion, consider a simple gene transcriptional network in isolation, where a protein X_i now acts as a repressor for the expression of the gene that produces the protein X_j. The system consists of the reversible reaction where X_i binds to a promoter P_i to form a promoter complex X_i:P_i. The promoter P_i then governs the rate at which protein X_j is produced. Again, mRNA dynamics are omitted. This can be described by the following set of chemical equations:

$$\phi \xrightarrow{u_i} \mathrm{X}_i \qquad \mathrm{X}_i \xrightarrow{\gamma_i} \phi$$
$$\mathrm{X}_i + \mathrm{P}_i \underset{k_{\mathrm{off}i}}{\overset{k_{\mathrm{on}i}}{\rightleftharpoons}} \mathrm{X}_i{:}\mathrm{P}_i \qquad \mathrm{P}_i \xrightarrow{\alpha_i} \mathrm{P}_i + \mathrm{X}_j.$$

This network can now be thought of as a repressor *module*. The input to this module is the *rate of production* u_i of the repressor protein X_i, which is assumed to arise from an extrinsic source, and the output from the module is the *rate of production* $y_i \coloneqq \alpha_i P_i$ of protein X_j.

Using the law of mass action kinetics, this repressor network can be described by the following system of equations:

$$\dot{\mathbf{x_i}} = f_i(\mathbf{x_i}, u_i) \qquad y_i = \alpha_i P_{\mathrm{tot}i} - \begin{bmatrix} 0 & \alpha_i \end{bmatrix} \mathbf{x_i},$$

where $\mathbf{x_i}$ and f_i have been defined in (1). The equation's equilibrium point, around a generic input u_i^*, is the same as was found in (2). The input-output static characteristic function is

$$y_i^* = \frac{\alpha_i k_{\mathrm{off}i} P_{\mathrm{tot}i} \gamma_i}{\gamma_i k_{\mathrm{off}i} + k_{\mathrm{on}i} u_i^*} \coloneqq h_i(u_i^*).$$

It can then be shown that the transfer function of this module from the input perturbation δu_i to the output perturbation δy_i is

$$R_i(s) = -A_i(s),$$

which means that the transfer functions of activator and repressor modules with similar parameters have the same magnitude with a π radian difference in phase. This module can be viewed in a similar fashion to the activator module shown in Figure 1.

III. Interconnections

Given a network that can be viewed as a set of cascade, parallel, and feedback interconnections, we can determine the linearized transfer function of the overall network using standard transfer function calculus techniques on the individual modules. We can also determine the equilibrium points of the overall network using the static characteristic functions of the individual modules. To make our intuition concrete, we present a few examples of basic interconnections in Tables I–IV. We present the cascade interconnection of two activators, two different parallel interconnections of two activators, and the feedback between an activator and a repressor. Due to lack of space, we leave it for the reader to construct more examples, such as longer cascades involving activators and respressors.

TABLE I

Cascade interconnection: A protein X_1 acts as an activator for the expression of the gene that produces protein X_2 and in turn, protein X_2 acts as an activator for the expression of the gene that produces protein X_3.

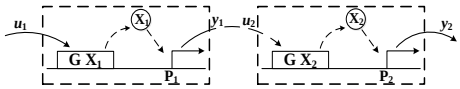
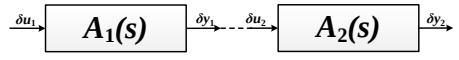
Biological Figure	
Systems Diagram	
Chemical Equations	$\phi \xrightarrow{u_1} \mathrm{X}_1$ $\mathrm{X}_1 \xrightarrow{\gamma_1} \phi \quad \mathrm{X}_2 \xrightarrow{\gamma_2} \phi$ $\mathrm{X}_1 + \mathrm{P}_1 \underset{k_{\mathrm{off1}}}{\overset{k_{\mathrm{on1}}}{\rightleftharpoons}} \mathrm{X}_1{:}\mathrm{P}_1 \quad \mathrm{X}_2 + \mathrm{P}_2 \underset{k_{\mathrm{off2}}}{\overset{k_{\mathrm{on2}}}{\rightleftharpoons}} \mathrm{X}_2{:}\mathrm{P}_2$ $\mathrm{X}_1{:}\mathrm{P}_1 \xrightarrow{\alpha_1} \mathrm{X}_1{:}\mathrm{P}_1 + \mathrm{X}_2 \quad \mathrm{X}_2{:}\mathrm{P}_2 \xrightarrow{\alpha_2} \mathrm{X}_2{:}\mathrm{P}_2 + \mathrm{X}_3$
System of Equations	$\dot{\mathbf{x}_1} = f_1(\mathbf{x}_1, u_1) \quad \dot{\mathbf{x}_2} = f_2(\mathbf{x}_2, y_1)$ $y_1 = \begin{bmatrix} 0 & \alpha_1 \end{bmatrix} \mathbf{x}_1 \quad y_2 = \begin{bmatrix} 0 & \alpha_2 \end{bmatrix} \mathbf{x}_2$ $(u_2 = y_1)$
Characteristic Functions	$y_1^* = g_1(u_1^*)$ $y_2^* = g_2(y_1^*) = g_2(g_1(u_1^*))$
Transfer Function	$\frac{\delta y_2(s)}{\delta u_1(s)} = A_2(s)A_1(s)$

TABLE II

Parallel interconnection with common input: Two distinct proteins X_1 and X_2 are produced by the same process at a common rate u_1, determined by some extrinsic source. X_1 now acts as an activator for the expression of the gene that produces protein X_3, while X_2 does the same for another protein X_4.

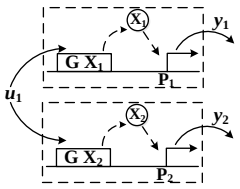
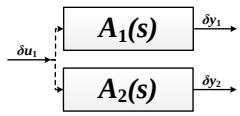
Biological Figure	
Systems Diagram	
Chemical Equations	$\phi \xrightarrow{u_1} \mathrm{X}_1 + \mathrm{X}_2$ $\mathrm{X}_1 \xrightarrow{\gamma_1} \phi \quad \mathrm{X}_2 \xrightarrow{\gamma_2} \phi$ $\mathrm{X}_1 + \mathrm{P}_1 \underset{k_{\mathrm{off1}}}{\overset{k_{\mathrm{on1}}}{\rightleftharpoons}} \mathrm{X}_1{:}\mathrm{P}_1 \quad \mathrm{X}_2 + \mathrm{P}_2 \underset{k_{\mathrm{off2}}}{\overset{k_{\mathrm{on2}}}{\rightleftharpoons}} \mathrm{X}_2{:}\mathrm{P}_2$ $\mathrm{X}_1{:}\mathrm{P}_1 \xrightarrow{\alpha_1} \mathrm{X}_1{:}\mathrm{P}_1 + \mathrm{X}_3 \quad \mathrm{X}_2{:}\mathrm{P}_2 \xrightarrow{\alpha_2} \mathrm{X}_2{:}\mathrm{P}_2 + \mathrm{X}_4$
System of Equations	$\dot{\mathbf{x}_1} = f_1(\mathbf{x}_1, u_1) \quad y_1 = \begin{bmatrix} 0 & \alpha_1 \\ 0 & \alpha_2 \end{bmatrix} \mathbf{x}_1$ $\dot{\mathbf{x}_2} = f_2(\mathbf{x}_2, u_1) \quad y_2 = \begin{bmatrix} 0 & \alpha_2 \end{bmatrix} \mathbf{x}_2$
Characteristic Functions	$y_1^* = g_1(u_1^*)$ $y_2^* = g_2(u_1^*)$
Transfer Function	$\begin{bmatrix} \frac{\delta y_1(s)}{\delta u_1(s)} \\ \frac{\delta y_2(s)}{\delta u_1(s)} \end{bmatrix} = \begin{bmatrix} A_1(s) \\ A_2(s) \end{bmatrix}$

TABLE III

PARALLEL INTERCONNECTION WITH SUMMED OUTPUTS: TWO DISTINCT PROTEINS X_1 AND X_2 ARE PRODUCED BY SOME PROCESSES AT RATES u_1 AND u_2, RESPECTIVELY, DETERMINED BY EXTRINSIC SOURCES. BOTH PROTEINS THEN INDEPENDENTLY ACT AS ACTIVATORS FOR THE EXPRESSION OF THE GENE THAT PRODUCES PROTEIN X_3.

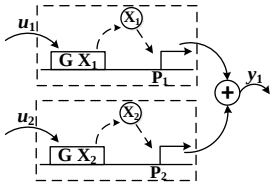
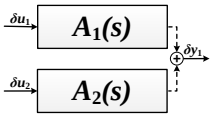
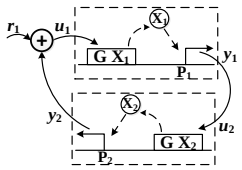
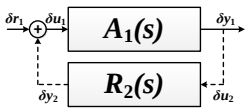
Biological Figure	
Systems Diagram	
Chemical Equations	$\phi \xrightarrow{u_1} \mathrm{X}_1 \quad \phi \xrightarrow{u_2} \mathrm{X}_2$ $\mathrm{X}_1 \xrightarrow{\gamma_1} \phi \quad \mathrm{X}_2 \xrightarrow{\gamma_2} \phi$ $\mathrm{X}_1 + \mathrm{P}_1 \underset{k_{\mathrm{off1}}}{\overset{k_{\mathrm{on1}}}{\rightleftharpoons}} \mathrm{X}_1{:}\mathrm{P}_1 \quad \mathrm{X}_2 + \mathrm{P}_2 \underset{k_{\mathrm{off2}}}{\overset{k_{\mathrm{on2}}}{\rightleftharpoons}} \mathrm{X}_2{:}\mathrm{P}_2$ $\mathrm{X}_1{:}\mathrm{P}_1 \xrightarrow{\alpha_1} \mathrm{X}_1{:}\mathrm{P}_1 + \mathrm{X}_3 \quad \mathrm{X}_2{:}\mathrm{P}_2 \xrightarrow{\alpha_2} \mathrm{X}_2{:}\mathrm{P}_2 + \mathrm{X}_3$
System of Equations	$\dot{\mathbf{x}_1} = f_1(\mathbf{x}_1, u_1)$ $\dot{\mathbf{x}_2} = f_2(\mathbf{x}_2, u_2)$ $y_1 = \begin{bmatrix}0 & \alpha_1\end{bmatrix}\mathbf{x}_1 + \begin{bmatrix}0 & \alpha_2\end{bmatrix}\mathbf{x}_2$
Characteristic Function	$y_1^* = g_1(u_1^*) + g_2(u_2^*)$
Transfer Function	$\begin{bmatrix}\frac{\delta y_1(s)}{\delta u_1(s)} \\ \frac{\delta y_1(s)}{\delta u_2(s)}\end{bmatrix} = \begin{bmatrix}A_1(s) \\ A_2(s)\end{bmatrix}$

TABLE IV

FEEDBACK INTERCONNECTION: A PROTEIN X_1 ACTS AS AN ACTIVATOR FOR THE EXPRESSION OF THE GENE THAT PRODUCES PROTEIN X_2 AND IN TURN, PROTEIN X_2 ACTS AS A REPRESSOR FOR THE EXPRESSION OF THE GENE THAT PRODUCES PROTEIN X_1. THE PRODUCTION RATE OF PROTEIN X_1 IS DETERMINED BY AN EXOGENOUS SOURCE, AS WELL AS BY THE REPRESSOR FEEDBACK. THE OUTPUT FROM THIS NETWORK WILL BE OBSERVED AT THE OUTPUT OF THE ACTIVATOR MODULE.

Biological Figure	
Systems Diagram	
Chemical Equations	$\phi \xrightarrow{r_1} \mathrm{X}_1$ $\mathrm{X}_1 \xrightarrow{\gamma_1} \phi \quad \mathrm{X}_2 \xrightarrow{\gamma_2} \phi$ $\mathrm{X}_1 + \mathrm{P}_1 \underset{k_{\mathrm{off1}}}{\overset{k_{\mathrm{on1}}}{\rightleftharpoons}} \mathrm{X}_1{:}\mathrm{P}_1 \quad \mathrm{X}_2 + \mathrm{P}_2 \underset{k_{\mathrm{off2}}}{\overset{k_{\mathrm{on2}}}{\rightleftharpoons}} \mathrm{X}_2{:}\mathrm{P}_2$ $\mathrm{X}_1{:}\mathrm{P}_1 \xrightarrow{\alpha_1} \mathrm{X}_1{:}\mathrm{P}_1 + \mathrm{X}_2 \quad \mathrm{P}_2 \xrightarrow{\alpha_2} \mathrm{P}_2 + \mathrm{X}_1$
System of Equations	$\dot{\mathbf{x}_1} = f_1(\mathbf{x}_1, r_1 + y_2)$ $y_1 = \begin{bmatrix}0 & \alpha_1\end{bmatrix}\mathbf{x}_1$ $\dot{\mathbf{x}_2} = f_2(\mathbf{x}_2, u_2) = f_2(\mathbf{x}_2, y_1)$ $y_2 = \alpha_2 P_{\mathrm{tot}2} - \begin{bmatrix}0 & \alpha_2\end{bmatrix}\mathbf{x}_2$
Characteristic Functions	$y_1^* = g_1(r_1^* + y_2^*) \quad y_2^* = h_2(y_1^*)$
Transfer Function	$\frac{\delta y_1(s)}{\delta r_1(s)} = \frac{A_1(s)}{1 - A_1(s)R_2(s)}$

IV. CASE STUDY: REPRESSILATOR

Elowitz and Liebler proposed the design of synthetic networks, to better understand the underlying design principles in a cellular network. They designed and constructed the *repressilator*, which is an oscillating network comprised of three naturally occurring transcriptional repressor systems. Depending on the values of the parameters in the network, the system could oscillate, or converge to a stable steady-state [15]. In this section, we divide this system into three modules and demonstrate how this approach can provide insight into the stability of the network, using standard methods used for analyzing interconnected SISO systems. One should note that though Elowitz and Liebler also saw the repressilator as a network of three interconnected modules, we now demonstrate how to model this formally and use systems theory to analyze the network.

In the repressilator network, the first repressor protein, LacI, inhibits the transcription of the tetR gene. The tetR protein then inhibits the transcription of the cI gene, which in turn inhibits the expression of the LacI gene. The network consists of a cascade of three repressor modules, and a feedback interconnection, as shown in Figure 2. The mRNA dynamics are *not* omitted in this section and therefore each module is made up of the following chemical reactions:

$$\phi \xrightarrow{u_i} \mathrm{mRNA}(\mathrm{X}_i) \qquad \mathrm{mRNA}(\mathrm{X}_i) \xrightarrow{\gamma_i} \phi$$
$$\mathrm{mRNA}(\mathrm{X}_i) \xrightarrow{\gamma_i \beta} \mathrm{mRNA}(\mathrm{X}_i) + \mathrm{X}_i \qquad \mathrm{X}_i \xrightarrow{\gamma_i \beta} \phi$$
$$2\mathrm{X}_i + \mathrm{P}_i \underset{k_{\mathrm{off}i}}{\overset{k_{\mathrm{on}i}}{\rightleftharpoons}} \mathrm{X}_i{:}\mathrm{X}_i{:}\mathrm{P}_i \qquad \mathrm{P}_i \xrightarrow{\alpha_i} \mathrm{P}_i + \mathrm{mRNA}(\mathrm{X}_j).$$

The symbols X_1, X_2 and X_3 denote the proteins LacI, tetR and cI, respectively. The input to this module is the *rate of production* u_i of the mRNA of the repressor protein X_i, and the output from the module is the *rate of production* $y_i = \alpha_i P_i$ of the mRNA of the next protein in the cascade. The symbols $k_{\mathrm{off}i}$ and $k_{\mathrm{on}i}$ denote the rate constants for the backward and forward equations of the binding reaction, respectively. The symbol α_i denotes the rate constant for the transcription of the mRNA of the next protein in the cascade, γ_i the rate constant for the decay of the mRNA of protein X_i, and β the ratio of the decay rate of protein X_i to the decay rate of its mRNA, when the binding-unbinding dynamics of the protein with the promoter region are assumed to be fast. The symbol $P_{\mathrm{tot}i}$ is used to describe the total concentration of bound and unbound promoters, i.e. $P_{\mathrm{tot}i} = X_i{:}X_i{:}P_i + P_i$. Each module is then represented by a set of differential equations as follows:

$$\begin{aligned}
\dot{mRNA}(X_i) &= u_i - \gamma_i mRNA(X_i) \\
\dot{X}_i &= \gamma_i \beta mRNA(X_i) - \gamma_i \beta X_i \\
&\quad - 2k_{\mathrm{on}i}(X_i)^2 P_i + 2k_{\mathrm{off}i}(P_{\mathrm{tot}i} - P_i) \\
\dot{P}_i &= -k_{\mathrm{on}i}(X_i)^2 P_i + k_{\mathrm{off}i}(P_{\mathrm{tot}i} - P_i) \\
y_i &= \alpha_i P_i.
\end{aligned}$$

Each module's input-output static characteristic function

around a generic input u_i^*, is given by

$$y_i^* = \alpha_i \frac{\gamma_i^2 k_{\text{off}i} P_{\text{tot}i}}{\gamma_i^2 k_{\text{off}i} + k_{\text{on}i}(u_i^*)^2}.$$

The linearized transfer function of each module from the perturbation at the input δu_i to the perturbation at the output δy_i,

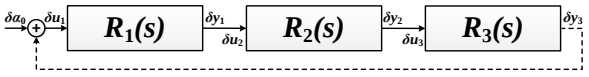
$$R_i(s) = -\frac{2\alpha_i \beta \gamma_i k_{\text{on}i}(\frac{\gamma_i k_{\text{off}i} P_{\text{tot}i} u_i^*}{\gamma_i^2 k_{\text{off}i} + k_{\text{on}i}(u_i^*)^2})}{(s+\gamma_i)(s^2 + \bar{\theta}_i s + \bar{\psi}_i)}, \qquad (3)$$

where

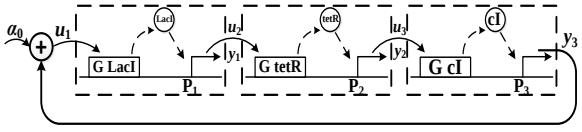
$$\bar{\theta}_i \coloneqq k_{\text{off}i} + k_{\text{on}i}((\frac{u_i^*}{\gamma_i})^2 + \frac{4\gamma_i k_{\text{off}i} P_{\text{tot}i} u_i^*}{\gamma_i^2 k_{\text{off}i} + k_{\text{on}i}(u_i^*)^2}) + \beta\gamma_i$$

$$\bar{\psi}_i \coloneqq \beta\gamma_i(k_{\text{off}i} + k_{\text{on}i}(\frac{u_i^*}{\gamma_i})^2).$$

The units and parameters chosen for our analysis are similar to those chosen in [15]. In the results that follow, time is scaled in units of the mRNA lifetime, protein concentrations are in terms of the number of repressors necessary to half-maximally repress the promoter (K_M) and mRNA concentrations are scaled by the average number of proteins produced per mRNA molecule (K_T). Elowitz and Liebler assume that the mRNA half-life is 20 minutes, K_M is 40 monomers per cell and K_T is 20 proteins per transcript. The Hill coefficient is assumed to be 2 for this analysis. The symbol α_0 corresponds to an exogenous rate of production of the mRNA of LacI, by mechanisms external to the repressilator. Like Elowitz and Liebler, we assume that the modules are each characterized by the same parameters. We then linearize the feedback loop about the input $\alpha_0^* = 0$, and hence obtain the same transfer function for each module in the network, i.e., $R_1(s) = R_2(s) = R_3(s)$. We define $R(s)$ to be the right-hand side of (3) without the negative sign, so $R(s) \coloneqq -R_1(s) = -R_2(s) = -R_3(s)$.



(a)



(b)

Fig. 2. (a) Systems level representation of repressilator network (b) Biological realization of repressilator network

Lemma 1: The number $\#CUP$ of closed-loop unstable (i.e., in the closed right-hand side plane) poles of the linearized repressilator network is equal to

$$\#CUP = 3\#OUP + \sum_{i=1}^{3} \#END[z_i],$$

where $\#OUP$ denotes the number of (open-loop) unstable poles of $R(s)$, $\#END[z]$ denotes the number of clockwise encirclements of the Nyquist plot of $R(s)$ around the point $z \in \mathbb{C}$, and $z_1 \coloneqq -1$, $z_2 \coloneqq e^{j\pi/3}$, $z_3 \coloneqq e^{-j\pi/3}$. □

Proof. To investigate stability of the repressilator, we consider the characteristic equation $F(s) = 1 - (-R(s))^3 = 1 + R(s)^3 = 0$. The number of unstable poles is thus given by the unstable solutions to the equation:

$$1 + R(s)^3 = 0 \quad \Leftrightarrow \quad \exists i \in \{1,2,3\},\ R(s) = z_i,$$

where $z_1 \coloneqq -1$, $z_2 \coloneqq e^{j\pi/3}$, $z_3 \coloneqq e^{-j\pi/3}$ are the three roots to the equation $z^3 = -1$. To count the number of unstable poles of the repressilator, we must then add the number of unstable poles of each of the three equations

$$R(s) = z_i, \quad i \in \{1,2,3\},$$

which can be done using Cauchy's argument principle by counting the number of clockwise encirclements of the point $z_i \in \mathbb{C}$ for the Nyquist contour of $R(s)$ [16]. ■

We first fix α_i, $P_{\text{tot}i}$, $k_{\text{on}i}$, $k_{\text{off}i}$ and γ_i and compute the equilibrium point of the repressilator network around $\alpha_0^* = 0$. We compute y_3^*, which can be shown to be equivalent to y_2^* and y_1^* when the parameters across each of the three modules are equal. It can easily be shown that for fixed α_i, $P_{\text{tot}i}$, $k_{\text{on}i}$, $k_{\text{off}i}$ and γ_i, the output equilibrium point of each of the repressor modules in the repressilator network linearized about $\alpha_0^* = 0$ is the same as the equilibrium point of the output from a single repressor module linearized around y_3^*. Since β does not affect the equilibrium points of the modules, we can then analyze the closed loop stability of the equilibrium points of the repressilator network linearized around $\alpha_0^* = 0$, by applying Lemma 1 to a single repressor module linearized around y_3^*, for different values of β.

Figure 3 shows the stability of the equilibrium points of the repressilator network, obtained by applying Lemma 1 to a single repressor module. It can be observed that for a given α_i, $P_{\text{tot}i}$, $k_{\text{on}i}$, $k_{\text{off}i}$ and γ_i, the equilibrium points of the repressilator network are stable for sufficiently small or sufficiently large values of β. The Nyquist plots in Figure 4 provide some intuition about this observation. It should be noted, however, that β small would imply that the mRNA dynamics are slower than, or on a similar timescale as the protein dynamics for each module, which is generally not true.

V. Retroactivity in Biochemical Networks

One could divide the chemical reactions from the activator module in Section II-A for $i = 1$ as a block with input corresponding to the production rate of X_1, that produces an output corresponding to the production rate of the complex X_1:P_1, as described by the reactions

$$\begin{aligned} \phi &\xrightarrow{u_1} \text{X}_1 & \qquad \phi &\xrightarrow{\bar{y}_2} \text{X}_1 + \text{P}_1 \\ \text{X}_1 &\xrightarrow{\gamma_1} \phi & \qquad \text{X}_1 + \text{P}_1 &\xrightarrow{k_{\text{on}1}} \text{X}_1\text{:P}_1, \end{aligned} \qquad (4)$$

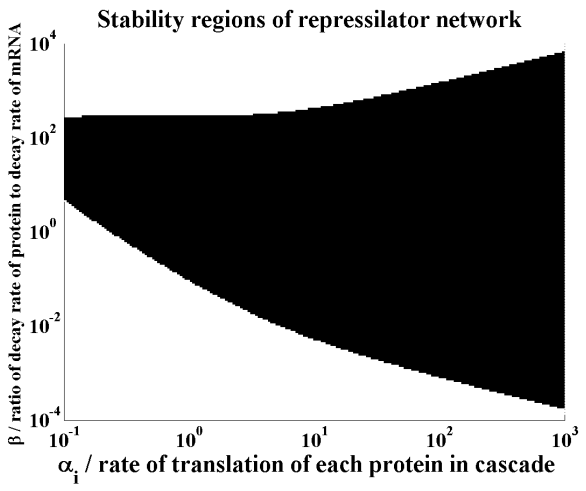
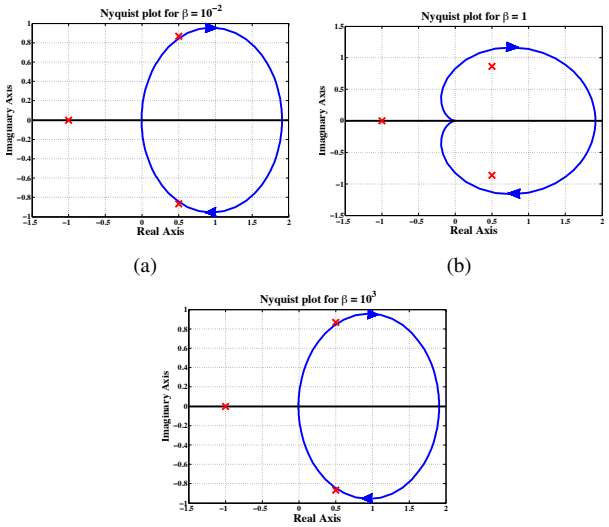


Fig. 3. With the parameters ($P_{\text{tot}i} = 100, k_{\text{on}i} = k_{\text{off}i} = 1000, \gamma_i = 1$) fixed, we vary β for different fixed values of α_i. A single repressor module is linearized about the solution u_i^* to the equation $u_i^* = \alpha_i \frac{\gamma_i^2 k_{\text{off}i} P_{\text{tot}i}}{\gamma_i^2 k_{\text{off}i} + k_{\text{on}i}(u_i^*)^2}$. We then derive the stability of the equilibrium points of the repressilator network by applying Lemma 1. The white regions indicate that the equilibrium points of the repressilator network are stable, while the dark region indicates the existence of some closed loop unstable poles.



(a) (b)

(c)

Fig. 4. $P_{\text{tot}i} = 100, k_{\text{on}i} = k_{\text{off}i} = 1000, \gamma_i = 1$ and $\alpha_i = 1$ in all the plots (a) Nyquist plot of $R(s)$ for $\beta = 10^{-2}$. No encirclement of any of the solutions to the equation $z^3 = -1$, implying repressilator network is stable. (b) Nyquist plot of $R(s)$ for $\beta = 1$. Nyquist contour encircles each of the points $z = e^{j\pi/3}$ and $z = e^{-j\pi/3}$ once, which indicates that the repressilator network will have two unstable poles. (c) Nyquist plot of $R(s)$ for $\beta = 10^3$. No encirclement of any of the solutions to the equation $z^3 = -1$, implying repressilator network is stable.

in cascade with a second block with input corresponding to the production rate of the complex X_1:P_1 and two outputs corresponding to the production rates of proteins X_1 and X_2, as described by the reactions

$$\begin{aligned} X_1 + P_1 \xrightarrow{k_{\text{on1}}} X_1{:}P_1 \qquad X_1{:}P_1 \xrightarrow{k_{\text{off1}}} X_1 + P_1 \\ X_1{:}P_1 \xrightarrow{\alpha_1} X_1{:}P_1 + X_2. \end{aligned} \tag{5}$$

The system of equations corresponding to these two blocks are, respectively, given by

$$\dot{X}_1 = \bar{u}_1 - \gamma_1 X_1, \qquad \bar{y}_1 = k_{\text{on1}} X_1 \tag{6}$$

and

$$\begin{aligned} \dot{X_1{:}P_1} &= \bar{u}_2(P_{\text{tot1}} - X_1{:}P_1) - k_{\text{off1}} X_1{:}P_1 \\ y_1 &= \alpha_1 X_1{:}P_1 \\ \bar{y}_2 &= k_{\text{off1}} X_1{:}P_1 - \bar{u}_2(P_{\text{tot1}} - X_1{:}P_1) \end{aligned} \tag{7}$$

where

$$\bar{u}_1 = u_1 + \bar{y}_2, \qquad \bar{u}_2 = \bar{y}_1,$$

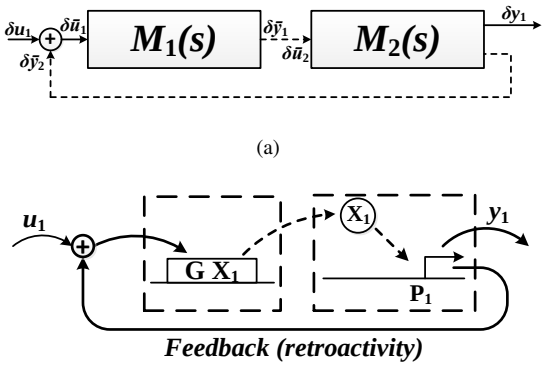
and can be viewed as the feedback loop shown in Figure 5.

The transfer function from $\delta\bar{u}_1$ to $\delta\bar{y}_1$ can be computed to be

$$\frac{\delta\bar{y}_1}{\delta\bar{u}_1}(s) = \frac{k_{\text{on1}}}{s + \gamma_1} =: M_1(s), \tag{8}$$

and the transfer functions from $\delta_{\bar{u}_2} = \delta_{\bar{y}_1}$ to δ_{y_1} and $\delta_{\bar{y}_2}$ are

$$\begin{bmatrix} \frac{\delta y_1}{\delta\bar{u}_2}(s) \\ \frac{\delta\bar{y}_2}{\delta\bar{u}_2}(s) \end{bmatrix} = \begin{bmatrix} \left(\frac{\alpha_1 k_{\text{off1}} P_{\text{tot1}}}{k_{\text{off1}} + \bar{u}_2^*}\right) \frac{1}{s + k_{\text{off1}} + \bar{u}_2^*} \\ -\left(\frac{k_{\text{off1}} P_{\text{tot1}}}{k_{\text{off1}} + \bar{u}_2^*}\right) \frac{s}{s + k_{\text{off1}} + \bar{u}_2^*} \end{bmatrix} =: M_2(s). \tag{9}$$



(a)

(b)

Fig. 5. (a) Systems representation of decomposition of an activator module (b) Biological diagram representing the decomposition of an activator module

From these two transfer functions, we can precisely compute the linearized transfer function from δu_1 to δy_1 to be the same as was computed in Section II-A.

In Del Vecchio et al's framework, the first block (4) is regarded as the main module that defines the dynamics of the protein X_1. In the second block (5), the protein X_1 binds with a promoter region to regulate the activation or repression of X_2. As highlighted in Figure 5, there is a feedback path in the interconnection between these two modules that necessarily affects the transfer function from u_1 to $\bar{y}_1$. This phenomenon is precisely a manifestation of the *retroactivity* property noted in [5]. Del Vecchio et al's sufficient conditions

for retroactivity to be small, i.e. (a) $k_{\text{off1}} >> k_{\text{on1}} P_{\text{tot1}}$ or (b) $X_1 >> P_{\text{tot1}}$, can be shown to attenuate retroactivity in our framework, without any assumptions on the timescales of the reactions. Either of these conditions being satisfied would make the numerical value of $\bar{u}_2^*$ large, which then makes the H_∞-norm of the transfer function $M_2(s)$ in (9) very small. This then means that the transfer function from δu_1 to $\delta \bar{y}_1$ in Figure 5(a) is approximately $M_1(s)$, which indicates the attenuation of retroactivity.

One can also see (6) and (7) as a feedback connection between two input-output modules for which standard transfer function calculus applies. However, to obtain the correct transfer function and equilibrium points, one must recognize that (7) is a single-input/*two-output* system. By not breaking (4)–(5) into two modules, we avoided having to work with MIMO systems in all the systems considered before this section, but this generalization may prove to be useful.

VI. Discussion

The method we have proposed can be generalized to break up more complex networks into modules. This would involve the creation of stable MIMO modules whose dynamics would not change upon interconnection. A generalized method to automatically decompose a biological network into these MIMO modules could be developed, with a similar objective as in [7], but using transfer function calculus as a basis for the decomposition. It would then be useful to find out if these modules are similar to functional modules that exist within a biological network or even a cell, such as the RNA processing unit or the DNA repair unit [17].

Breaking up a complex biological network into modules has been proven to be useful in the identification of parameters and pathways in the network [6], [11]. However, it has been shown that retroactivity could make such identification of parameters and pathways more challenging [12]. Since we define biological modules in such a way that they are not affected by retroactivity, it is conceivable that representing a biological network as an interconnection of such modules might be a better way to conduct system identification. A major challenge, however, is similar to the one that is faced when attempting to compute the Polymerases per Second (PoPS) unit, that is used to measure the inputs and outputs of a BioBrick device. To the authors' knowledge, it is not possible to experimentally measure this unit [18]. A future project would be to find a way to estimate the values of the inputs and outputs from our modules, given experimental data.

Recent work has also shown that gene regulation is an inherently noisy process [14] and therefore, a stochastic approach may be more appropriate to model the dynamics of chemical reactions. We hence, aim to generalize these results taking into account stochasticity and noise.

VII. Conclusion

We have shown that gene regulatory networks can be viewed as an interconnection of functionally isolated modules. We have used the four most common interconnections in systems theory to show that the input-output dynamics of these modules do not change upon interconnection. We also showed that these modules could be used to predict global network properties (such as stability), using the repressilator network by Elowitz and Liebler as an example [15]. Finally, we concluded by arguing that, in our framework, retroactivity can be viewed as an interconnection of two modules with feedback.

References

[1] N. B. C. Bianca, *Towards a mathematical theory of complex biological systems*, ser. Series in Mathematical Biology and Medicine. World Scientific Publishing Co. Pte. Ltd., 2011, vol. 11.

[2] L. H. Hartwell, J. J. Hopfield, S. Liebler, and A. W. Murray, "From molecular to modular cell biology," *Nature*, vol. 402, no. 6761, pp. C47–52, December 1999.

[3] H. M. Sauro, "Modularity defined," *Molecular Systems Biology*, vol. 4, no. 166, March 2008.

[4] T. Schlitt and A. Brazma, "Current approaches to gene regulatory network modelling." *BMC Bioinformatics*, vol. 8, no. S-6, 2007.

[5] D. Del-Vecchio, A. J. Ninfa, and E. D. Sontag, "Modular cell biology: retroactivity and insulation," *Molecular Systems Biology*, vol. 4, no. 161, 2008.

[6] B. N. Kholodenko, A. Kiyatkin, F. J. Bruggeman, E. Sontag, H. V. Westerhoff, and J. B. Hoek, "Untangling the wires: a strategy to trace functional interactions in signaling and gene networks." *Proc Natl Acad Sci U S A*, vol. 99, no. 20, pp. 12 841–6, 2002.

[7] J. Anderson, Y.-C. Chang, and A. Papachristodoulou, "Model decomposition and reduction tools for large-scale networks in systems biology," *Automatica*, vol. 47, no. 6, pp. 1165–1174, June 2011.

[8] J. P. Hespanha, *Linear Systems Theory*. Princeton Press, 2009.

[9] D. Angeli and E. Sontag, "Monotone control systems," *IEEE Trans. Automat. Control*, vol. 48, no. 10, pp. 1684–1698, 2003.

[10] ——, "Interconnections of monotone systems with steady-state characteristics," in *Optimal control, stabilization and nonsmooth analysis*, ser. Lecture Notes in Control and Inform. Sci. Berlin: Springer, 2004, vol. 301, pp. 135–154.

[11] M. Andrec, B. N. Kholodenko, R. M. Levy, and E. D. Sontag, "Inference of signaling and gene regulatory networks by steady-state perturbation experiments: structure and accuracy," *J. Theoret. Biol.*, vol. 232, no. 3, pp. 427–441, 2005.

[12] E. Sontag, "Modularity, retroactivity, and structural identification," in *Design and Analysis of Biomolecular Circuits*, H. Koeppl, G. Setti, M. di Bernardo, and D. Densmore, Eds. Springer New York, 2011, pp. 183–200.

[13] Y.-J. J. Shin and L. Bleris, "Linear control theory for gene network modeling." *PloS one*, vol. 5, no. 9, Sept. 2010.

[14] A. Singh and J. P. Hespanha, "Optimal feedback strength for noise suppression in auto-regulatory gene networks," *Biophysical Journal*, vol. 96, no. 10, pp. 4013–4023, May 2009.

[15] M. B. Elowitz and S. Leibler, "A synthetic oscillatory network of transcriptional regulators," *Nature*, vol. 403, no. 6767, pp. 335–8, Jan. 2000.

[16] R. C. Dorf and R. H. Bishop, *Modern Control Systems*. Pearson Prentice-Hall, 2004.

[17] J. Karr, J. Sanghvi, D. Macklin, M. Gutschow, J. Jacobs, B. Bolival, N. Assad-Garcia, J. Glass, and M. Covert, "A whole-cell computational model predicts phenotype from genotype." *Cell*, vol. 150, no. 5, pp. 389–401, 2012-07-20 00:00:00.0.

[18] J. R. Kelly, A. J. Rubin, J. H. Davis, C. M. Ajo-Franklin, J. Cumbers, M. J. Czar, K. de Mora, A. L. Glieberman, D. D. Monie, and D. Endy, "Measuring the activity of BioBrick promoters using an in vivo reference standard." *Journal of biological engineering*, vol. 3, no. 1, Mar. 2009.