Limitations and Trade-offs in Gene Expression due to Competition for Shared Cellular Resources

Andras Gyorgy and Domitilla Del Vecchio

***Abstract*— Gene circuits share transcriptional and translational resources in the cell. The fact that these common resources are available only in limited amounts leads to unexpected couplings in protein expressions. As a result, our predictive ability of describing the behavior of gene circuits is limited. In this paper, we consider the simultaneous expression of proteins and describe the coupling among protein concentrations due to competition for RNA polymerase and ribosomes. In particular, we identify the limitations and trade-offs in gene expression by characterizing the attainable combinations of protein concentrations. We further present two application examples of our results: we show that even in the absence of regulatory linkages, genes can seemingly behave as repressors, and surprisingly, as activators to each other, purely due to the limited availability of shared cellular resources.**

I. INTRODUCTION

One of the major bottlenecks in systems and synthetic biology is context-dependence [1], as it hinders our ability to accurately predict the behavior of complex systems from that of the composing modules [2]. This lack of modularity is particularly important when engineering biological systems using smaller components [3], as it often leads to a lengthy and *ad hoc* re-design process every time the context changes [4]. Context-dependence arises due to a number of different factors: unknown regulatory linkages; loading effects due to known regulatory interactions between components, a phenomenon known as retroactivity [5], [6]; metabolic burden [7]; cell growth [8]; and competition for shared cellular resources [9].

In this paper, we focus on the effects of competition for transcriptional and translational resources on gene expression. Since these resources are available only in limited amounts, they have to be reallocated every time new genes are introduced into the cell, or when the activity of already present genes changes. Due to the reallocation of these common resources, the over-expression of one gene can affect the growth rate of the cell [8], and it can decrease the expression of other genes [10]. As a result, the expression of different genes become coupled, even in the absence of regulatory linkages among them. To accurately predict and control the behavior of gene circuits, we must determine the distribution of shared resources, that is, the cellular economy of gene expression.

This work was supported by AFOSR grant FA9550-12-1-0129 and NIGMS grant P50 GM098792.

A. Gyorgy is with the Department of Electrical Engineering and Computer Science, MIT, Cambridge, MA 02139, USA gyorgy@mit.edu

D. Del Vecchio is with the Department of Mechanical Engineering, MIT, Cambridge, MA 02139, USA ddv@mit.edu

Here, we characterize how the expression of different exogenous genes become coupled due to competition for RNA polymerase (RNAP) and ribosomes. We focus on RNAP and ribosomes as their availability is considered to be the major limiting factor in transcription [11] and translation [10], respectively. We prove that due to the limited availability of these cellular resources, the attainable protein concentrations lie within the intersection of simplexes, and we show how these simplexes depend on various biochemical parameters, such as ribosome binding site (RBS) strength and DNA copy number. Building upon our results, we show that even in the absence of regulatory linkages, genes can seemingly repress and activate each other, as a result of the reallocation of limited resources.

Our work is closely related to recent efforts investigating the effects of shared cellular resources on gene circuits. In particular, in [12] and [13] the authors detail the effects of the limited availability of ribosomes causing translational crosstalk, a phenomenon verified experimentally in [14] in cell-free systems. A general framework for studying the effects of resource competition is presented in [15] using Metabolic Control Analysis [16], yielding response coefficients that describe local flux sensitivities in a gene network. Our work complements these results as we consider the role of both RNAP and ribosomes to characterize the global limitations and trade-offs in protein expression for n genes. Some of these results have been validated *in vivo* for two genes [17].

This paper is organized as follows. In Section II, the system of interest is introduced, together with the motivation and research question: *Having n genes, what are the limitations and trade-offs in gene expression due to competition for shared cellular resources?* In Section III, we determine the attainable protein concentrations and characterize how various biochemical parameters affect the interdependence in gene expression. In Section IV, we present two implications of the limited availability of RNAP and ribosomes on gene expression. Finally, we conclude our results and present future research directions in Section V.

II. SYSTEM MODEL AND PROBLEM FORMULATION

We consider a system in which n genes are expressed. In particular, each gene is first transcribed by RNAP to mRNA, then mRNA is translated by ribosomes to protein (Fig. 1A). Furthermore, we focus on the case when the transcription of each gene is regulated by a transcription factor (TF) as follows. In the case of gene i expressing protein p_i, TF u_i

first binds to the empty promoter b_i^* forming the promoter complex b_i. Then, the binding of RNAP x to b_i can form the transriptionally active promoter complex c_i, resulting in the production of mRNA m_i at rate γ_i (encompassing the elongation reactions). Finally, mRNA decays at rate δ_i. Consequently, the reactions describing the transcriptional processes for gene i are as follows:

$$\mathrm{u}_i+\mathrm{b}_i^* \underset{\mu_i^-}{\overset{\mu_i^+}{\rightleftharpoons}} \mathrm{b}_i,\ \mathrm{b}_i+\mathrm{x} \underset{\kappa_i^-}{\overset{\kappa_i^+}{\rightleftharpoons}} \mathrm{c}_i,\ \mathrm{c}_i \xrightarrow{\gamma_i} \mathrm{b}_i+\mathrm{x}+\mathrm{m}_i,\ \mathrm{m}_i \xrightarrow{\delta_i} \emptyset.$$

Translation of m_i is initialized by the ribosome y binding to the RBS of the mRNA m_i, forming the translationally active complex d_i. The decay of mRNA when bound to the ribosome occurs with rate constant $a_i\delta_i$ where $0 < a_i \leq 1$ ($a_i \to 0$ represents the case when the ribosome-bound mRNA is protected from degradation, whereas $a_i = 1$ models the scenario when ribosomes provide no protection against degradation, which is considered in what follows). Protein p_i decays at rate λ_i, whereas elongation and production are lumped together in one step with effective production rate constant π_i. Therefore, the reactions describing the translation processes for gene i are given by

$$\mathrm{m}_i+\mathrm{y} \underset{k_i^-}{\overset{k_i^+}{\rightleftharpoons}} \mathrm{d}_i \xrightarrow{a_i\delta_i} y, \qquad \mathrm{d}_i \xrightarrow{\pi_i} \mathrm{m}_i+\mathrm{y}+\mathrm{p}_i, \qquad \mathrm{p}_i \xrightarrow{\lambda_i} \emptyset.$$

Consequently, the corresponding differential equation model for $i = 1, 2, \ldots, n$ is given by

$$\begin{aligned}
\dot{b}_i &= \left(\mu_i^+ u_i b_i^* - \mu_i^- b_i\right) - \left(\kappa_i^+ x b_i - \kappa_i^- c_i\right) + \gamma_i c_i,\\
\dot{c}_i &= \left(\kappa_i^+ x b_i - \kappa_i^- c_i\right) - \gamma_i c_i,\\
\dot{m}_i &= \gamma_i c_i - \delta_i m_i - \left(k_i^+ m_i y - k_i^- d_i\right) + \pi_i d_i,\\
\dot{d}_i &= \left(k_i^+ m_i y - k_i^- d_i\right) - \pi_i d_i - a_i\delta_i d_i,\\
\dot{p}_i &= \pi_i d_i - \lambda_i p_i.
\end{aligned} \tag{1}$$

In the above model, c_i encompasses both the promoter bound and transcribing RNAP, and similarly d_i captures both the RBS bound and translating ribosomes. Furthermore, the parameters δ_i and λ_i encompass both degradation and cell growth, and in this paper, we assume that the rate at which cells growth remains the same, which is the case, for instance, in the experiments in [17]. It must be pointed out that if an overexpressed protein is, for instance, toxic to the cell, the growth rate may decrease [8].

A. RNAP and Ribosome Demand at the Steady State

Introduce the dissociation constants $\kappa_i = (\kappa_i^- + \gamma_i)/\kappa_i^+$ and $k_i = (k_i^- + \pi_i + \delta_i)/k_i^+$ for $i = 1, 2, \ldots, n$. Given that protein production and decay are much slower than binding and unbinding reactions [18], we have $\gamma_i \ll \kappa_i^-$ and $\pi_i, \delta_i \ll k_i^-$, so that $\kappa_i \approx \kappa_i^-/\kappa_i^+$ and $k_i \approx k_i^-/k_i^+$. The stronger the binding of RNAP to the promoter, the smaller κ_i, and similarly, the stronger the binding of ribosome to the RBS, the smaller k_i. Next, define the dissociation constant $\mu_i = \mu_i^-/\mu_i^+$ of the TF u_i to the promoter of gene i, and let

$$\epsilon_i = \frac{\frac{u_i}{\mu_i}\left(1+\frac{x}{\kappa_i}\right)}{1+\frac{u_i}{\mu_i}\left(1+\frac{x}{\kappa_i}\right)}, \qquad \text{for } i = 1, 2, \ldots, n. \tag{2}$$

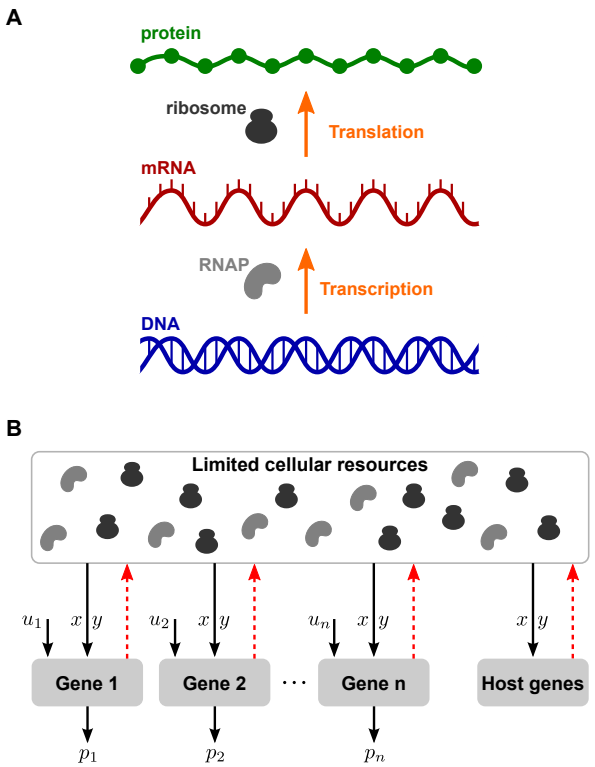

Fig. 1. Genes compete for transcriptional and translational resources. (A) DNA is transcribed into mRNA by RNAP, which is then translated to protein by ribosome. (B) The sharing of limited resources couples the expression of genes, even in the absence of regulatory linkages, as a result of loading (red) imposed by each gene on the pool of shared resources.

Assuming that DNA concentration is constant [19], we have that $\eta_i = b_i^* + b_i + c_i$, where η_i is the total concentration of the promoter of gene i. We further have $\epsilon_i = (b_i + c_i)/\eta_i$, so that $\epsilon_i \in [0, 1)$ is the fraction of the promoter of gene i activated by u_i. At the steady state of (1), we have

$$c_i = \epsilon_i \eta_i \frac{x}{x+\kappa_i}, \qquad d_i = \epsilon_i \frac{\gamma_i \eta_i}{\delta_i} \frac{x}{x+\kappa_i} \frac{y}{y+k_i}, \tag{3}$$

whereas the concentration of protein p_i is given by

$$p_i = \epsilon_i \frac{\pi_i}{\lambda_i} \frac{\gamma_i \eta_i}{\delta_i} \frac{x}{x+\kappa_i} \frac{y}{y+k_i}, \qquad \text{for } i = 1, 2, \ldots, n. \tag{4}$$

We call c_i and d_i in (3) the *RNAP and ribosome demand* of gene i at the steady state, respectively, as they represent the concentration of RNAP and ribosomes bound to the promoter and mRNA, respectively. The protein concentrations p_i for $i = 1, 2, \ldots, n$ in (4) are implicitly coupled as the free concentration x and y of RNAP and ribosomes, respectively, depend on the demand by the genes (Fig. 1B), as we detail in the next section.

B. Modeling the Limited Availability of RNAP & Ribosomes

Combining the results of [20], [21], [22], [23], it is shown in [24] that in the case when the growth rate of cells is the same, the limited availability of RNAP and ribosomes can

be modeled by the conservation laws

$$X = (1 + W_x)x + \sum_{i=1}^{n} \epsilon_i \eta_i \frac{x}{x + \kappa_i}, \tag{5}$$

$$Y = y + \sum_{i=1}^{n} \epsilon_i \frac{\gamma_i \eta_i}{\delta_i} \frac{x}{x + \kappa_i} \frac{y}{y + k_i}. \tag{6}$$

where X and Y denote the available pool of RNAP and ribosomes, respectively, and $W_x > 0$ (capturing the non-specific binding of RNAP to the chromosome).

C. Problem Formulation

Let $\epsilon = (\epsilon_1, \epsilon_2, \ldots, \epsilon_n)^T$ and $u = (u_1, u_2, \ldots, u_n)^T$, and write (2) as $\epsilon = E(u, x)$. With this, (5) and (6) can be written as $X = F_\epsilon(\epsilon, x)$ and $Y = G_\epsilon(\epsilon, x, y)$ respectively, and (4) with $p = (p_1, p_2, \ldots, p_n)^T$ as $p = H_\epsilon(\epsilon, x, y)$. Furthermore, define the sets $\mathcal{U} = [0, \infty)^n$ and

$$\begin{aligned} \mathcal{P} = \{p \mid\ & p = H_\epsilon(E(u,x),x,y),\ X = F_\epsilon(E(u,x),x), \\ & Y = G_\epsilon(E(u,x),x,y),\ x \in [0, X],\ y \in [0, Y], \\ & u \in \mathcal{U}\}, \end{aligned} \tag{7}$$

so that $\mathcal{P}$ is the set of protein concentrations attainable at the steady state. Therefore, we call $\mathcal{P}$ the *realizable region*. Here, we seek an explicit characterization of $\mathcal{P}$ solely in terms of p, instead of the definition in (7) involving u, x and y in the form of implicit constraints. As a result, we can answer the following questions. *How does the concentration of protein p_j change upon activation of gene i for $j \neq i$? To what extent is it possible to increase the concentration of p_i without affecting the concentration of p_j?* In other words, we seek to characterize the limitations and trade-offs in protein production due to the limited availability of RNAP and ribosomes.

III. REALIZABLE REGION

We characterize the realizable region $\mathcal{P}$ through a series of intermediate results. In particular, we first focus on the activation level ϵ_i of gene i for $i = 1, 2, \ldots, n$. From [24], in order to find the realizable region $\mathcal{P}$, it is sufficient to consider (4)–(6) for $\epsilon \in \mathcal{E} = [0, 1)^n$, instead of considering (4)–(6) with (2) for $u \in \mathcal{U}$. Then, we consider a biologically reasonable approximation [24] of (4)–(6) and characterize the corresponding set $\mathcal{S}$ of attainable protein concentrations. Finally, we prove that $\mathcal{P} \subseteq \mathcal{S}$.

A. Approximate Model & Approximate Realizable Region $\mathcal{S}$

As an intermediate step to characterize the realizable region $\mathcal{P}$, consider the (biologically reasonable, see [24]) approximations $x \ll \kappa_i$ and $y \ll k_i$ for $i = 1, 2, \ldots, n$. Using $x + \kappa_i \approx \kappa_i$ and $y + k_i \approx k_i$ in (4)–(6) yields

$$p_i = \frac{Q_i \epsilon_i}{1 + \sum_{i=1}^{n} R_i \epsilon_i}, \qquad \text{for } i = 1, 2, \ldots, n \tag{8}$$

with

$$\begin{aligned} Q_i &= \frac{1}{1 + W_x} \frac{\pi_i}{\lambda_i} \frac{\gamma_i \eta_i}{\delta_i} \frac{1}{\kappa_i k_i} XY, \\ R_i &= \frac{1}{1 + W_x} \left(\frac{\gamma_i \eta_i}{\delta_i} \frac{1}{\kappa_i k_i} X + \frac{\eta_i}{\kappa_i} \right). \end{aligned} \tag{9}$$

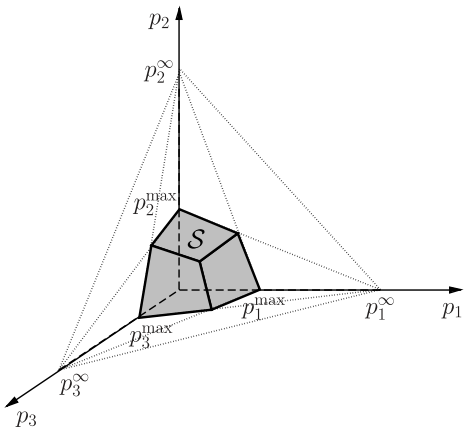

Fig. 2. In the case of $n = 3$, $\mathcal{S}$ in (11) is the intersection of the simplexes $\mathcal{S}_1$, $\mathcal{S}_2$ and $\mathcal{S}_3$ (Lemma 1), where $\mathcal{S}_i$ is defined in (10) as the simplex given by the origin, $p_i^{\max}$ on the p_i-axis and p_j^∞ on the p_j-axis for $j \neq i$.

Furthermore, let $\hat{A} : \mathbb{R}^n \to \mathbb{R}^n$ be the function mapping ϵ to p according to (8), so that $p = \hat{A}(\epsilon)$. Next, define $p_i^{\max} = Q_i/(1 + R_i)$ and $p_i^\infty = Q_i/R_i$ and introduce the simplex $\mathcal{S}_i$ for $i = 1, 2, \ldots, n$ as

$$\mathcal{S}_i = \left\{ p \mid p \geq 0 \text{ and } \frac{p_i}{p_i^{\max}} + \sum_{\substack{j=1 \\ j \neq i}}^{n} \frac{p_j}{p_j^\infty} < 1 \right\}. \tag{10}$$

Lemma 1. *Let*

$$\mathcal{S} = \{p \mid p = \hat{A}(\epsilon),\ \epsilon \in \mathcal{E}\}. \tag{11}$$

Then, we obtain $\mathcal{S} = \cap_{i=1}^n \mathcal{S}_i$ where $\mathcal{S}_i$ is defined in (10)*.*

Proof: Here we present a sketch, for details see [24].

We first show $\mathcal{S} \subseteq \cap_{i=1}^n \mathcal{S}_i$ as follows. By introducing $\mathcal{E}_i = \{\epsilon \mid \epsilon_i \in [0, 1) \text{ and } \epsilon_j \in [0, \infty) \text{ for } j \neq i\}$ we can show that $p = \hat{A}(\epsilon) \in \mathcal{S}_i$ by (10) for $\epsilon \in \mathcal{E}_i$. Combining this together with the fact that $\epsilon \in \mathcal{E} = \cap_{i=1}^n \mathcal{E}_i$ yields that $\mathcal{S} \subseteq \cap_{i=1}^n \mathcal{S}_i$.

Second, we prove $\cap_{i=1}^n \mathcal{S}_i \subseteq \mathcal{S}$ by showing that for any $p \in \cap_{i=1}^n \mathcal{S}_i$ there exists an $\epsilon \in \mathcal{E}$ such that $p = \hat{A}(\epsilon)$. In particular, pick $p \in \cap_{i=1}^n \mathcal{S}_i$ and define for $i = 1, 2, \ldots, n$

$$P_i = \frac{p_i}{1 - \sum_{\substack{j=1 \\ j \neq i}}^{n} \frac{p_j}{p_j^\infty}} \qquad \text{and} \qquad \epsilon_i = \frac{P_i}{Q_i - R_i P_i} \tag{12}$$

such that $p = \hat{A}(\epsilon)$. Then it is easy to show that $\epsilon_i \in [0, 1)$ for $i = 1, 2, \ldots, n$, so that $\epsilon \in \mathcal{E}$, concluding the proof. ■

The realizable region $\mathcal{S}$ of protein concentrations when $x \ll \kappa_i$ and $y \ll k_i$ is given as $\mathcal{S} = \cap_{i=1}^n \mathcal{S}_i$ by Lemma 1, where $\mathcal{S}_i$ is the n-dimensional simplex defined by the following $n + 1$ vertices: the origin, $p_i^{\max}$ on the p_i-axis and p_j^∞ on the p_j-axis for $j \neq i$ (see Fig. 2). Furthermore, the dependence of $\mathcal{S}_i$ on the biochemical parameters is given by the expressions of $p_i^{\max}$ and p_i^∞. For instance, both $p_i^{\max}$ and p_i^∞ increase as k_i decreases (stronger RBS for gene i), and $p_i^{\max}$ increases while p_i^∞ remains unaffected as κ_i decreases (stronger promoter for gene i).

B. The Realizable Region $\mathcal{P}$ Lies Inside $\mathcal{S}$

We next show that even when the approximations $x \ll \kappa_i$ and $y \ll k_i$ do not hold, the set of attainable protein concentrations $\mathcal{P}$ from (7) lie within $\mathcal{S}$ given in (11).

Theorem 1. *Considering $\mathcal{P}$ and $\mathcal{S}$ defined in* (7) *and* (11)*, respectively, we obtain that $\mathcal{P} \subseteq \mathcal{S}$.*

Proof: Here we present a sketch, for details, see [24]. Fix $\epsilon \in \mathcal{E}$, let $f_\epsilon(\epsilon)$ and $g_\epsilon(\epsilon)$ be such that $X = F_\epsilon(\epsilon, f_\epsilon(\epsilon))$ and $Y = G_\epsilon(\epsilon, f_\epsilon(\epsilon), g_\epsilon(\epsilon))$. Then introduce $p = H_\epsilon(\epsilon, f_\epsilon(\epsilon), g_\epsilon(\epsilon))$, the value of p given by (4) satisfying the constraints (5)–(6). In what follows, we show that $p \in \mathcal{S}_i$ for $i = 1, 2, \ldots, n$, since it implies that $p \in \mathcal{S}$ as $\mathcal{S} = \cap_{i=1}^n \mathcal{S}_i$ by Lemma 1.

To show that $p \in \mathcal{S}_i$ for $i = 1, 2, \ldots, n$, define

$$\alpha_i = \frac{\kappa_i}{x + \kappa_i}, \qquad \beta_i = \frac{k_i}{y + k_i}, \qquad \epsilon_i' = \alpha_i \beta_i \epsilon_i, \quad (13)$$

so that we can write p_i in (4) as

$$p_i = \frac{Q_i \epsilon_i'}{1 + \sum_{i=1}^n R_i' \epsilon_i'} \quad (14)$$

with Q_i from (9) and with $R_i' = [\gamma_i \eta_i X/(\kappa_i k_i \delta_i) + \eta_i/(\beta_i \kappa_i)]/(1 + W_x)$, and let $\hat{p} = (\hat{p}_1, \hat{p}_2, \ldots, \hat{p}_n)^T$ where $\hat{p}_i$ is given by (8). With this, we can show that $0 \le p_i < \hat{p}_i$. This together with the fact that

$$\frac{\hat{p}_i}{p_i^{\max}} + \sum_{\substack{j=1 \\ j \neq i}}^n \frac{\hat{p}_j}{p_j^\infty} - 1 < 0, \quad (15)$$

from Lemma 1 implies that

$$\frac{p_i}{p_i^{\max}} + \sum_{\substack{j=1 \\ j \neq i}}^n \frac{p_j}{p_j^\infty} < \frac{\hat{p}_i}{p_i^{\max}} + \sum_{\substack{j=1 \\ j \neq i}}^n \frac{\hat{p}_j}{p_j^\infty} < 1. \quad (16)$$

As a result, $p \in \mathcal{S}_i$ for $i = 1, 2, \ldots, n$ by (10), concluding the proof. ∎

Introduce x_0 and y_0 such that $X = F_\epsilon(0, x_0)$ and $Y = G_\epsilon(0, x_0, y_0)$, that is, x_0 and y_0 denote the concentration of free RNAP and ribosomes, respectively, when none of the genes in Fig. 1B are activated ($\epsilon_i = 0$ for $i = 1, 2, \ldots, n$). Next, define

$$\mathcal{B} = \{p \mid p = H(u, x_0, y_0),\ u \in \mathcal{U}\}, \quad (17)$$

representing the set of attainable protein concentrations without considering competition for RNAP and ribosomes (so that $x = x_0$ and $y = y_0$), see Fig. 3 for a particular example when $n = 2$. Since $\mathcal{S} \subset \mathcal{B}$ in Fig. 3, if $(p_1, p_2) \in \mathcal{B} \setminus \mathcal{S}$ then $(p_1, p_2) \notin \mathcal{P}$. As a result, without considering competition for RNAP and ribosomes we would erroneously conclude that the protein concentrations (p_1, p_2) are attainable.

IV. Practical Implications of the Limited Availability of RNAP and Ribosomes

Here, we show that proteins can seemingly behave both as repressors and as activators, purely as an effect of the limited availability of RNAP and ribosomes.

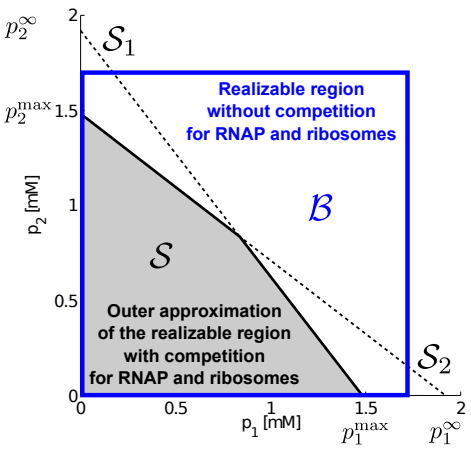

Fig. 3. The set of attainable protein concentrations without considering competition for RNAP and ribosomes is $\mathcal{B}$ given in (17). When considering the limited availability of RNAP and ribosomes given by (5)–(6), the set of attainable protein concentrations $\mathcal{P}$ in (7) lie inside the outer approximation $\mathcal{S}$ from (11). The set $\mathcal{S}$ is the intersection of the simplexes $\mathcal{S}_1$ and $\mathcal{S}_2$ in (10) (triangles with dashed border). Simulation parameters: $\eta_1 = \eta_2 = 100$nM, $\lambda_1 = \lambda_2 = 1\text{hr}^{-1}$, $\pi_1 = \pi_2 = 1500\text{hr}^{-1}$, $\delta_1 = \delta_2 = 10\text{hr}^{-1}$, $\gamma_1 = \gamma_2 = 500\text{hr}^{-1}$ $\kappa_1 = \kappa_2 = 150$nM, $k_1 = k_2 = 1000$nM, $W_x = 12$, $X = Y = 1300$nM, $x_0 = 100$nM and $y_0 = 1300$nM. For details on the parameters, see [24].

A. Lateral Inhibition with Two Genes

Consider two genes, and for simplicity, focus on the biologically reasonable approximations [24] $x \ll \kappa_i$ and $y \ll k_i$ for $i = 1, 2$, so that the realizable set of protein concentrations $\mathcal{P}$ is equal to $\mathcal{S}$ by Lemma 1. We investigate how p_2 changes as the expression of p_1 increases.

Fix the activation ϵ_2 of gene 2 ($\epsilon_2 = \epsilon_2^*$), while increasing the activation ϵ_1 of gene 1. Without considering competition for the shared resources, the set of attainable protein concentrations is $\mathcal{B}$ given in (17), see Fig. 4A. In this case, the concentration x and y of free RNAP and ribosomes, respectively, are independent of the value of ϵ_1 and ϵ_2. As a result, p_1 increases while p_2 remains unaffected when increasing the activation ϵ_1 of gene 1 by (8). That is, the attainable pairs (p_1, p_2) lie along a horizontal line (Fig. 4A). However, due to the limited availability of resources, p_2 decreases by (8) as the activation ϵ_1 of gene 1 increases, since some of the resources have to be reallocated from gene 2 to gene 1.

Referring to (8), the pair (p_1, p_2) satisfies the linear constraint

$$\underbrace{\left(A_1 + \frac{B_1}{X}\right)}_{\alpha} p_1 + \underbrace{\left(A_2 + \frac{B_2 + C/\epsilon_2^*}{X}\right)}_{\beta} p_2 = Y \quad (18)$$

with $A_i = \lambda_i/\pi_i$, $B_i = \delta_i k_i A_i/\gamma_i$ and $C = B_2 \kappa_2 k_2 (1 + W_x)/\eta_2$. Since $\frac{\partial p_1}{\partial \epsilon_1} > 0$ and $\frac{\partial p_2}{\partial \epsilon_1} < 0$ from (8), the pair (p_1, p_2) moves along the line (18) from left to right by increasing ϵ_1 (red line in Fig. 4B). The top boundary of $\mathcal{S}$ in Fig. 4B is given by (18) when $\epsilon_2 = 1$ (gene 2 is fully activated), and similarly, the right boundary of $\mathcal{S}$ corresponds to the case when $\epsilon_1 = 1$ (gene 1 is fully activated).

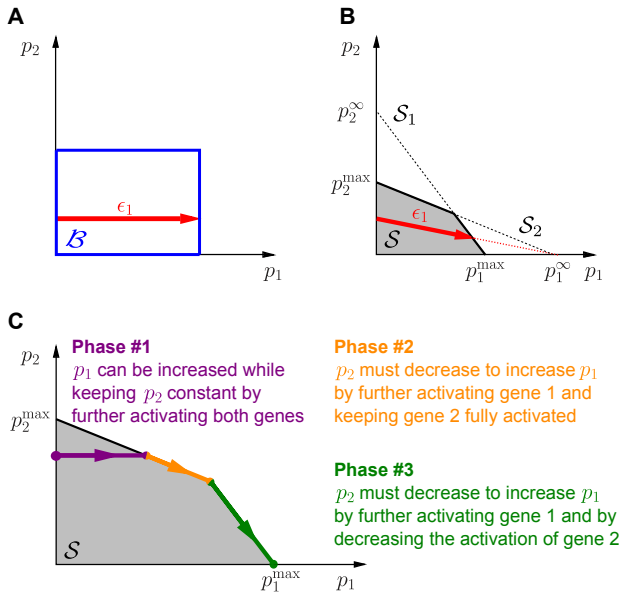

Fig. 4. Limitations and trade-offs in protein concentrations in the case of two genes. (A) Without competition for RNAP and ribosomes, the concentration of p_2 would not decrease when activating gene 1 (red). (B) When the genes are competing for RNAP and ribosomes, resources need to be reallocated from p_2 for the production of p_1. As a result, the concentration of p_2 decreases upon activation of gene 1 (red). (C) We can first increase the expression of gene 1 without decreasing p_2 (purple), but once reaching the boundary of the realizable region $\mathcal{S}$, the expression of gene 2 must decrease to further increase p_1 (orange and green).

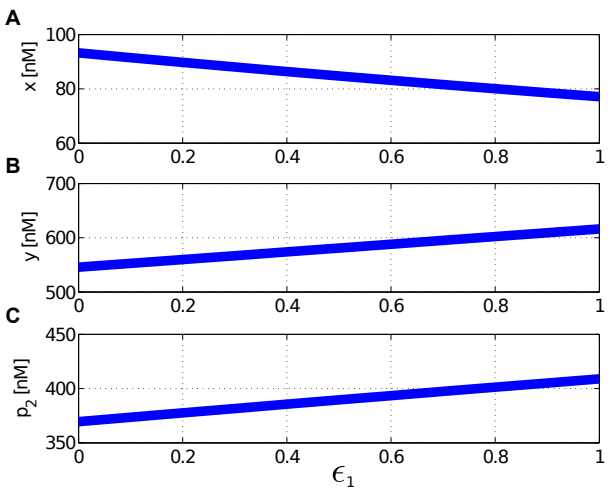

Fig. 5. Lateral activation in the case of $n = 3$ genes upon activation of gene 1. (A) The free concentration x of RNAP decreases. (B) The free concentration y of ribosomes increases as the activation level of gene 1 increases, since they are freed up from gene 3. (C) The concentration p_2 of p_2 increases as the activation level of gene 1 increases (without using the approximations $\kappa_2 \ll x \ll \kappa_1, \kappa_3$ and $y \ll k_1, k_2, k_3$). Simulation parameters: $X = Y = 1300$nM, $W_x = 12$, $\gamma_1 = \gamma_2 = \gamma_3 = 500\text{hr}^{-1}$, $\pi_1 = \pi_2 = \pi_3 = 1500\text{hr}^{-1}$, $\delta_1 = \delta_2 = \delta_3 = 10\text{hr}^{-1}$, $\lambda_1 = \lambda_2 = \lambda_3 = 1\text{hr}^{-1}$, $\eta_1 = 100$nM, $\eta_2 = 10$nM, $\eta_3 = 100$nM, $\kappa_1 = 100$nM, $\kappa_2 = 10$nM, $\kappa_3 = 1000$nM, $k_1 = 10$M, $k_2 = 1$mM and $k_3 = 1\mu$M. For details on the parameters, see [24].

The linear constraint in (18) can be interpreted as an isocost line [25], a concept introduced in microeconomics to describe the combinations of two products that can be purchased with a limited budget. Here, the products are p_1 and p_2 with prices α and β, respectively, whereas the budget Y is the concentration of available ribosomes. Increasing the availability of resources (RNAP and ribosomes) allows for purchasing more products: the value of p_1 and p_2 can be increased simultaneously. In particular, increasing the concentration X of available RNAP molecules decreases the prices α and β, whereas increasing the concentration of available ribosomes Y increases the budget. Furthermore, the isocost line describes how changing the biochemical parameters of a gene affects the extent of competition due to the limited availability of resources. In particular, the slope of the isocost line is $-\alpha/\beta$ by (18), so that producing an extra p_1 decreases the concentration of p_2 by α/β. The "more expensive" p_1 compared to p_2 (the greater α/β), the more p_2 have to be sacrificed in order to purchase an additional unit of p_1. For instance, α decreases with the dissociation constant k_1, so that stronger RBS for gene 1 makes the isocost line flatter by (18), verified *in vivo* in [17].

Without competition for shared resources, increasing the activation ϵ_1 of gene 1 does not affect p_2 (Fig. 4A). However, due to the limited availability of RNAP and ribosomes, the expression of gene 2 decreases when activating p_1 according to (18), as $\frac{\partial p_2}{\partial \epsilon_1} < 0 < \frac{\partial p_1}{\partial \epsilon_1}$ by (8). As a result, to increase p_2 and keep p_1 unaffected (phase #1 in Fig. 4C), we must increase the activation of both genes: by (8), increasing ϵ_1 yields greater expression of p_1, and the resulting decrease in p_2 can be compensated by increasing ϵ_2. However, when gene 2 becomes fully activated ($\epsilon_2 \to 1$), compensation is no longer possible, so that further increasing the activation ϵ_1 of gene 1 decreases p_1 (phase #2 in Fig. 4C). Finally, when gene 1 becomes fully activated ($\epsilon_1 \to 1$), the concentration of p_1 cannot be further increased while keeping gene 2 fully activated. Instead, we must decrease ϵ_2 so that resources can be reallocated to the expression of p_1 (phase #3 in Fig. 4C).

B. Lateral Activation with Three Genes

Since genes compete for the shared resources, one would expect that activating one gene decreases the expression of a different one. Here, we show that this is not always the case, and that counter-intuitively, unconnected genes can behave as activators to each other due to the limited availability of resources. To this end, consider $\kappa_2 \ll x \ll \kappa_1, \kappa_3$ and $y \ll k_1, k_2, k_3$, so that the promoter of gene 2 is saturated with RNAP. Furthermore, we focus on the case when genes 2 and 3 are fully activated ($\epsilon_2, \epsilon_3 \to 1$). Considering (5)–(6) when $\kappa_2 \ll x \ll \kappa_1, \kappa_3$ and $y \ll k_1, k_2, k_3$, and taking the derivative of p_2 in (4) with respect to ϵ_1 yields

$$\operatorname{sgn}\left(\frac{\mathrm{d}p_2}{\mathrm{d}\epsilon_1}\right) = \operatorname{sgn}\left(\frac{\gamma_3\eta_3}{\delta_3\kappa_3 k_3} - \left(1 + W_x + \frac{\eta_3}{\kappa_3}\right)\frac{\gamma_1}{\delta_1 k_1}\right).$$

As a result, $\frac{\mathrm{d}p_2}{\mathrm{d}\epsilon_1} > 0$ if, for instance, k_1 is sufficiently large (the RBS of the mRNA encoding p_1 is sufficiently weak). In this case, activating gene 1 increases the concentration of p_2, despite gene 2 being already fully activated ($\epsilon_2 \to 1$).

This seemingly counter-intuitive result can be explained as follows. Activating gene 1 results in an increased demand for RNAP, thus less RNAP is available for the other two genes by (5). However, since the promoter of gene 2 is particularly strong ($\kappa_2 \ll x$), it stays saturated with RNAP, that is, the concentration of mRNA encoding p_2 remains about the same. By contrast, the promoter of gene 3 is weak ($\kappa_3 \gg x$), so that less mRNA encoding p_3 is produced, i.e., the demand for ribosomes by gene 3 decreases by (3). In the meantime, if the RBS of the mRNA encoding p_1 is weak (k_1 is sufficiently large), the demand for ribosomes by gene 1 is negligible by (3). Consequently, the ribosomes not used by gene 3 can be used by gene 2. This lateral activation phenomenon is demonstrated in Fig. 5.

V. DISCUSSION

In this paper, we have characterized how the concentration of proteins become coupled due to competition for shared cellular resources, even in the absence of regulatory linkages. In particular, we showed that the realizable region $\mathcal{P}$ of protein concentrations lies within $\mathcal{S}$ given by (10)–(11), which is a biologically reasonable [24] outer approximation. We further demonstrated that the coupling in protein concentrations due to competition for RNAP and ribosomes can be interpreted using isocost lines, a concept introduced in microeconomics to describe the attainable combinations of products having a limited budget. Finally, we presented the counter-intuitive phenomenon of lateral activation: inducing the expression of one protein can increase the production of a second one, by reallocating resources from a third, serving as a buffer for shared resources.

The presented simplex-based method aids the rational and predictable design of complex biocircuits as follows. First, it provides a simple tool to check whether the desired behavior of a large gene circuit is realizable. If the transcriptional/translational resources are insufficient, the simplexes could guide the partitioning of circuits into smaller components, each realizable within a single cell. Second, exploiting the analytic expressions of the simplexes, we can predict how tuning various biochemical parameters affect unwanted coupling, so that we can mitigate these adverse effects.

A natural extension of the results presented here is considering regulatory linkages among genes, thus enabling the description of how the limited availability of resources couples the expression of different proteins in arbitrary gene networks. We are further working on the extension of the presented framework to describe the dynamic behavior of gene circuits. A particularly interesting research direction is combining the results of [6], describing the effects of sharing transcription factors on the dynamics of modules, and the result presented here, characterizing the stationary effects of the limited availability of transcriptional and translational machinery. As a result, one could account for two of the major causes of context-dependence in systems and synthetic biology in a unified mathematical framework, allowing a more detailed understanding of natural systems, and the design of multi-module systems with predictable behavior.

REFERENCES

[1] S. Cardinale and A.P. Arkin, Contextualizing context for synthetic biology - identifying causes of failure of synthetic biological systems, *Biotechnol. J.*, vol. 7, no. 7, 2012, pp. 856-66.

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

[3] P.E.M. Purnick and R. Weiss, The second wave of synthetic biology: from modules to systems, *Nature Reviews Molecular Cell Biology*, vol. 10, no. 6, 2009, pp. 410-22.

[4] A.L. Slusarczyk, A. Lin and R. Weiss, Foundations for the design and implementation of synthetic genetic circuits, *Nature Reviews Genetics*, vol. 13, no. 6, 2012, pp. 406-20.

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

[6] A. Gyorgy and D. Del Vecchio, Modular composition of complex gene transcription networks, *PLoS Computational Biology*, vol. 10, no. 3, 2014, e1003486

[7] W.E. Bentley , N. Mirjalili, D.C. Andersen, R.H. Davis and D.S. Kompal, Plasmid-encoded protein: the principal factor in the "metabolic burden" associated with recombinant bacteria, *Biotechnology and Bioengineering*, vol. 35 no. 7, 1990, pp. 668-81.

[8] M. Scott, C.W. Gunderson, E.M. Mateescu, Z. Zhang and T. Hwa, Interdependence of cell growth and gene expression: origins and consequences, *Science*, vol. 330, no. 6007, 2010, pp. 1099-1102.

[9] N.A. Cookson, W.H. Mather, T. Danino, O. Mondragon-Palomino, R.J. Williams, L.S. Tsimring and J. Hasty, Queueing up for enzymatic processing: correlated signaling through coupled degradation, *Molecular Systems Biology*, vol. 7, 2011, 561.

[10] J. Vind, M.A. Sorensen, M.D. Rasmussen and S Pedersen, Synthesis of proteins in *Escherichia coli* is limited by the concentration of free ribosomes: expression from reporter genes does not always reflect functional mRNA levels, *J Mol Biol*, vol. 231, no. 3, 1993, pp. 678-88.

[11] G. Churchward, H. Bremer and R. Young, Transcription in bacteria at different DNA concentrations, *J. Bacteriol.*, vol. 150, 1982, pp. 572-81.

[12] W.H. Mather, J. Hasty, L.S. Tsimring and R.J. Williams, Translational cross talk in gene networks, *Biophysical Journal*, vol. 104. no. 11, 2013, pp. 2564-72.

[13] R. Algar, T. Ellis and G.-B. Stan, Modelling the burden caused by gene expression: an *in silico* investigation into the interactions between synthetic gene circuits and their chassis cell, *arXiv:1309.7798*, 2013.

[14] D. Siegal-Gaskins, V. Noireaux and R.M. Murray, Biomolecular resource utilization in elementary cell-free gene circuits, *In Proc of the American Control Conference*, 2013.

[15] D. De Vos, F.J. Bruggeman, H.V. Westerhoff and B.M. Bakker, How molecular competition influences fluxes in gene expression networks, *PLoS One*, vol. 6, 2011, e28494.

[16] M.A. Savageau, *Biochemical systems analysis. A study of function and design in molecular biology*, Addison-Wesley Publishing Co.; 1976.

[17] J.I. Jimenez, A. Gyorgy, H.-H. Huang, H. Chung, R. Weiss and D. Del Vecchio, Isocost lines describe the cellular economy of genetic circuits, *In Preparation*

[18] U. Alon, *An Introduction to systems biology - Design principles of biological circuits*, Chapman and Hall; 2006.

[19] T. Akerlund, K. Nordstrom and R. Bernander, Analysis of cell size and DNA content in exponentially growing and stationary-phase batch cultures of Escherichia coli, *J Bacteriol*, vol. 177, 1995, pp. 6791-97.

[20] S. Klumpp and T. Hwa, Growth-rate-dependent partitioning of RNA polymerases in bacteria, *PNAS*, vol. 105, 2008, pp. 20245-50.

[21] H. Bremer and P.P. Dennis, Modulation of chemical composition and other parameters of the cell by growth rate, *Escherichia coli and Salmonella: cellular and molecular biology*, vol. 2, 1996, pp. 1553-69.

[22] A.D. Tadmor and T. Tlusty, A coarse-grained biophysical model of *E. coli* and its application to perturbation of the rRNA operon copy number, *PLOS Computational Biology*, vo. 4, 2008.

[23] H. Bremer, P.P. Dennis and M. Ehrenberg, Free RNA polymerase and modeling global transcription in *Escherichia coli*", *Biochimie*, col. 85, no. 6, 2003, pp. 597-609.

[24] A. Gyorgy and D. Del Vecchio, Technical Report on Limitations and trade-offs in gene expression due to competition for shared cellular resources, DSpace@MIT, http://hdl.handle.net/1721.1/90249, 2014.

[25] R.S. Pindyck and D.L. Rubinfeld, *Microeconomics*. Pearson Practice Hall, NJ; 2005.