Quantifying Resource Competition and its Effects in the TX-TL System

Andras Gyorgy and Richard M. Murray

Abstract— Without accounting for the limited availability of shared cellular resources, the standard model of gene expression fails to reliably predict experimental data obtained *in vitro*. To overcome this limitation, we develop a dynamical model of gene expression explicitly modeling competition for scarce resources. In addition to accurately describing the experimental data, this model only depends on a handful of easily identifiable parameters with clear physical interpretation. Based on this model, we then characterize the combinations of protein concentrations that are simultaneously realizable with shared resources. As application examples, we demonstrate how the results can be used to explain similarities/differences among different *in vitro* extracts, furthermore, we illustrate that accounting for resource usage is essential in circuit design considering the toggle switch.

I. INTRODUCTION

One of the fundamental goals of synthetic biology is to engineer complex behaviors both at the cellular and population levels. Unfortunately, parts designed separately often fail to function once interconnected due to context-dependence. Sources of context-dependence include interactions among parts due to spatial co-localization [1], dependence on the host organism and strain [2], growth-dependence [3], enviromental dependence [4], and unwanted couplings due to the composition of modules [5], [6], [7]. As a result, while there are great successes in creating more and more complex circuits, these efforts often involve numerous iterative cycles of building and testing components.

While these steps are slow and expensive *in vivo*, they are significantly faster and cheaper *in vitro*. As a result, cell-free transcription-translation (TX-TL) systems offer a promising avenue for synthetic biology [8]. Additionally, these systems do not suffer from unwanted coupling between the synthetic parts and the behavior of the host organism, for instance, issues related to cell growth. Unfortunately, however, protein expression requires the availability of resources (RNAP, nucleotides, tRNAs, ribosomes, ATP, etc.) that are shared among genes, and as a result, protein expression levels are coupled even in the absence of regulatory linkages *in vitro* [9], just as *in vivo* [10], [11]. To reliably analyze and predict circuit behavior, these couplings need to be accounted for, otherwise circuits need to be continuously re-designed and re-tuned.

According to the most commonly used model of gene expression, mRNA m_i is transcribed and degraded with rate

constants α_i and δ_i, respectively, and protein x_i is translated and degraded with rate constants β_i and γ_i, respectively [12]. With this, mRNA and protein dynamics are given by

$$\dot{m}_i = \alpha_i D_i - \delta_i m_i \qquad \text{and} \qquad \dot{x}_i = \beta_i m_i - \gamma_i x_i, \quad (1)$$

where D_i denotes the concentration of DNA encoding the gene of protein x_i. Unfortunately, the above model shows poor agreement with experimental results (Fig. 1), as the model fails to account for the limited availability of shared transcriptional and translational resources.

To overcome this issue, we explicitly account for the limited availability of RNAP and ribosomes, as experimental evidence suggests that transcription and translation are limited by the availability of these resources [13], [14]. As a result, the first contribution of this paper is a model that can reliably predict protein expression with scarce resources even in the presence of multiple proteins, validated experimentally, extending the results of [9]. Building upon this model, the second contribution of this paper is the analytic characterization of protein concentrations that are simultaneously realizable with shared resources considering arbitrary network topology even during transient behavior, complementing the results in [10], [15]. We then demonstrate how the results presented here can be utilized to study and standardize cell-free extracts and to design genetic circuits.

Our work complements the continuing endeavour of the rational design of complex biocircuits. Notable examples focusing on the issue of shared resources study competition for transcription factors [16], for proteases [17], and for transcriptional and translational resources [9], [10]. A more general approach for studying the effects of resource

A. Gyorgy is with Department of Electrical Engineering and Computer Science, University of California, Berkeley, Berkeley CA 94704, USA `gyorgy@berkeley.edu`

R.M. Murray is with the Department of Control and Dynamical Systems, California Institute of Technology, Pasadena, CA 91125, USA `murray@cds.caltech.edu`
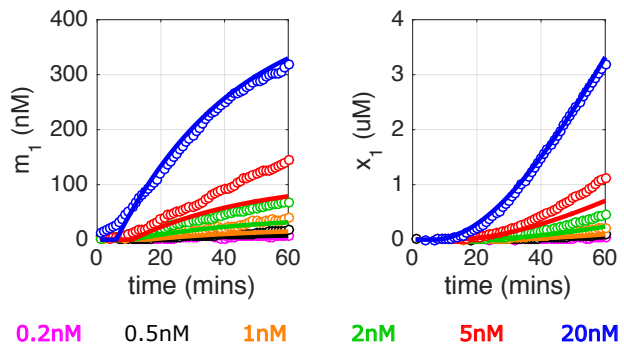
Fig. 1. Solid lines depict simulation results following parameter identification using `idnlgrey` and `pem` in MATLAB and the model in (1), empty circles represent measurement data from [9], colors indicate different DNA concentrations D_1 (see Figure 1 in [9] for experimental details).

competition is presented in [18] using Metabolic Control Analysis (MCA) [19], a tool that is closely related to Biochemical Systems Analysis [20]. MCA and its extensions compute sensitivities around the steady state, providing a local description. Conversely, the results presented here are not restricted to near-equilibrium behavior.

This paper is organized as follows. We first introduce a general mechanistic model for gene transcription networks accounting for the limited availability of RNAP and ribosomes and formulate the two main questions of the paper in Section II. We then answer these questions in Section III together with their experimental verification. Application examples are then considered in Section IV.

II. MODEL AND PROBLEM FORMULATION

After presenting the mathematical model of gene expression when shared resources are only available in limited amounts, we formulate the two main questions of the paper.

A. Mathematical Model of Gene Expression

Transcription of protein x_i is modeled by the reactions

$$b_i + p \underset{\kappa_i^-}{\overset{\kappa_i^+}{\rightleftharpoons}} c_i \qquad \text{and} \qquad c_i \xrightarrow{\alpha_i} b_i + p + m_i, \qquad (2)$$
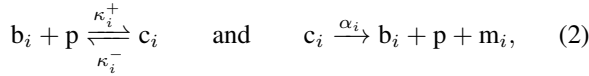
where b_i, c_i and m_i denote the promoter of x_i, the promoter bound with RNAP p, and the resulting transcript, respectively. Similarly, translation of protein x_i is described by the reactions

$$m_i + r \underset{k_i^-}{\overset{k_i^+}{\rightleftharpoons}} d_i \qquad \text{and} \qquad d_i \xrightarrow{\beta_i} m_i + r + x_i, \qquad (3)$$
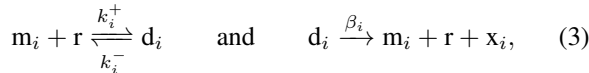
where d_i denotes the complex of the transcript bound to the ribosome r. Finally, dilution and degradation of mRNAs and proteins are captured by the reactions

$$m_i \xrightarrow{\delta_i} \emptyset, \qquad d_i \xrightarrow{\delta_i} r \qquad \text{and} \qquad x_i \xrightarrow{\gamma_i} \emptyset. \qquad (4)$$
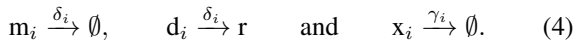
Note that proteases are missing from the standard TX-TL system and there is no dilution, so that $\gamma_i = 0$ generally [8].

From (2)–(4), the dynamics describing the expression of protein x_i are given by

$$
\begin{aligned}
\dot{b}_i &= -\left(\kappa_i^+ p b_i - \kappa_i^- c_i\right) + \alpha_i c_i, \\
\dot{c}_i &= \left(\kappa_i^+ p b_i - \kappa_i^- c_i\right) - \alpha_i c_i, \\
\dot{m}_i &= \alpha_i c_i - \delta_i m_i - \left(k_i^+ m_i r - k_i^- d_i\right) + \beta_i d_i, \qquad (5)\\
\dot{d}_i &= \left(k_i^+ m_i r - k_i^- d_i\right) - \beta_i d_i - \delta_i d_i, \\
\dot{x}_i &= \beta_i d_i - \gamma_i x_i,
\end{aligned}
$$

together with the conservation law $D_i = b_i + c_i$, where D_i is the total concentration of the DNA encoding protein x_i, assumed to be constant [21]. So far we considered the case when protein x_i is expressed constitutively. In the more general case when the expression of protein x_i is regulated, let $\epsilon_i(x) \in [0,1]$ denote the fraction of promoter accessible for RNAP (i.e., the fraction that is activated/not repressed). In this case, we have that $\epsilon_i(x) D_i = b_i + c_i$ instead of $D_i = b_i + c_i$, see [15] for details together with the specific form of $\epsilon_i(x)$ in the most common cases.

B. Shared Resources

Take the case of n proteins being simultaneously expressed. Let p_T and r_T denote the total concentration of RNAP and ribosomes, respectively. Considering (2)–(3), the concentration of RNAP sequestered for the transcription of m_i is c_i from (2), whereas the concentration of ribosomes allocated for the translation of x_i is d_i from (3). Consequently, we obtain the conservation laws

$$p_T = p + \sum_{i=1}^{n} c_i \qquad \text{and} \qquad r_T = r + \sum_{i=1}^{n} d_i \qquad (6)$$

for RNAP and ribosomes, respectively. It is through these conservation laws that the expression of different proteins x_i become coupled according to (5), even in the absence of regulatory linkages among them.

C. Problem Formulation

In this paper we address two main questions. First, we seek a model that reliably predicts protein expression levels in the presence of multiple proteins competing for the same resources such that it depends only on parameters that are easy to measure/identify. Second, based on this model, we wish to characterize the combinations of protein concentrations $x_i(t)$ for $i = 1, 2, \ldots, n$ and $t \geq 0$ that are simultaneously realizable with shared resources.

III. RESULTS

To address the two main questions formulated in the previous section, we first exploit the presence of multiple time-scales and the typical range of parameter values to obtain a reduced order model of (5)–(6).

A. Model Order Reduction and Parameter Identification

Introduce $\kappa_i = (\kappa_i^- + \alpha_i)/\kappa_i^+$ and $k_i = (k_i^- + \beta_i + \delta_i)/k_i^+$. Given that protein production and decay are much slower than binding and unbinding reactions [22], we have $\alpha_i \ll \kappa_i^-$ and $\beta_i, \delta_i \ll k_i^-$, so that $\kappa_i \approx \kappa^-/\kappa_i^+$ and $k_i \approx k_i^-/k_i^+$. Therefore, at the quasi-steady state of (5) we obtain

$$\dot{m}_i = \alpha_i c_i - \delta_i m_i - \delta_i d_i \qquad \text{and} \qquad \dot{x}_i = \beta_i d_i - \gamma_i x_i, \quad (7)$$

together with $0 = -\left(\kappa_i^+ p b_i - \kappa_i^- c_i\right) + \alpha_i c_i$ and $0 = \left(k_i^+ m_i r - k_i^- d_i\right) - \beta_i d_i - \delta_i d_i$, yielding

$$c_i = \epsilon_i D_i \frac{\frac{p}{\kappa_i}}{1 + \frac{p}{\kappa_i}} \qquad \text{and} \qquad d_i = \frac{m_i r}{k_i}. \qquad (8)$$

The dynamics of m_i and x_i for $i = 1, 2, \ldots, n$ are coupled through the solutions p and r of (6). From (6) it follows that $r = r_T/(1 + \sum_{j=1}^{n} \frac{m_j}{k_j})$, however, the analytic solution for p is not tractable for $n > 2$. To overcome this limitation, note that *in vivo* we have $p \ll \kappa_j$ and $r \ll k_j$ [10], [15], and since the *in vitro* TX-TL system is a diluted version of its *in vivo* counterpart, the concentrations p and r are even lower than *in vivo* [8]. Combining (8) with the approximations $p \ll \kappa_j$ and $r \ll k_j$, (7) becomes

$$
\begin{aligned}
\dot{m}_i &= \alpha_i c_i(x) - \delta_i m_i, \\
\dot{x}_i &= \beta_i d_i(m) - \gamma_i x_i,
\end{aligned}
\qquad (9)
$$

with $x = (x_1, x_2, \ldots, x_n)^T$, $m = (m_1, m_2, \ldots, m_n)^T$ and

$$c_i(x) = p_T \frac{\frac{D_i}{\kappa_i}\epsilon_i(x)}{1 + \sum_{j=1}^{n}\frac{D_j}{\kappa_j}\epsilon_j(x)}, \qquad (10)$$

$$d_i(m) = r_T \frac{\frac{m_i}{k_i}}{1 + \sum_{j=1}^{n}\frac{m_j}{k_j}}. \qquad (11)$$

While (9) bears resemblance to (1), it is fundamentally different as (10)–(11) capture the coupling in both transcription and translation due to shared resources. In particular, there is no coupling among mRNA concentrations if $\sum_{j=1}^{n} D_j/\kappa_j \ll 1$, and similarly, there are no unwanted interactions among protein expressions if $\sum_{j=1}^{n} m_j/k_j \ll 1$ additionally, and in this case, (9)–(11) simplifies to (1).

To validate (9)–(11), we consider the experimental data from [9]. First, we focus on the case of a single protein x_1 being expressed with different DNA concentrations D_1. Following parameter identification, the simulation results are presented in Fig. 2A, whereas the identified parameters can be found in Table I ($\gamma_1 = 0$ as proteins are not degraded/diluted). Since $m_1/k_1 < 0.04 \ll 1$ we expect that when neglecting competition for ribosomes, i.e., ignoring the denominator in (11), the resulting model should still reliably predict mRNA and protein concentrations, which is indeed the case (the results are indiscernable from those presented in Fig. 2A). Conversely, considering that D_1/κ_1 can be as high as 0.4, which is comparable to 1, we expect that neglecting RNAP sequestration, i.e., ignoring the denominator in (10), yields much worse predictive capabilities, evidenced by the results presented in Fig. 2B.

So far we only considered the expression of one protein. To further test the validity of (9)–(11), we next turn our attention to the case of two proteins expressed simultaneously: a reporter protein x_1 and a load protein x_2 (see Figure 5 in [9] for experimental details). The parameters of these two proteins are identified separately, see Table I for the obtained parameter values using the experimental data from [9]. The results in Fig. 3 correspond to the case when they are expressed together. Experimental data from [9] is depicted in blue, which is in good agreement with the simulation data obtained from (9)–(11) by considering both proteins (red curve). Conversely, the simulation data considering only the reporter protein x_1 without the load protein x_2 is depicted in green ($D_2 = 0$). The significant gap between this curve and the experimental data in blue underscores the importance of accounting for the limited availability of shared resources.
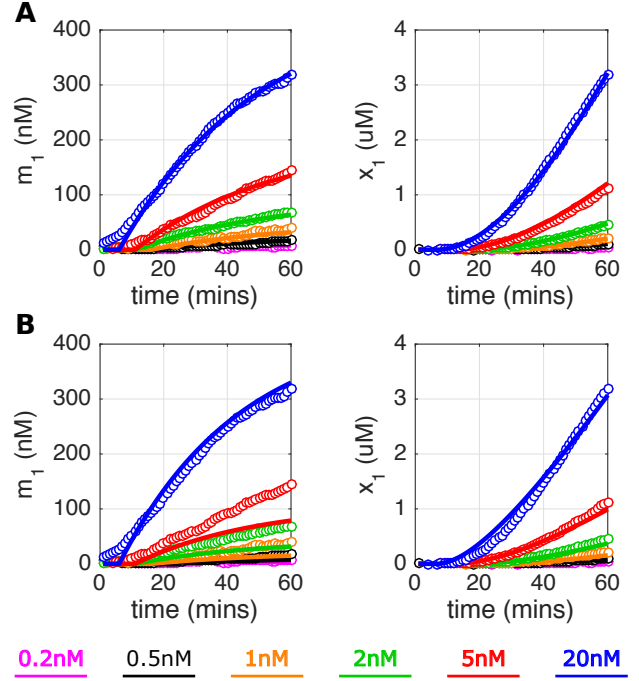

Fig. 2. Solid lines depict simulation results following parameter identification using `idnlgrey` and `pem` in MATLAB and the model in (9)–(11), empty circles represent measurement data from [9], colors indicate different DNA concentrations D_1 (see Figure 1 in [9] for experimental details). (A) Both RNAP and ribosome sequestration are modeled. (B) RNAP sequestration is neglected.
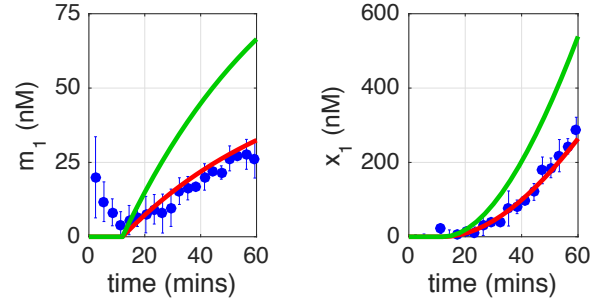

Fig. 3. Blue circles represent measurement data from [9] with standard deviation obtained from triplicates in the case of $D_1 = 2$nM and $D_2 = 10$nM (see Figure 5 in [9] for experimental details). The red curve denotes the simulation results from (9)–(11) considering both proteins with parameters in Table I, whereas the green curve corresponds to the case when only x_1 is expressed ($D_2 = 0$).

B. Simultaneously Realizable Protein Concentrations

Building upon the model in (9)–(11) reliably predicting protein expression levels with shared resources, we next focus on characterizing the protein concentrations that can be expressed simultaneously. To simplify notation, we write $\epsilon_i(t)$ instead of $\epsilon_i(x(t))$, and similarly, $c_i(t)$ and $d_i(t)$ instead of $c_i(x(t))$ and $d_i(m(t))$ from (10)–(11). Furthermore, since our main focus is the TX-TL system, we assume that $\gamma_i = 0$ for $i = 1, 2, \ldots, n$, corresponding to the case when proteins are not degraded/diluted [8].

TABLE I

PARAMETER VALUES OBTAINED VIA IDENTIFICATION

Parameter	Value	Unit	Parameter	Value	Unit
δ_1	0.97	1/h	δ_2	1.39	1/h
$\alpha_1 p_T$	610.20	nM/h	$\alpha_2 p_T$	843.07	nM/h
$\beta_1 r_T$	140.18	nM/h	$\beta_2 r_T$	3421.89	nM/h
κ_1	8.21	nM	κ_2	7.6786	nM
k_1	7.75	μM	k_2	213.69	μM

Proposition 1. *Let $\epsilon_i(t) \equiv \epsilon_i^* \in [0,1]$ for $i = 1, 2, \ldots, n$, and let $m_i(t)$ and $x_i(t)$ denote the solution of*

$$\dot{m}_i = \alpha_i c_i(t) - \delta_i m_i(t) \qquad m_i(0) = 0,$$
$$\dot{x}_i = \beta_i d_i(t) \qquad x_i(0) = 0, \qquad (12)$$

where $c_i(t)$ and $d_i(t)$ are given in (10)–(11). Then

$$x_i(t) \leq \beta_i r_T [t - g_i(t, \omega_i)] \qquad t \geq 0$$

with

$$\omega_i = \frac{\alpha_i p_T}{\delta_i k_i} \frac{\frac{D_i}{\kappa_i} \epsilon_i^*}{1 + \sum_{j=1}^n \frac{D_j}{\kappa_j} \epsilon_j^*},$$
$$g_i(t, \omega) = \frac{\ln[(1+w)e^{\delta_i t} - \omega]}{\delta_i (1 + \omega)}. \qquad (13)$$

If $\delta_i = \delta$ for $i = 1, 2, \ldots, n$, we further have

$$x_i(t) = \beta_i r_T \frac{\omega_i}{\sum_{j=1}^n \omega_j} \left[t - g_i\left(t, \sum_{j=1}^n \omega_j \right) \right] \qquad t \geq 0.$$

Proof. Define $\alpha_i^* = \frac{D_i}{\kappa_i} \epsilon_i^* / (1 + \sum_{j=1}^n \frac{D_j}{\kappa_j} \epsilon_j^*)$, so that $\dot{m}_i = \alpha_i^* - \delta_i m_i$ from (12), yielding

$$m_i(t) = \frac{\alpha_i^*}{\delta_i} \left(1 - e^{-\delta_i t} \right) \qquad t \geq 0. \qquad (14)$$

As $m_i(t) \geq 0$ for $t \geq 0$, from (12) we further have $\dot{x}_i \leq \beta_i r_T \frac{m_i}{k_i} / (1 + \frac{m_i}{k_i})$, so that from (14) we obtain that

$$x_i(t) \leq \beta_i r_T \int_0^t \frac{\frac{\alpha_i^*}{\delta_i k_i}(1 - e^{-\delta_i \tau})}{1 + \frac{\alpha_i^*}{\delta_i k_i}(1 - e^{-\delta_i \tau})} d\tau$$
$$= \beta_i r_T [t - g_i(t, \omega_i)] \qquad t \geq 0$$

with ω_i and $g_i(.,.)$ defined in (13).

Similarly, in the special case when $\delta_i = \delta$ for $i = 1, 2, \ldots, n$ we have from (12) that

$$x_i(t) = \beta_i r_T \int_0^t \frac{\frac{\alpha_i^*}{\delta k_i}(1 - e^{-\delta \tau})}{1 + (1 - e^{-\delta \tau}) \sum_{j=1}^n \frac{\alpha_j^*}{\delta k_j}} d\tau$$
$$= \beta_i r_T \frac{\omega_i}{\sum_{j=1}^n \omega_j} \left[t - g_i\left(t, \sum_{j=1}^n \omega_j \right) \right] \qquad t \geq 0.$$

$\square$

Proposition 1 provides us with two results. In the special case when the mRNA degradation rates are the same, we obtain the exact solution for $x_i(t)$. In the general case, however, we only obtain an upper bound for the trajectories. Importantly, even the upper bound turns out to be an extremely good approximation of $x_i(t)$, evidenced by Fig. 4A for the setup with the reporter protein x_1 and the load protein x_2 considered in Fig. 3. This is not surprising since we obtain the upper bound in the proof of Proposition 1 by considering the approximation $1 + \sum_{j=1}^n m_j / \kappa_j \ll 1 + m_i / \kappa_i$, which is expected to introduce little approximation error based on the identified parameters in Table I.

While Proposition 1 provides us with detailed knowledge about $x_i(t)$ individually for $i = 1, 2, \ldots, n$, it only applies to
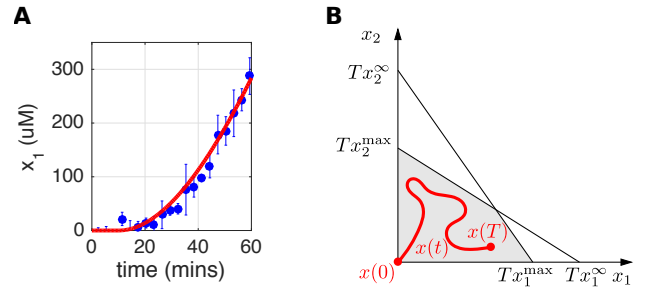

Fig. 4. Realizable protein concentrations. (A) Blue circles represent measurement data from [9] with standard deviation obtained from triplicates in the case of $D_1 = 2$nM and $D_2 = 10$nM (see Figure 5 in [9] for experimental details). The red curve denotes the upper bound $\beta_i r_T [t - g_i(t, \omega_i)]$ from Proposition 1. (B) With two proteins, the trajectory $x(t)$ of (15) is constrained within the grey region, given by the intersection of the simplexes $\mathcal{S}_1(T)$ and $\mathcal{S}_2(T)$ from (16).

the special case when each protein has a constant activation level (i.e., $\epsilon_i(t) \equiv \epsilon_i^*$ for $t \geq 0$ and $i = 1, 2, \ldots, n$). This is the case when proteins have no regulatory interactions among them, which is the case in the experimental setup in [9]. However, in the overwhelming majority of cases, we are interested in protein dynamics with regulatory interactions among them. Therefore, in what follows we focus on the case when the activation levels $\epsilon_i(t)$ are time-varying signals. To characterize the realizable protein concentrations in this case, we exploit the fact that protein dynamics are slower than mRNA dynamics [22], so that we have $m_i(t) = \alpha_i c_i(t) / \delta_i$ at the quasi-steady state of (9)–(11), together with

$$\dot{x}_i = \frac{A_i \epsilon_i(t)}{1 + \sum_{j=1}^n B_j \epsilon_j(t)} - \gamma_i x_i \qquad x_i(0) = 0, \qquad (15)$$

where $A_i = \frac{\beta_i \alpha_i D_i p_T r_T}{\delta_i \kappa_i k_i}$ and $B_j = \frac{D_j}{\kappa_j}(1 + \frac{\alpha_j p_T}{\delta_j k_j})$.

Proposition 2. *Consider $\epsilon_j(t)$ for $j = 1, 2, \ldots, n$ for $t \in [0, T]$ such that $\epsilon_j(t) \in [0, 1]$, and let $x(t)$ denote the corresponding solution of (15) when $\gamma_i = 0$ for $i = 1, 2, \ldots, n$. Then we have $x(t) \in \mathcal{S}(T) = \cap_{i=1}^n \mathcal{S}_i(T)$ with*

$$\mathcal{S}_i(T) = \left\{ x \mid x \geq 0 \text{ and } \frac{x_i}{x_i^{\max}} + \sum_{j=1, \, j \neq i}^n \frac{x_j}{x_j^\infty} \leq T \right\}, \quad (16)$$

where $x_i^{\max} = A_i / (1 + B_i)$ and $x_i^\infty = A_i / B_i$.

Proof. Since $x_i(t)$ increases monotonically from (15) when $\gamma_i = 0$, it is sufficient to prove that $x(T) \in \mathcal{S}(T)$ by showing that $x(T) \in \mathcal{S}_i(T)$ for $i = 1, 2, \ldots, n$. To this end, define

$$z_i = \frac{x_i}{x_i^{\max}} + \sum_{j=1, \, j \neq i}^n \frac{x_j}{x_j^\infty}, \qquad (17)$$

so that $z_i(0) = 0$. Furthermore, from (15) it follows that $\dot{z}_i(t) = 1 + [1 - \epsilon_i(t)]/[1 + \sum_{j=1}^n B_j \epsilon_j(t)]$, so that $\epsilon_j(t) \in [0, 1]$ for $j = 1, 2, \ldots, n$ yields $\dot{z}_i(t) \in [0, 1]$, thus $z_i(T) \in [0, T]$. From (17) we than obtain $x_i / x_i^{\max} + \sum_{j=1, \, j \neq i}^n x_j / x_j^\infty \leq T$, whereas $x_i(T) \geq 0$ follows directly from (15). Therefore, $x(T) \in \mathcal{S}_i(T)$ for $i = 1, 2, \ldots, n$, proving that $x(t) \in \cap_{i=1}^n \mathcal{S}_i(T)$ for $t \in [0, T]$. $\square$

According to Proposition 2, the protein concentrations that are simultaneously realizable are constrained within the intersection of the simplexes $\mathcal{S}_i(T)$ for $i = 1, 2, \ldots, n$. These simplexes and thus their intersection grow in time, see Fig. 4B. Since $x(t) \in \cap_{i=1}^{n} \mathcal{S}_i(T)$ for $t \in [0, T]$ holds for all $\epsilon_j(t) \in [0, 1]$ time-varying signals, thus for all possible underlying regulatory networks, in some cases the set $\cap_{i=1}^{n} \mathcal{S}_i(T)$ is conservative. However, it is not a trivial and overly conservative outer approximation, as it can also be extremely tight. To illustrate this, let $\epsilon_i(t) \equiv 1$ and $\epsilon_j(t) \equiv 0$ for $j \neq i$ and $t \in [0, T]$, yielding $x_i(T) = T x_i^{\max}$ and $x_j(T) = 0$, a point on the boundary of the set $\cap_{i=1}^{n} \mathcal{S}_i(T)$.

IV. APPLICATION EXAMPLES

Next we illustrate how the quantification of resource usage can be leveraged throughout two application examples.

A. Comparing TX-TL Extracts

We demonstrated in Fig. 2 and Fig. 3 that we can reliably predict both mRNA and protein expression levels using the model in (9)–(11). Additionally, the parameters that we identified (see Table I) can be used to characterize and compare parts, and more importantly, different TX-TL extracts. To illustrate this, consider Fig. 5, in which we compare two different extracts: 'e16' (blue) and 'e27' (orange) expressing the construct Pr-deGFP-MGapt from [9].

Let 'TL peak' denote the highest protein concentration achieved, and let 'TL peak time' be the time to reach 90% of this value starting from zero ('TX peak' and 'TX peak time' are defined similarly for mRNA levels). Looking at 'TL peak' and 'TL peak time', one might conclude that 'e16' and 'e27' behave similarly, except that 'e27' runs longer, thus the higher 'TL peak'. Considering 'TX peak time' further suggests a strong similarity between the two extracts, the only difference being the fact that 'e27' runs longer. Looking at 'TX peak', however, there is a stark difference: while 'e27' runs considerably longer, it produces far less mRNA.

To understand how these two extracts are different, we consider the parameters that we identified (Fig. 5). In particular, κ is smaller in 'e27', so that the promoter behaves as if it was stronger than in 'e16'. However, αp_T is much smaller in 'e27' than in 'e16', so that transcription is much slower (less RNAP and/or smaller transcription rate) in 'e27' than in 'e16'. As for translation, the two extracts are indeed similar, as $\beta r_T / k$ characterizing translational speed is almost identical. Summarizing the above, while 'e16' and 'e27' behave almost identical at the translational level, there is a stark difference at the transcriptional level, revealed by the identified parameters using the model in (9)–(11).

B. Rational Design of Circuits

From (15), the dynamics of the toggle switch from [23] can be written for $i, j = 1, 2$ as

$$\frac{dx_i}{dt} = \frac{A_i \epsilon_i(x_j)}{1 + B_1 \epsilon_1(x_2) + B_2 \epsilon_2(x_1)} - \gamma_i x_i \qquad i \neq j \quad (18)$$

with $\epsilon_i(x_j) = K_{D,i}/(K_{D,i} + x_j^2)$, where $K_{D,i}$ denotes the dissociation constant of x_j binding to the promoter of x_i.
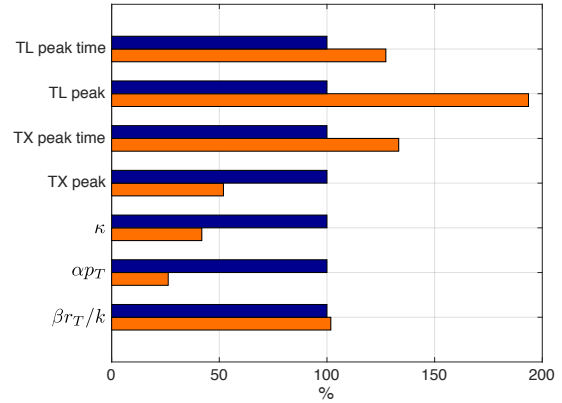

Fig. 5. Parameter values obtained by expressing the circuit Pr-CFP-MGa (see [9] for details) in two extracts: 'e16' (blue) and 'e27' (orange), both obtained by following the extract preparation protocol detailed in [8]. Parameter values are normalized so that the ones corresponding to 'e16' represent 100%.

To simplify the notation and analysis, we focus on the case when $K_{D,1} = K_{D,2} = K_D$, $A_1 = A_2 = A$, $B_1 = B_2 = B$ and $\gamma_1 = \gamma_2 = \gamma$, so that introducing

$$z_i = \frac{\gamma}{A} x_i, \qquad K = \left(\frac{\gamma}{A}\right)^2 K_D \qquad \text{and} \qquad \tau = t\gamma \quad (19)$$

yields the dimensionless description for $i, j = 1, 2$

$$\frac{dz_i}{d\tau} = \frac{\epsilon_i(z_2)}{1 + B[\epsilon_1(z_2) + \epsilon_2(z_1)]} - z_i \qquad i \neq j. \quad (20)$$

Neglecting resource usage ($B = 0$) yields the standard dimensionless description for $i, j = 1, 2$

$$\frac{dz_i}{d\tau} = \epsilon_i(z_j) - z_i \qquad i \neq j. \quad (21)$$

Proposition 3. *The system in* (20) *is bistable if and only if* $K < \frac{1}{4(B+1)^2}$.

Proof. The system in (20) has five fixed points. Two of them are always imaginary, one is always real, and the other two are real if and only if $K < \frac{1}{4(B+1)^2}$. Taking the Jacobian at these latter two fixed points, its eigenvalues are negative, so that the corresponding fixed points are stable nodes.

As for the third fix point, we use index theory to show that it is a saddle. Consider the square $[0, 1/(1 + B)]^2$. Along its sides, the vector field (20) is pointing inwards, so that its index is $I = +1$. The square $[0, 1/(1 + B)]^2$ contains the two fixed points given above, both having index $+1$. Consequently, the third fixed point must be contained in the square $[0, 1/(1 + B)]^2$, and it must have index -1, so that it is a saddle. Choosing $\mu = K(B + 1)^2$ as bifurcation parameter, the system (20) goes through a supercritical pitchfork bifurcation at $\mu = 1/4$. Therefore, (20) is bistable if and only if $K < \frac{1}{4(B+1)^2}$. □

Proposition 4. *The system in* (21) *is bistable if and only if* $K < \frac{1}{4}$.

Proof. Similar to the proof of Proposition 3. □

According to Propositions 3–4, the toggle switch with and without resource usage is bistable if the dimensionless dissociation constant K is sufficiently small. However, this threshold is not the same for (20) and (21). Considering (19), while for (21) it is sufficient to have

$$K_D < \frac{1}{4}\left(\frac{A}{\gamma}\right)^2, \qquad (22)$$

for (20) bistability is achieved only if

$$K_D < \frac{1}{4}\left(\frac{A}{\gamma}\right)^2 \frac{1}{(1+B)^2}. \qquad (23)$$

For instance, considering Table I, we obtain that $1/(1 + B)^2 \approx 0.05$. That is, while without accounting for resource usage we would conclude that K_D needs to satisfy only (22), in reality, it needs to be about 20 times smaller (tighter repression) so that it satisfies (23) to yield a bistable toggle.

V. CONCLUSION

In this paper, we characterized the effects of the limited availability of transcriptional and translational resources on mRNA and protein expression in the cell-free TX-TL system. The main motivation of this work is the apparent discrepancy between the prediction of the standard model in (1) and the experimental data in Fig. 1.

To overcome this limitation, we first derived the model in (9)–(11) that accounts for the limited availability of RNAP and ribosomes. This model reliably predicts both the expression level of a single protein (Fig. 2) and that of a reporter protein expressed simultaneously together with a load protein (Fig. 3). Second, we characterized the combination of protein concentrations that are simultaneously realizable with shared resources (Fig. 4). Both these results contribute to realizing the full potential of the TX-TL system as a platform for rapid prototyping and characterization of synthetic gene circuits.

Additionally, we demonstrated that by identifying both extract and part parameters using (9)–(11), we can compare various extracts and understand how similar/different they are. As a result, batch-to-batch variation can be minimized, furthermore, different extracts can be chosen for different purposes with desirable properties (e.g., more translational capacity at the expense of less transcriptional capacity). Complementing this, we illustrated that accounting for competition for shared resources is essential when designing synthetic circuits. In particular, we demonstrated that for the toggle switch to become bistable, we need to employ significantly stronger promoters than what would be suggested by models neglecting the effects of shared resources.

A natural extension of the results presented here is describing the second phase of translation/transcription, i.e., after the peak in transcription (see [8] for details). To this end, it is essential to understand why there is a transition between these two phases, and which factors determine the timing of this transition. Additionally, the results presented here, especially the parameter values that can be identified using the model in (9)–(11), can serve as guides for fine-grained modeling of the reactions including more details, such as ATP and nucleotides [24].

REFERENCES

[1] R.S. Cox, M.G. Surette and M.B. Elowitz, Programming gene expression with combinatorial promoters, *Mol. Syst. Biol.*, vol. 3, 2007.

[2] F.K. Balagadde, L. You, C.L. Hansen, F.H. Arnold and S.R. Quake, Long-term monitoring of bacteria undergoing programmed population control in a microchemostat, *Science*, vol. 309, 2005, pp. 137-40.

[3] 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, 2010, pp. 1099-1102.

[4] J. Neupert, D. Karcher and R. Bock, Design of simple synthetic RNA thermometers for temperature-controlled gene expression in *Escherichia coli*, *Nucleic Acids Res.*, vol 36, 2008, pp. e124.

[5] E. Franco, E. Friedrichs. J. Kim, R. Jungmann, R.M. Murray, E.Winfree and F.C. Simmel, Timing molecular motion and production with a synthetic transcriptional clock", *PNAS*, vol. 108, 2011.

[6] D. Del Vecchio, A.J. Ninfa and E.D. Sontag, Modular cell biology: retroactivity and insulation, *Mol. Syst. Biol.*, vol. 4, 2008.

[7] A. Gyorgy and D. Del Vecchio, Modular composition of gene transcription networks, *PLoS Comp. Biol.*, vol. 10, pp. e1003486.

[8] M.K. Takahashi, C.A. Hayes, J. Chappell, Z.S. Sun, R.M. Murray, V. Noireaux and J.B. Lucks, Characterizing and prototyping genetic networks with cell-free transcription-translation reactions, *Methods*, vol. 86, 2015, pp. 60-72.

[9] D. Siegal-Gaskins, Z.A. Tuza, J. Kim, V. Noireaux and R.M. Murray, Gene Circuit Performance Characterization and Resource Usage in a Cell-Free "Breadboard", *ACS Synth. Biol.*, vol. 3, 2014, pp. 41625.

[10] A. Gyorgy, J.I. Jimenez, J. Yazbek, H.-H. Huang, H. Chung, R. Weiss and D. Del Vecchio, Isocost lines describe the cellular economy of genetic circuits, *Biophys. J.*, vol. 109, 2015, pp. 639-46.

[11] F. Ceroni, R. Algar, G.B. Stan and T. Ellis, Quantifying cellular capacity identifies gene expression designs with reduced burden, *Nature Methods*, vol. 12, 2015, pp. 415–18.

[12] D. Del Vecchio and R.M. Murray, *Biomolecular Feedback Systems*, Princeton University Press, 2014.

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

[14] 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, 1993, pp. 678-88.

[15] A. Gyorgy and D. Del Vecchio, "Limitations and trade-offs in gene expression due to competition for shared cellular resources", *IEEE Conference on Decision and Control*, 2014.

[16] R.C. Brewster, F.M. Weinert, H.G. Garcia, D. Song, M. Rydenfelt and R. Phillips, The transcription factor titration effect dictates level of gene expression, *Cell*, vol. 156, 2014, pp. 1312-23.

[17] 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, *Mol. Syst. Biol.*, vol. 7, 2011.

[18] 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, pp. e28494.

[19] D.A. Fell, Metabolic control analysis: a survey of its theoretical and experimental development, *Biochem. J.*, vol. 286, 1992, pp. 313-30.

[20] M.A. Savageau, *Biochemical systems analysis: a study of function and design in molecular biology*, AddisonWesley, 1976.

[21] 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. 6791-7.

[22] U. Alon, *An introduction to systems biology – Design principles of biological circuits*, Chapman and Hall, 2006.

[23] T.S. Gardner, C.R. Cantor and J.J. Collins, Construction of a genetic toggle switch in *Escherichia coli*, *Nature*, vol. 403, 2000, pp. 339-42.

[24] Z.A. Tuza, V. Singhal, J. Kim and R.M. Murray, "An in silico modeling toolbox for rapid prototyping of circuits in a biomolecular breadboard system", *IEEE Conference on Decision and Control*, 2013.