Submitted, OUP Synthetic Biology, May 2019
http://www.cds.caltech.edu/~murray/preprints/ham19-oupsynbio_s.pdf

Analysis of primitive genetic interactions for the design of a genetic signal differentiator.

Wolfgang Halter[1,*] Richard M. Murray[2] Frank Allgöwer[1]

[1] Institute for Systems Theory and Automatic Control, University of Stuttgart, 70569 Stuttgart, Germany and [2] California Institute of Technology, Pasadena, CA 91125, USA

*Corresponding author: E-mail: wolfgang.halter@ist.uni-stuttgart.de

Abstract

We study the dynamic and static input output behavior of several primitive genetic interactions and their effect on the performance of a genetic signal differentiator. In a simplified design, several requirements for the linearity and time-scales of processes like transcription, translation and competitive promoter binding were introduced. By experimentally probing simple genetic constructs in a cell-free experimental environment and fitting semi-mechanistic models to these data, we show that some of these requirements can be verified, while others are only met with reservations in certain operational regimes. Analyzing the linearized model of the resulting genetic network we conclude that it approximates a differentiator with relative degree one. Taking also the discovered non-linearities into account and using a describing function approach, we further determine the particular frequency and amplitude ranges where the genetic differentiator can be expected to behave as such.

Key words: genetic circuit design; combinatorial promoters; signal differentiator

1 Introduction

The systematic design of functional genetic circuits is one of the key challenges in the field of synthetic biology. Usually, the goal is to add a desired function to a cellular organism. As the complexity of these functions has been increasing steadily [1], it becomes increasingly difficult to design the topology of the genetic network and decide what kind of genetic interactions to use. One way to approach this synthesis problem is by adapting methods from the fields of systems and control theory [2], e.g. by starting with a description of the desired part as a linear transfer function, finding the necessary fundamental input/output functions which realize this transfer function and then realizing the evolving network topology with primitive genetic interactions. The key to this approach is to determine how fundamental linear I/O functions like gain, integrator, sum and difference can be realized using only primitive genetic interactions such as transcription, translation, combinatorial promotors, post-transcriptional modification or pairwise interactions of DNA, mRNA or protein molecules.

This design workflow follows the ideas of [3], where the authors showed that any arbitrary *linear* input/output system can be realized exactly using only zeroth and first order biochemical reactions. We adressed the question of replacing the zeroth and first order biochemical reactions with general genetic interactions in [4]. Therein, several requirements were introduced to conclude that the processes of transcription and translation can be interpreted as gain and integration respectively and that combinatorial promoters may be used to realize the difference of two concentrations. In [4], and also in this work, we use these results to design a genetic signal differentiator, i.e. a genetic part whose output indicates the temporal derivative of its input. Such a module would be of particular interest in context of a genetic PID controller that could be used to regulate production processes within a cell. While for this purpose the genetic realization of the more important integral feedback has been studied extensively [5, 6, 7, 8, 9, 10], differential operators in a biological context have been investigated rather sporadically [11, 12] and have only recently moved into the focus of synthetic biology [13]. In latter work, the authors introduce a differentiator module based on mechanisms borrowed from the *E. coli* chemotaxis regulatory network. This mechanism is based on active enzyme-like degradation and the assumption that this degradation operates at saturation of the enzyme. In contrast to the results of [13], the topology presented in [4] is not based on a known biological exam-

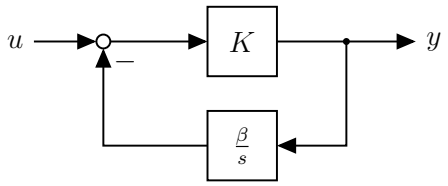

Figure 1: Ideal approximation of a differentiator, from [4].

ple but is derived from scratch, using an adjusted version of the general design framework of [3]. This leads to a differentiator module of similar complexity but different assumptions and requirements which need to be guaranteed.

In this work, we combine control theoretic concepts, mathematical models and observations from experiments to verify and adapt the requirements introduced in [4]. We find that, in cell-free extract, transcription can be considered as a PT1 element, i.e. a delayed gain, while translation indeed can be seen as an integrator. Further, we show that combinatorial promoters are not very well suited to realize the difference of two signals and that the dynamics are very much dependent on the operation conditions. Lastly, we study how not meeting the requirements affects the performance of the genetic signal differentiator and reveal the operating conditions under which the differentiator behaves as expected and where this is not the case.

In the following, we first introduce the desired signal differentiator, one possible topology to realize this part and the necessary requirements for primitive genetic interactions by recapitulating the results established in [4]. After, we introduce mathematical models of protein synthesis as well as the cell-free experimental environment which is used to generate the experimental data. Subsequently, the requirements on time-scales and linear operation regimes of the processes of transcription and translation are verified by fitting the model to a series of experimental data and analyzing the resulting parameters, leading to transfer function representations of these two processes. Using another series of experiments, we determine the input-output steady-state map of a combinatorial promoter and discuss the limited capability of such promoters to realize the difference of two signals. Finally, the impact of the discovered discrepancies on the performance of the genetic differentiator is studied both in time and frequency domain, using a describing function approach for the latter.

2 Background

First, we briefly recapitulate the results from [4] before we analyze, verify and adjust the requirements we introduced therein.

In the field of control theory, one can study linear systems in two different domains. First, in the time domain, by looking at the states of a system and the temporal derivatives thereof which define a system of ordinary differential equations (ODEs). And second, in the frequency domain, by looking at transfer functions which are complex valued functions and describe how different frequency components of an input signal are modified by a system. These two domains are connected via the Laplace transformation and particularly the frequency domain is very useful for the design and analysis of linear systems. An ideal differentiator would be given by the transfer function $G(s) = s$ with Laplace variable s. However, as is well known in the control community, an exact realization of such an ideal differentiator is not possible due to the lack of causality. For a system to be causal, its output must not depend on future values of the input signal. This is not the case for the differentiator. In case the system is given in form of a rational transfer function, i.e. $G(s) = \frac{N(s)}{D(s)}$, one can easily check for this property by examining the degrees of the polynomials $N(s)$ and $D(s)$: causality is given if the degree of $N(s)$ is not bigger than the degree of $D(s)$.

The desired function thus can only be approximated, e.g. by adding an additional low-pass filter to the ideal differentiator, leading to the desired transfer function

$$G(s) = \frac{Ks}{s + K} \tag{1}$$

where K is the bandwidth of the filter. One possibility to realize this transfer function is by the circuit depicted in Fig. 1, with a (preferably large) gain K in the forward path and a weighted integrator in the feedback path. Ideally, one chooses $\beta = 1$ to recover (1). Thus, in order to approximate the differentiator, three basic functions are needed: a gain, an integrator and the signal difference between input and feedback. Finding genetic realizations of these basic functions is the main challenge in designing the differentiator. In particular, it is expected that this cannot be achieved in an exact way, thus it is necessary to determine how inaccuracies in the basic parts influence the behavior of the assembled circuit. For an initial guess for finding

such functions, a semi-mechanistic model of transcription and translation [14] was used in [4] to conclude that the processes of transcription and translation can approximately be seen as a gain and integrator respectively and that a combinatorial promoter may be used to realize the difference of two signals. In the remainder of this section, we briefly recapitulate these deductions.

In the process of protein synthesis, the genetic information is read from DNA (with concentration D_i) and transcribed into mRNA (M_i), then, mRNA molecules are translated into proteins (P_i). In the following, the subscript i stands for the i-th gene (G_i) in a network with I distinct genes. With

$$\mathbf{P} = \begin{bmatrix} P_1 & \dots & P_I \end{bmatrix}^\top$$

representing all proteins present in the genetic network, the dynamics of mRNA and protein concentrations of gene i are described by

$$\dot{M}_i = f_i(\mathbf{P}, \Theta_i, \Phi) - p_i(M_i, \Theta_i, \Phi) \tag{2a}$$

$$\dot{P}_i = g_i(M_i, \Theta_i, \Phi) - q_i(P_i, \Theta_i, \Phi) \tag{2b}$$

where $f_i(\mathbf{P}, \Theta_i, \Phi)$ and $g_i(M_i, \Theta_i, \Phi)$ are the respective production and $p_i(M_i, \Theta_i, \Phi)$ and $q_i(P_i, \Theta_i, \Phi)$ the respective degradation rates. These rates are possibly dependent on protein and mRNA concentrations, certain gene specific parameters $\Theta_i \in \mathbb{R}^N$ like DNA concentrations (D_i) or initiation and degradation rates, as well as several environmental parameters $\Phi \in \mathbb{R}^L$ which include, among others, the total amount of RNA polymerase (RNAP), ribosomes and endonucleases, the transcription and translation elongation rates, and other host dependent variables. For better readability the arguments Θ_i and Φ are omitted in the remainder.

In [4], we introduced the topology depicted in Fig. 2 as one approach to realize the transfer function (1). Therein, the input is considered to be a transcription factor, i.e. $u = P_u$, which activates gene G_1 and inhibits another gene G_2. Each of these genes produces a transcription factor which suppresses its own production. While G_1 has the purpose of capturing positive gradients of the input signal, G_2 is designed to capture negative ones. The output of the part is then given as the difference between the mRNA concentrations of the two genes, i.e.

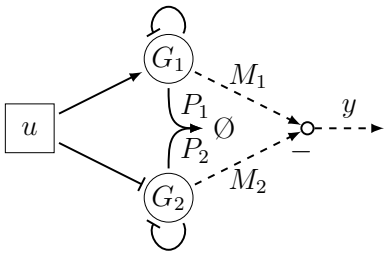

Figure 2: Genetic differentiator: Genes G_1 and G_2 tracking positive and negative slopes of u. Proteins produced by G_1 and G_2 neutralize each other. Difference of associated mRNAs indicate output y.

$y = M_1 - M_2$. Further, for the purpose of a minimal signal representation, the transcription factors P_1 and P_2 undergo an annihilation reaction.

Several simplifications and requirements for the processes of transcription, translation and degradation were introduced to finally arrive at the desired model equations

$$\dot{M}_1 = \alpha D_1 \cdot \kappa(P_u - P_1) - \nu_1 M_1 \tag{3a}$$

$$\dot{P}_1 = \beta M_1 - \delta_1 P_1 - \delta_{12} P_1 P_2 \tag{3b}$$

$$\dot{M}_2 = \alpha D_2 \cdot \kappa(k_0 - P_u - P_2) - \nu_2 M_2 \tag{3c}$$

$$\dot{P}_2 = \beta M_2 - \delta_2 P_2 - \delta_{12} P_1 P_2 \tag{3d}$$

with the function

$$\kappa(x) = \begin{cases} x & x > 0 \\ 0 & x \leq 0 \end{cases} \tag{4}$$

assuring strictly positive transcription rates. In the following, we focus on G_1, the gene for capturing positive gradients, and recapitulate the requirements for the biological processes necessary to arrive at (3). Subsequently, the connection between (3) and (1) will be discussed. We note that the focus on G_1 is without any loss of generality as the following requirements can be adjusted with minimal effort to arrive at the equations for G_2.

Requirement 1. *M_1 and P_1 are subject to first order degradation, i.e.*

$$p_1(M_1) = \nu_1 M_1 \tag{5a}$$

$$q_1(P_1) = \delta_1 P_1. \tag{5b}$$

with degradation rate constants $\nu_1, \delta_1 \in \Theta_1$.

Although degradation rates p_i and q_i are usually dependent on protease and endonuclease levels we require first order degradation dynamics to assure linearity with respect to mRNA and protein levels.

Requirement 2. *The operation regime is such that f_1 and g_1 are both approximately linear in D_1 and M_1, respectively.*

This requirement is rectified by results like the ones presented in [15], where particularly the linearity of g_i in M_i is shown. Alternatively, similar simplifications have been applied by following a linearization approach as pursued in [16]. In general, however, although the transcription rate f_i increases monotonically with DNA concentration D_i, it cannot grow arbitrarily large but is subject to saturation effects for large enough DNA or transcription factor concentrations, see e.g. [14, 17].

Requirement 3. *There exists a combinatorial promoter which is piecewise linear in two inputs, such that*

$$f_1\left([P_u, P_1]^\top\right) \sim \kappa\left(P_u - P_1\right)$$

with $\kappa(\cdot)$ like in Eq. (4).

With this requirement, we demand that the combined effect of the two transcription factors is proportional to the difference of their concentrations, as long as $P_u > P_1$, and zero, otherwise. In other words, f_1 as a function of $[P_u, P_1]^\top$, needs to fulfill the fundamental additivity property of linear functions in the regime $P_u > P_1$. This further means that, as we are considering a combinatorial promoter, P_u has to act as an activator for G_1 while P_1 acts as an inhibitor. Consequently, instead of forming the difference between input P_u and integral feedback P_1

by using direct interactions between the two species, we move the difference operation to the promoter function.

Now if Requirements 2 and 3 hold, we find

$$f_1([P_u, P_1]^\top) \approx \alpha D_1 \cdot \kappa(P_u - P_1) \tag{6a}$$

$$g_1(M_1) \approx \beta M_1, \tag{6b}$$

where α and β stand for lumped production rate parameters. Thus, with Requirements 1 to 3, we arrive at the first part of Eq. (3). Note that, when considering both genes G_1 and G_2, this means that the transcription and translation rate constants α and β are assumed to be equal for both genes. Also, it is required that $P_u > P_1$ for the part to work properly. For this reason, the annihilation reaction between P_1 and P_2 was introduced, see [4] for more details.

Finally, concerning an appropriate choice of parameters, another requirement can be deduced from typical degradation rates given e.g. in [18].

Requirement 4. *The degradation of mRNA is much faster than the one of protein, i.e.* $\nu_1 \gg \delta_1$.

With that in mind, one can apply a quasi steady state approximation of the mRNA dynamics and further assume that $\delta_1 \approx 0$ to arrive at

$$\tilde{M}_1 \approx \frac{\alpha}{\nu_1} D_1(P_u - P_1)$$
$$\dot{P}_1 \approx \beta \tilde{M}_1$$

where $\tilde{M}_i$ stands for the steady state mRNA concentration. Thus, we conclude that the process of transcription can be interpreted as a gain while translation approximately realizes an integrator. With the signal entering the transcription process chosen as the residual of input P_u and integral feedback P_1, the presented model thus realizes Eq. (1).

In [4], we verified this structure by simulating the system based on the much more detailed model described in [14]. This detailed model mainly aims at taking the finite amounts of RNAP and ribosomes as well as the time delay of transcription and translation into account, however, chosen parameters only reflected average parameters from literature. Further, saturation effects

and nonlinearities of the promoter dynamics were neglected.

After recapitulating the results of [4] and realizing the limitations of the used models, we now adjust our modeling approach and focus on analyzing and verifying the requirements by conducting a series of experiments using a cell-free experimental system [19].

3 Materials and methods

In this section, a brief overview on the experimental technique as well as the subsequently used models is provided.

3.1 TX-TL experimental platform

For the purpose of establishing a reliable, efficient and fast prototyping environment for genetic circuits, various cell-free TX-TL systems have been developed and optimized during the past decade [19, 20, 21, 22, 23, 24]. The main advantages of cell-free over classical cell based *in-vitro* systems are that cellular systems impose certain physical constraints on the gene circuits and the incorporation of the desired genes is comparably time consuming. Cell-free extracts on the other hand provide a well reproducible platform for rapid testing of arbitrary gene circuits. Such an extract for instance can be produced from *Escherichia coli* (*E. coli*) bacteria by bead-beating cell resuspensions, see [23] for more details on the production of *E. coli* extract. As DNA formatting and transformation as well as cell growth are thus decoupled from the actual testing of the circuit, testing cycles can be speed up significantly from several days for testing in original cells to only a few hours for testing in cell-free extract.

However, regeneration of resources required for mRNA and protein synthesis is an issue in cell-free environments, which is why the dynamics of mRNA and protein production are subject to some overlayed degradation dynamics of the extract. Therefore, the experiments are only meaningful for a limited experiment duration and we only consider observations within the first 200 minutes after initiation of the experiment. However, even in this limited time frame, degradation of resources will be visible in the experimental data. Since this mechanism is not considered in the mathematical models, the identified parameters will be biased. Production parameters tend to be underestimated while degradation parameters tend to be overestimated.

For every TX-TL experiment, the DNA subject to testing is suspended in water and mixed

with cell extract and an energy buffer. This buffer contains amino acids, NTPs, tRNAs and other small molecules necessary for mRNA and protein synthesis. The reaction volume was chosen to 5μL. Usually, one or more genetic constructs encode a fluorescent reporter protein such as GFP. After initialization of the experiment, the mixture is incubated at 29°C inside a Biotek plate reader, which assesses the level of fluorescent protein every few minutes. While the concentration of a fluorescent protein like GFP can be assessed directly, measuring the amount of mRNA requires an additional mechanism. We therefore make use of the malachite green dye (20μM) and a corresponding aptamer sequence (MGapt) which is added to the 3′ untranslated region (UTR) of the gene. The dye binds to a binding pocket of this sequence and changes its emission properties upon binding, therefore again enabling us to monitor a fluorescence signal which is proportional to the mRNA concentration [25]. However, measurements of the mRNA signal due to binding of the malachite green dye revealed only a poor signal to noise ratio, therefore an additional data pre-processing step was introduced by fitting a Gaussian process to the experimental data. Details on the pre-processing procedure can be found in Supplementary Data A.

In this work, we distinguish between gene and extract specific parameters. Gene specific parameters include variables like the affinity of the particular promoter sequence towards RNAP and other proteins and by definition are considered to be independent of the environment the experiment is conducted in, i.e. hold in different batches of cell-extract as well as inside living cells. In contrast, remaining parameters like the concentration of RNAP or transcription and translation elongation rates are denoted as extract or environment dependent, thus may vary even between different batches of cell-free extract. The experiments presented in this work have all been conducted using the same batch of TX-TL extract.

All genetic parts were originally given as plasmids. Using polymerase chain reactions and appropriate primer sequences, only the relevant linear double-stranded gene sequence was extracted from these plasmids and used in the TX-TL experiments. By addition of protein gamS, the degradation of linear DNA is prevented [26]. Information about the used genetic constructs can be found in Supplementary Data B.

3.2 Modeling protein synthesis

Throughout this work, different promoters are discussed and analyzed for various purposes. Therefore, the different mechanisms and modeling framework used for simulating the temporal evolution of mRNA and proteins are introduced. We therein build uppon the dynamics given in Eq. (2), however avoid using as strict simplifications as the ones outlined in Section 2.
In the following, complexes of two chemical species A and B are denoted with $A{:}B$ and conserved quantities are indicated by a bar, e.g. $\overline{R}$, the total amount of RNAP.

It is a well established result [18, 27], that the production rate of mRNA f_i is proportional to the concentration of promoter which is bound to a corresponding RNAP holoenzyme and not blocked by any inhibitors, e.g.

$$f_i(\mathbf{P}) = \alpha \cdot D_i{:}R{:}\sigma_{70}(\mathbf{P}, \Theta_i, \Phi) \tag{7}$$

where the concentration of complex $D_i{:}R{:}\sigma_{70}$ may be depending on other proteins $\mathbf{P}$, gene specific parameters Θ_i and extract specific parameters Φ.
In this example, sigma factor 70 (σ_{70}) first has to bind to RNAP to form the holoenzyme before this complex then binds the promoter region. The sigma factor therein has a very high specificity towards certain promoters, enabling the cell to switch between different transcriptional programs depending on which sigma factor is expressed. Note that compared to (6a), this is a more realistic model for mRNA production but prohibits making the same deductions for the genetic differentiator.

The basic mechanisms of interest for us are binding and unbinding reactions happening at the promoter sequence of DNA. Usually, as in [18, 28], the amount of $D_i{:}R{:}\sigma_{70}$ is approximated by Michaelis-Menten like equations, assuming that either DNA or RNAP holoenzyme is in abundance. In contrast to that, we won't make this assumption but particularly take the binding and unbinding reactions into account in order to consider both competition for shared cellular resources and saturation effects at the promoter. For simple setups where only self-competition occurs, we derive a closed form expression for the steady state concentration of the respective biochemical complexes.

3.2.1 Holoenzyme formation

When RNAP R is bound to a sigma factor σ_x, this complex is referred to as the RNAP holoenzyme. As discussed briefly in the previous section, such a holoenzyme binds to the promoter sequence of a gene and initiates the transcription process. Therefore, sigma factors are a crucial component for this process and without the right sigma factor, transcription cannot initiate. According to [29], RNAP alone is sufficient for transcription elongation, however, initiation requires sigma factors. We therefore assume that the formation of holoenzyme is independent of the holoenzyme binding to the promoter sequence, meaning that sigma factor and RNAP can bind and unbind irrespective of the fact if RNAP is bound to DNA or not.

We therefore have to consider the reactions

$$R + \sigma_x \underset{k^-_{\sigma x}}{\overset{k^+_{\sigma x}}{\rightleftharpoons}} R{:}\sigma_x \tag{8a}$$

$$D_i{:}R + \sigma_x \underset{k^-_{\sigma x}}{\overset{k^+_{\sigma x}}{\rightleftharpoons}} D_i{:}R{:}\sigma_x \tag{8b}$$

for each sigma factor and DNA species present in the system in order to account for the competition for RNAP. To simplify (8), we introduce

$$\overline{R{:}\sigma_x} = R{:}\sigma_x + \sum_i D_i{:}R{:}\sigma_x$$

$$X{:}R = R + \sum_i D_i{:}R$$

the total amount of R bound to σ_x as well as the total amount of R which is *not* bound to its respective sigma factor. Then, (8) can be combined to

$$X{:}R + \sigma_x \underset{k^-_{\sigma x}}{\overset{k^+_{\sigma x}}{\rightleftharpoons}} \overline{R{:}\sigma_x}. \tag{9}$$

In most cases, only dissociation constants

$$K_{\sigma x} = \frac{k^-_{\sigma_x}}{k^+_{\sigma_x}}$$

are identifiable and it is assumed that binding reactions are fast compared to the transcription elongation steps and thus in quasi steady state. Therefore, for notational simplicity, we will

reduce the notation to using dissociation constants instead of on and off rates in the remainder of this work.

Note that in (9) $X{:}R$ and σ_x denote both the unbound chemical species. If only one sigma factor is present in the system, the amount of $\overline{R{:}\sigma_x}$ can be calculated analytically as a function of the dissociation constant $K_{\sigma x}$ and the total amounts of RNAP and sigma factor respectively, viz. by application of the following proposition.

Proposition 1. *Given the entities A, B and $A{:}B$ and the reaction*

$$A + B \overset{K}{\rightleftharpoons} A{:}B.$$

If none of the entities participates in any other chemical reaction, the steady state of $A{:}B$ can be expressed in terms of the total amounts of A and B as

$$A{:}B = \frac{1}{2}\left(K + \overline{A} + \overline{B} - \sqrt{(K + \overline{A} + \overline{B})^2 - 4\overline{A}\overline{B}}\right) \tag{10}$$

with $\overline{A} = A + A{:}B$ and $\overline{B} = B + A{:}B$.

The proof can be found in Supplementary Data C. It is noted that usually, i.e. for the deduction of Michaelis-Menten kinetics, it is assumed that either $\overline{A} \gg \overline{B}$ or $\overline{B} \gg \overline{A}$ holds while Proposition 1 gives exact solutions for any values of $\overline{A}$ and $\overline{B}$. In cases when only a single sigma factor is present and its total concentration is constant over the time course of the experiment, we will later on use the amount $\overline{R{:}\sigma_x}$ as a fitting parameter and omit the binding reaction in order to reduce the complexity of the fitting problem. However, in cases where the concentration of sigma factor varies over time, we either use the exact formula from Proposition 1, or if there is more than one sigma factor, we directly implement the binding reactions as fast reactions and accept the increased computational complexity.

3.2.2 Promoter binding

After formation of $R{:}\sigma_x$, the RNAP holoenzyme binds to the promoter sequence and starts transcribing the information encoded as DNA. A promoter is called constitutive , if this binding

of RNAP happens spontaneously and is not influenced by any activators or inhibitors, i.e.

$$D_i + R{:}\sigma_x \overset{K_{iH}}{\rightleftharpoons} D_i{:}R{:}\sigma_x. \tag{11}$$

In such cases, given that the promoter does not interact with other holoenzymes, Proposition 1 can be applied again to simplify the modeling formalism.

In contrast to a constitutive promoter, binding of RNAP can also be inhibited by other proteins, leading to a combinatorial promoter with competitive binding mechanism, i.e. by the additional reaction

$$D_i + P_j \overset{K_{ij}}{\rightleftharpoons} D_i{:}P_j \tag{12}$$

which now competes with (11).

3.2.3 Translation and degradation rates

Similarly to the transcription rate (7), the rate of translation is given by

$$g_i(M_i) = \beta \cdot M_i{:}Q(M_i, \Theta_i, \Phi), \tag{13}$$

where $M_i{:}Q$ stands for the concentration of ribosomes (Q) bound to the ribosome binding site of mRNA M_i. We assume unregulated ribosomal binding and that the ribosome binding site sequences used for the constructs are of equal strength. Thus, the reactions for forming the complex $M_i{:}Q$ are the same as for the formation of holoenzyme and consequently, in case of only one mRNA species present, Proposition 1 can be applied again. Whenever more than one mRNA species is considered, competition for ribosomes occurs and binding reactions are implemented.

Degradation of mRNA and protein is mainly influenced by third party molecules such as endonucleases (E) and proteases. It is known [30] that latter species is quasi non-existent in TX-TL extract, thus we keep the first order degradation for proteins as in (5b). Endonucleases, on the other hand, are present in limited quantities, thus loading effects need to be considered. We explicitly assume that the binding of ribosomes and endonucleases is independent of each other, i.e. can be seen as two distinct processes where ribosomes and endonucleases do not

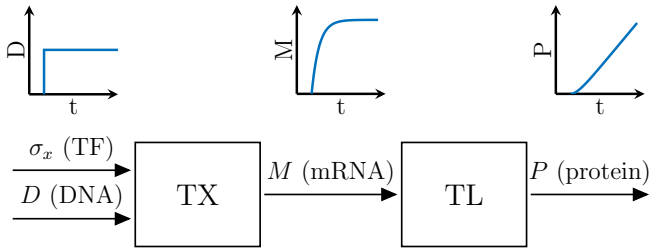

Figure 3: Scheme of probing protein synthesis with step in DNA and expected responses.

compete for mRNA. Thus, once more we define

$$p_i(M_i) = \gamma \cdot M_i{:}E(M_i, \Theta_i, \Phi) \tag{14}$$

and apply Proposition 1 whenever only self-competition occurs.

4 Results

Given the foundational work summarized in Section 2, it is yet unclear to what extent Requirements 1 to 4 can be verified. In particular, we are interested in answering the question of whether the processes of transcription and translation indeed can be regarded as a gain and integrator respectively (Requirements 1, 2 and 4) and further, whether one can find a suitable combinatorial promotor which satisfies all linearity requirements in order to verify Requirement 3.

4.1 I/O behavior of transcription and translation

First, we analyze the time-scales and linearity of transcription and translation. Therefore, the input-/output (I/O) behavior of these processes are characterized by experimentally probing a simple gene with different input steps as depicted schematically in Fig. 3. By observing the response to different step sizes in the input, the non-linearity of the promoter dynamics can be identified. The gene we study is equipped with a σ_{70} dependent constitutive promoter and expresses GFP. By fitting a suitable model to the experimental data and analyzing the corresponding parameters, Requirements 1 and 4 will be verified.

4.1.1 Experimental setup

There are two possibilities to realize a step-like input of varying height at the transcriptional level using promoters like introduced in Section 3.2: either by varying the amount of sigma factor (i.e. the transcription factor) while keeping the DNA concentration constant, or alternatively, changing the DNA concentration itself. While varying DNA amounts is straightforward, the sigma factor input additionally required purified protein which may be biologically unstable and is more difficult to obtain than DNA.

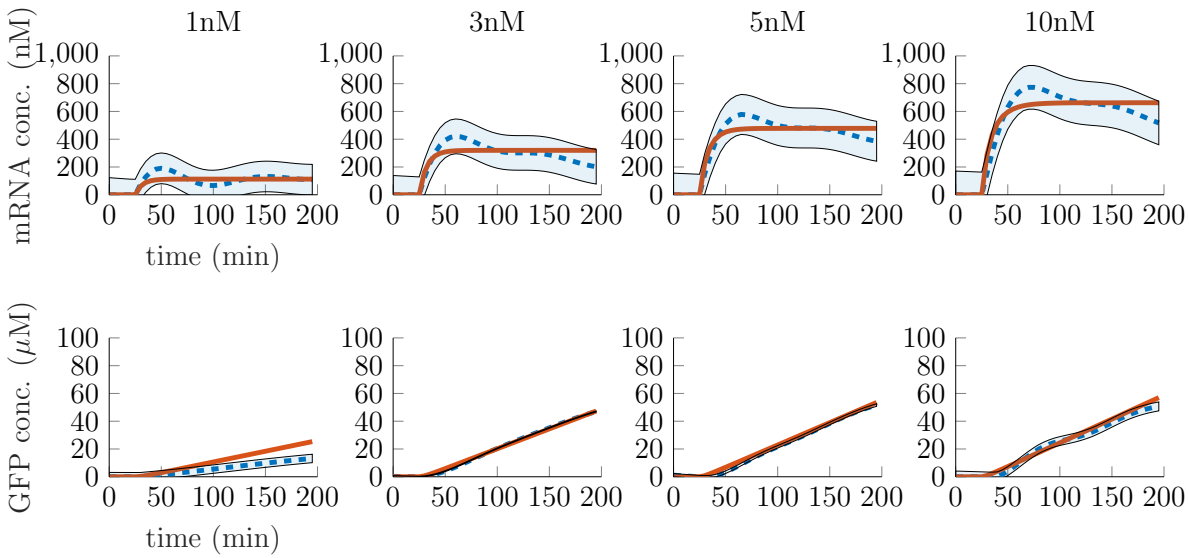

Figure 4: Mean and 95% confidence interval of experimental step-responses (blue, dotted mean, shaded confidence interval) and simulated step responses of the fitted nonlinear model (red, solid).

Depending on the choice of input, i.e. sigma factor or DNA, different dynamical effects can be expected when probing the system with steps of different height. As discussed before in Section 3.2, the mRNA production rate is proportional to the complex $D_i{:}R{:}\sigma_{70}$, wherein the concentration depends on the total amounts of DNA, RNAP and sigma factor. In case the concentration of sigma factor is considered as input, the corresponding model needs to incorporate both the formation of holoenzyme as well as the binding of holoenzyme to the DNA. Thus both binding rates would need to be considered. In contrast, when varying the DNA concentration, the binding reaction of holoenzyme can be neglected and the amount of total holoenzyme $\overline{R{:}\sigma_{70}}$ can be introduced instead.

This approach reduces the complexity of the fitting problem by focusing on the identifi-

Table 1: Values of the parameters obtained by fitting the nonlinear model to step response data.

parameter	unit	value	description
α	min^{-1}	21.54	transcription rate const.
β	min^{-1}	2.35	translation rate const.
γ	min^{-1}	0.18	mRNA deg. const.
δ	min^{-1}	1.19e−8	protein deg. const.
K_{1H}	nM	0.82	dissoc. const. for D_1 and $R{:}\sigma_{70}$
K_{MQ}	nM	72.26	dissoc. const. for M_1 and Q
K_{ME}	nM	102.20	dissoc. const. for M_1 and E
$\overline{R{:}\sigma_{70}}$	nM	4.26	total RNAP holoenzyme
$\overline{Q}$	nM	165.94	total ribosomes
$\overline{E}$	nM	650.30	total endonuclease

cation of promoter binding kinetics only. Thus, for the identification of the I/O behavior of transcription and translation, we first limit ourselves to step inputs in form of varying DNA concentrations and study the sigma factor dependent holoenzyme formation in a separate experiment, discussed in Section 4.2.

We choose four different DNA concentrations for probing the system: 1nM, 3nM, 5nM and 10nM. Three technical replicates were conducted. The data obtained by this process is depicted in Fig. 4. Therein, blue dashed lines stand for the mean of mRNA (upper column) and protein (lower column) concentrations and the 95% confidence intervals are illustrated as shaded blue regions respectively.

4.1.2 Corresponding model

We denote the index of the gene under study with $i = 1$ and accordingly the amount of GFP with P_1. According to Section 3.2 and particularly Eqs. (2), (5b), (7), (13) and (14), the corresponding model is determined by the complexes

$$\begin{aligned} D_1{:}R{:}\sigma_{70} &= D_1{:}R{:}\sigma_{70}\left(\overline{D}_1, \overline{R{:}\sigma_{70}}, K_{1H}\right) \\ M_1{:}Q &= M_1{:}Q\left(\overline{M}_1, \overline{Q}, K_{MQ}\right) \\ M_1{:}E &= M_1{:}E\left(\overline{M}_1, \overline{E}, K_{ME}\right) \end{aligned}$$

which are calculated using Proposition 1, depending on the total amounts of DNA, mRNA, RNAP holoenzyme, ribosomes and endonucleases as well as the respective dissociation con-

stants. We again note that the model can capture the dynamics only in a limited time frame as the degradation of extract is not taken into account. For fitting the model to the given data, we introduce a maximum likelihood objective function, see e.g. [31], and apply several rounds of both patternsearch and fmincon optimization algorithms implemented in Matlab. The resulting parameters given in Table 1 give rise to the red trajectories depicted in Fig. 4.

For the process of translation, we observe that the protein degradation rate δ is evaluated to be of magnitude 10^{-8} and therefore, compared to γ, practically zero.

Conclusion 1. *As required in Requirement 4, the degradation of mRNA is much faster than the one of protein.*

In order to check the linearity Requirements 1 and 2, we study the entities $D_1{:}R{:}\sigma_{70}$, $M_1{:}E$ and $M_1{:}Q$ as functions of the fitted parameters over the relevant range of DNA and mRNA concentrations as depicted in Fig. 5. This way, one can visualize the non-linear nature of the production reactions of mRNA and protein as well as the degradation of mRNA. Although these results clearly indicate that the processes of transcription and translation do not behave linearly in their inputs in general, they allow us to define operation regimes as those required in Requirement 2, i.e. where the linearity requirement holds at least approximately.

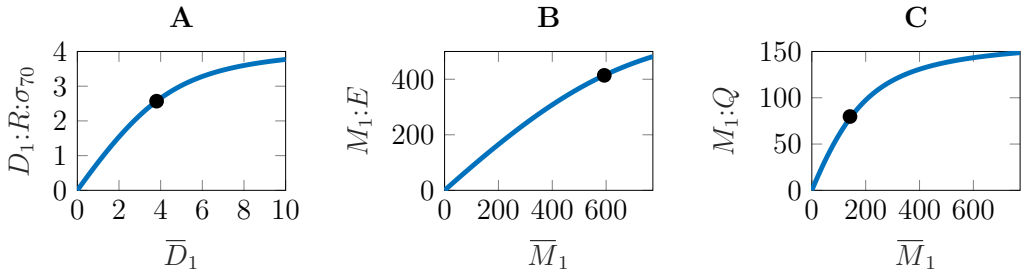

Figure 5: Amount of active complexes for transcription (A), mRNA degradation (B) and translation (C) over relevant range of DNA and mRNA respectively.

In that sense, we now introduce a relative measure of nonlinearity and define the ϵ-linear-range of a function $f : \mathbb{R} \to \mathbb{R}$ as the largest interval $[0, \xi^\star]$ for which this nonlinearity measure is just ϵ. For the nonlinearity measure we follow the methods introduced in [32]. Let $\|f(x)\|_{L_2[0,\xi]}$ be the truncated L2 norm of $f(x)$, defined by

$$\|f(x)\|_{L_2[0,\xi]} = \sqrt{\frac{1}{\xi}\int_0^\xi f(x)^2 dx}.$$

To approximate f, we use the linear function mx. Note that we forced the intercept of the linear function to take the value 0 to assure strictly positive values of the linear function on the interval $[0, \xi]$. For a given f, the best linear approximation in the interval $[0, \xi]$ is then found as the argument $m = m^\star$ which minimizes

$$L(\xi, m) = \|(f(x) - mx)\|_{L_2[0,\xi]}, \tag{15}$$

the absolute L2 norm of the residual between function $f(x)$ and the linear function mx. The value of $L(\xi, m^\star)$ now can be seen as an absolute measure for the nonlinearity of f on the interval $[0, \xi]$, however, this measure depends on the magnitude of the function f. Thus, in order to compare this measure across different functions, we normalize (15) by the L2 norm of f, i.e.

$$L_{\text{rel}}(\xi, m) = \frac{\|(f(x) - mx)\|_{L_2[0,\xi]}}{\|f(x)\|_{L_2[0,\xi]}}$$

to find our relative measure of nonlinearity.

Consequently, $\xi^\star$ is found as the solution of

$$\begin{aligned} \max \quad & \xi \\ \text{s.t.} \quad & \min_m L_{\text{rel}}(\xi, m) \leq \epsilon. \end{aligned} \tag{16}$$

In the given case, when one allows for a 5% error, i.e. $\epsilon = 0.05$, one obtains the linear ranges indicated as black points in Fig. 5.

Conclusion 2. *Linearity of production and degradation terms, as requested in Requirements 1 and 2, can be verified with* 95% *accuracy with*

$$\begin{aligned} D_1{:}R{:}\sigma_{70} &\approx A_{tx} \cdot \overline{D_1} && \text{for } \overline{D_1} \in [0, 3.805] \\ M_1{:}E &\approx A_{deg} \cdot \overline{M_1} && \text{for } \overline{M_1} \in [0, 593.1] \\ M_1{:}Q &\approx A_{tl} \cdot \overline{M_1} && \text{for } \overline{M_1} \in [0, 141.7] \end{aligned}$$

and $A_{tx} = 0.726$, $A_{deg} = 0.752$, $A_{tl} = 0.602$.

4.1.3 Linearized model and transfer functions

Given the linear operation regimes indicated in Conclusion 2, one can now derive linear models for transcription and translation which are then valid in the respective regimes. In the control community, the standard approach to approximate a nonlinear model with a linear one is to locally linearize the nonlinear function at on specific value. In case of the nonlinear mRNA degradation rate p_1 for example, a linearization around some fixed value $\overline{M}_1^0$ would yield

$$p_1(\overline{M}_1) \approx p_1(\overline{M}_1^0) + \left.\frac{dp_1}{d\overline{M}_1}\right|_{\overline{M}_1^0} \cdot (\overline{M}_1 - \overline{M}_1^0).$$

This approach assures that the linear function evaluated at $\overline{M}_1^0$ has the same value as the original nonlinear one, and that the difference between the two functions is small in a neighborhood around $\overline{M}_1^0$. Thus, the quality of the linear model on a certain interval strongly depends on the chosen value $\overline{M}_1^0$. In our case, particularly the values of $\overline{M}_1$ may vary across a wide range. Further, it should be made sure that in the case when neither DNA nor mRNA or protein is present, the temporal derivatives of these species also is equal to zero, i.e. that

$$\dot{\overline{M}}_1(\overline{D}_1 = 0, \overline{M}_1 = 0) = \dot{\overline{P}}_1(\overline{M}_1 = 0, \overline{P}_1 = 0) = 0$$

holds. This will only be achieved if all linear functions go through the origin. To assure this, one would consequently have to perform the linearization at $\overline{D}_1 = \overline{M}_1 = \overline{P}_1 = 0$, leading to potentially large deviations between the linear and nonlinear models at larger values of the independent variables. Therefore, instead of using this standard approach, we directly use the approximations of Conclusion 2 where we already made sure that the linear approximation is as good as possible over a given interval of the independent variable.

We thus obtain the linear model

$$\dot{\overline{M}}_1 \approx \alpha A_{\mathrm{tx}} \overline{D}_1 - \gamma A_{\mathrm{deg}} \overline{M}_1$$
$$\dot{\overline{P}}_1 \approx \beta A_{\mathrm{tl}} \overline{M}_1 - \delta \overline{P}_1$$

and when defining $\overline{D}_1$ and $\overline{M}_1$ as input and output of the transcription module, $\overline{M}_1$ and $\overline{P}_1$ as input and output of the translation module, the corresponding transfer functions

$$G_{\text{tx}}(s) = \frac{\alpha A_{\text{tx}}}{s + \gamma A_{\text{deg}}} \tag{17}$$

$$G_{\text{tl}}(s) = \frac{\beta A_{\text{tl}}}{s + \delta} \tag{18}$$

are obtained. We conclude that due to the fact that δ is very small, translation can indeed be seen as integration as long as DNA and mRNA concentrations are in the appropriate operation regime. However, the initial assumption that transcription can be seen as a gain needs to be adjusted as mRNA degradation cannot be neglected, leading to a PT1 element instead of a gain.

So far, we studied and characterized time-scales and linearity of the processes of transcription and translation in context of an *E. coli* cell-free extract and mainly focused on possible limitations caused by the promoter and mRNA binding kinetics. We therefore bypassed nonlinear effects of RNAP holoenzyme formation by changing DNA concentrations instead of using σ_{70} as input and found that at least during the first 200 minutes of a TX-TL experiment, resource limitations do have an effect on transcription, translation and mRNA degradation. By studying different step responses, the linear operation regimes were identified. We now turn towards inhibitor binding dynamics and in particular towards the problem of how to realize a signal difference using combinatorial promoters.

4.2 Signal difference and combinatorial promoters

In order to approximate the derivative of a signal by implementing the scheme depicted in Fig. 1, we remember that the input into the gain (i.e. transcription) has to be the residual between the reference and feedback signal.

There are various ways to realize a signal difference in biology, a widely used one being sequestration-based mechanisms between the signaling molecules, e.g. binding and degradation of the complex like elaborated in [7, 8]. When dealing with RNA or DNA, such a mechanism can be realized in a straight-forward way by e.g. the use of antisense strands. When it comes to proteins or metabolites, engineering a sequestration mechanism for an arbitrary protein or

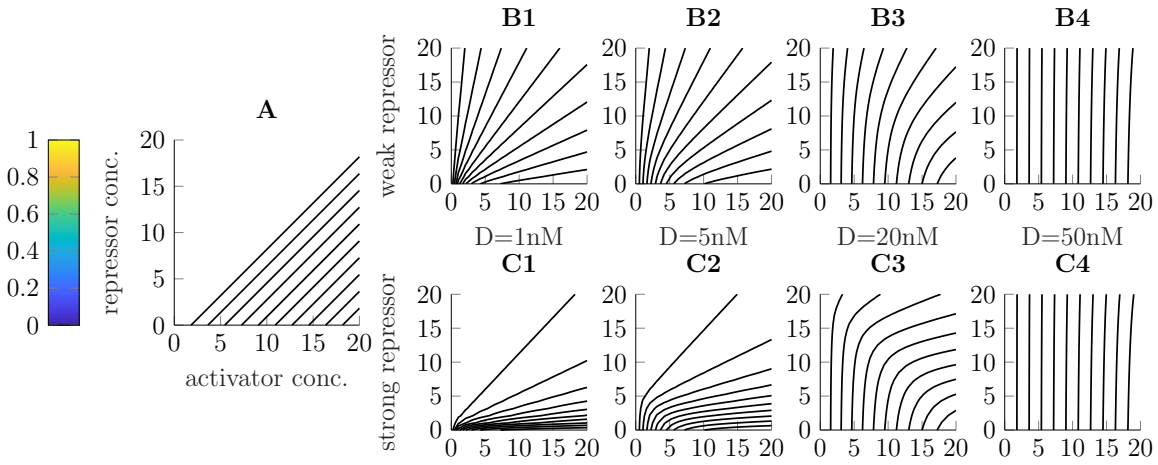

Figure 6: Level sets of promoter activity $D_i{:}R{:}\sigma_x$ over varying levels of sigma factor and inhibitor. **A** Desired behavior for non-negative signal difference. **B1-B4** Simulated values for varying DNA concentrations under a weak repressor. **C1-C4** Simulated values for varying DNA concentrations under a strong repressor.

metabolite may be possible but in general is more challenging. Thus, one is rather restricted to the use of existing pairs of proteins which undergo binding reactions, e.g. sigma factors and anti sigma factors. Combinatorial promoters as an alternative mechanism may offer a higher flexibility during the prototyping process as various inhibitor operator sequences are already known for transcriptional regulation. Therefore, it is in principle possible to compare the concentrations of any two transcription factors by combination of these operator sequences with different promoters. It is one of the goals of this work to investigate whether this approach can actually be used for the purpose of subtraction in a biological context.

Following such an approach, the desired behavior of the steady state of promoter dynamics is depicted in Fig. 6**A** where the steady state of $D_i{:}R{:}\sigma_x$ is color-coded over varying concentrations of $(\overline{\sigma_x})$ and inhibitor $(\overline{P}_j)$. Due to non-negativity of concentrations, no activity is desired whenever the concentration of inhibitor exceeds the one of activator (upper left triangle resembling zero). Otherwise, it is aspired that $D_i{:}R{:}\sigma_x$ is proportional to the difference $\overline{\sigma_x} - \overline{P}_j$, illustrated by the parallel and equidistant level sets in Fig. 6**A**.

Applying Proposition 1 and assuming that the total amount of RNAP holoenzyme is fixed, the amount of $D_i{:}R{:}\sigma_x$ depends on the chosen DNA concentration as well as on dissociation constants K_{iH} and K_{ij} of the RNAP holoenzyme and inhibitor respectively. If for instance we assume that $K_{iH} = K_{ij} = 1$ and look at the relative amount of activated DNA $D_i{:}R{:}\sigma_x/\overline{D}_i$, varying holoenzyme and inhibitor in the same range results in qualitatively different steady-

Table 2: Values of the parameters obtained by fitting the nonlinear model to the time-series responses of the combinatorial promoter.

parameter	unit	value	description
$K_{\sigma 70}$	nM	1.8e−6	dissoc. const. for R and σ_{70}
$K_{\sigma 28}$	nM	5.3e−3	dissoc. const. for R and σ_{28}
K_{tetR}	nM	8.1e−3	dissoc. const. for D_2 and $tetR$
K_{aTc}	nM	2.74	dissoc. const. for $tetR$ and aTc
K_{2H}	nM	1.084e4	dissoc. const. for D_2 and R:σ_{28}
K_{s28H}	nM	33.86	dissoc. const. for D_{s28} and R:σ_{70}
K_{tetRH}	nM	2.97e3	dissoc. const. for D_{tetR} and R:σ_{70}
$\bar{R}$	nM	283.14	total RNAP
$\bar{\sigma}_{70}$	nM	3.36	total sigma factor 70

state maps depending on how much $\overline{D}_i$ is chosen, as depicted in Fig. 6 **B1**-**B4**. For high DNA, sigma factor acts quasi linearly on the promoter while the inhibitor does not play a role at all. On the other hand, for small amounts of DNA, the inhibitor has a large effect and distorts the steady-state map such that the level sets converge to each other at the origin. Also, suppression due to the repressor does not seem strong enough as in all cases, D_i:R:$\sigma_x \gg 0$ for $\sigma_x < P_j$.

In contrast to that, Fig. 6 **C1**-**C4** show the same conditions, except that now $K_{iH} = 10 \cdot K_{ij}$, i.e. the inhibitor binds 10 times stronger to the promoter than RNAP holoenzyme does. In that case, only minimal transcriptional activity is expected when there is less sigma factor than repressor. Further, although level sets are curved, for medium amounts of DNA, e.g. 20nM, they are comparably equidistant and the steady-state map is almost symmetric.

This means that, while we have to acknowledge that exact realization of the difference of two signals is not possible with combinatorial promoters, some crucial properties can be approximated by choosing dissociation constants and DNA amounts carefully.

For that purpose and also for detangling the RNAP holoenzyme binding reaction, we study a a gene with a *pTar* intitiation sequence combined with a *tetO* inhibitor operator which expresses GFP. The *pTar* promoter is sensitive towards an RNAP holoenzyme consisting of RNAP bound to σ_{28}, while the operator sequence *tetO* enables binding and inhibition through Tet repressor proteins (tetR). We denote the concentration of this gene as D_2 and GFP concentration as P_2. To avoid usage of purified protein, both σ_{28} and $tetR$ are produced in the TX-TL system from respective constitutive (i.e. σ_{70} dependent) DNAs D_{s28} and D_{tetR}. While the amount of $\overline{D}_{s28}$ is varied to achieve different activation levels, inhibition is influenced by adding different amounts of anhydrotetracycline (aTc) which binds to $tetR$ and thus alleviates its association with the

promoter. The concentration of $\overline{D}_{tetR}$ is kept at a constant level of 1nM. The combinatorial promoter then produces GFP, dependent on the concentrations of σ_{28} and unblocked $tetR$. The time-series of this experiment can be found in Supplementary Data D.

According to the experimental setup, several chemical species compete for the same resources, thus Proposition 1 cannot be applied anymore and the binding reactions themselves had to be implemented as fast reactions. For brevity reasons, the binding reactions are not listed here. We focus on mRNA and protein dynamics, i.e. the ODEs

$$\begin{aligned}
\dot{\overline{M}}_{s28} &= \alpha \cdot D_{s28}{:}R{:}\sigma_{70} - \gamma \cdot M_{s28}{:}E \\
\dot{\bar{\sigma}}_{s28} &= \beta \cdot M_{s28}{:}Q - \delta \cdot \bar{\sigma}_{s28} \\
\dot{\overline{M}}_{tetR} &= \alpha \cdot D_{tetR}{:}R{:}\sigma_{70} - \gamma \cdot M_{tetR}{:}E \\
\dot{\overline{tetR}} &= \beta \cdot M_{tetR}{:}Q - \delta \cdot \overline{tetR} \\
\dot{\overline{M}}_{2} &= \alpha \cdot D_{2}{:}R{:}\sigma_{28} - \gamma \cdot M_{2}{:}E \\
\dot{P}_{2} &= \beta \cdot M_{2}{:}Q - \delta \cdot P_{2}.
\end{aligned}$$

Fitting these equations to the data, we obtain the parameters listed in Table 2 and the trajectories depicted in Supplementary Data D. In the fitting process, the optimization is constrained such that the amount of complex $\overline{R{:}\sigma_{70}}$ is similar to the value fitted in the first experiment where binding of sigma factor has been neglected, see Table 1.

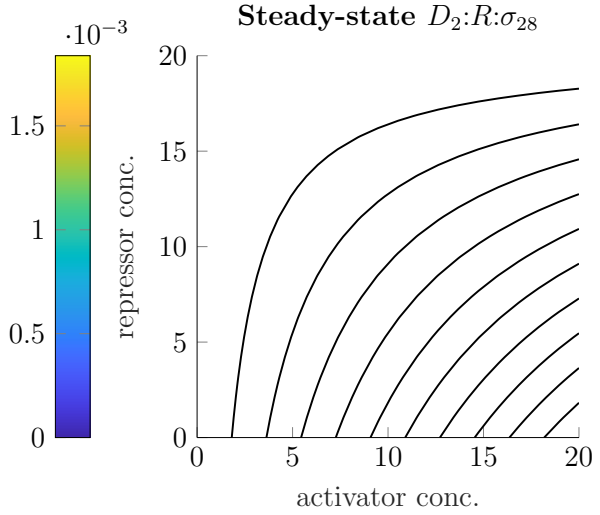

Figure 7: Promoter activity of *pTar-tetO*, obtained from fitted model.

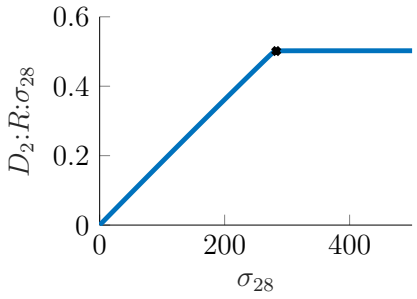

Figure 8: Amount of active transcription complex over relevant range of sigma factor concentration for *pTar* promoter.

The values given in Table 2 indicate that the total amount of RNAP is much bigger than the one of σ_{70} and further, that binding between these two species is very strong. Although σ_{28} also binds strongly to RNAP, its affinity is still smaller than the one of σ_{70}. The excessive amount of RNAP and the much higher binding affinity of σ_{70} thus leads to a decoupling of the two binding reactions.

We also note that the binding of $R{:}\sigma_{28}$ to the *pTar* promoter apparently has a very low affinity which leads to low GFP levels compared to the input step experiments. Together with the fact that the repressor *tetR* binds the *pTar* promoter very strongly, this leads to the steady-state promoter map depicted in Fig. 7, where the amount of active promoter for 20nM of DNA and varying activator and inhibitor concentrations is determined based on the reactions from Section 3.2 and parameters from Table 2.

Although there is leakage for medium amounts of inhibitor and activator and the level sets are not completely linear, the determined promoter dynamics are comparable to the desired behavior of Fig. 6 **A**.

Conclusion 3. *Using combinatorial promoters, the difference between two signals can only be realized to a limited extent.*

Given these results the transcription dynamics of Section 4.1 can now be extended with the appropriate promoter dynamics and σ_{28} as input. As pointed out before, the strong binding affinities of the sigma factors lead to a quite linear but bi-modal input-output behavior, as depicted in Fig. 8, compared to the one depicted in Fig. 5 **A**. Therein, the active $D_2{:}R{:}\sigma_{28}$ complex linearly follows the amount of σ_{28} until the concentration of RNAP is matched. Consequently, the transcriptional gain A_{tx} changes due to the change of input and using the same

linear approximation as defined in (16), one now finds

$$D_2{:}R{:}\sigma_{28} \approx \tilde{A}_{tx} \cdot \sigma_{28}$$
$$\text{with} \quad \tilde{A}_{tx} = 0.0018. \tag{19}$$

4.3 Implications for the closed loop

Initially, with Requirements 1 to 4, we expected the process of transcription to behave like a gain, translation to behave like an integrator and combinatorial promoters to provide the difference of two signals. Now, several observations were made which differ from our initial view.

First, although mRNA degradation is indeed much faster than protein degradation, the simplification to a simple gain is not justified and the temporal dynamics of mRNA production should be taken into account instead, leading to a PT1 behavior instead of a gain.

Second, both production and degradation rates are subject to saturations due to finite amounts of resources of the transcriptional and translational machinery in the cell-free extract. For small inputs however, these rates can be seen as linear functions of their inputs and the linear operation regimes have been determined explicitly.

Third, when realizing the difference of two signals by using combinatorial promoters, one only obtains an approximation of the difference and the quality of the estimate depends on the magnitudes of the inputs.

Now that these deviations from our initial requirements have been identified and characterized, their effect on functionality and performance of the synthetic genetic differentiator postulated in [4] can be studied. For that purpose, two different models are compared with the ideal realizable differentiator from Eq. (1) in both time and frequency domain.

The first model is given by the closed loop of the models G_{tx} and G_{tl} given in (17) and (18) respectively and adapted with the new transcriptional gain (19). This results in a linear model like depicted in Fig. 9 where no saturation effects are taken into account and perfect signal difference is assumed. However, the slow mRNA production and resulting PT1 behavior is taken into account and parameters of G_{tx} and G_{tl} resemble realistic values as they were obtained from experimental data. With the simplification $\delta = 0$, the transfer function of the

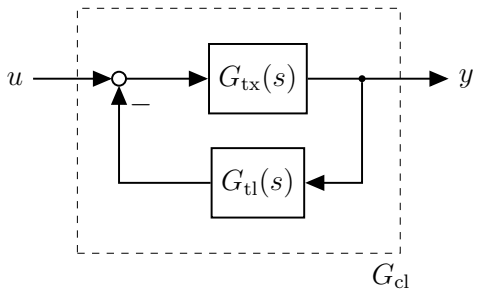

Figure 9: Topology of the linearized model given as the closed loop of G_{tx} and G_{tl}.

closed loop system thus is given by

$$G_{cl}(s) = \frac{\alpha \tilde{A}_{\mathrm{tx}} s}{s^2 + \gamma A_{\mathrm{deg}} s + \alpha\beta \tilde{A}_{\mathrm{tx}} A_{\mathrm{tl}}}. \tag{20}$$

The second model is considered as the detailed nonlinear model and is based on the reactions introduced in Section 3.2, thus taking all saturation effects, non-linearities and time-delays into account. It consists of a gene G_1 with a combinatorial promoter like the one studied in Section 4.2, i.e. sensitive to a σ_{28} holoenzyme and $tetR$ inhibitor, producing this very same inhibitor, therefore realizing the circuit from Fig. 1. The concentration of σ_{28} is considered as input signal. In order to capture both positive and negative gradients, the same approach as introduced in [4] is used, leading to a network topology like in Fig. 2 where G_2 is of similar structure as G_1 but with negative influence of the input on the transcription rate. The following additional mechanisms are necessary to realize this topology:

a) Additionally to σ_{28}, a second sigma factor σ_{xx} is introduced to be present at a constant level. While $R{:}\sigma_{28}$ activates transcription of G_1 and $R{:}\sigma_{xx}$ activates the one of G_2, both holoenzymes bind to both genes, leading to a competition and negative influence of one to the other.

b) Self inhibition of the two genes is achieved by two different inhibitors, e.g. $tetR$ and $tetR^\star$.

c) The two inhibitors $tetR$ and $tetR^\star$ undergo an annihilation reaction at rate $\mu = 0.1/(\mathrm{nM} \cdot \mathrm{min})$ which was chosen arbitrarily.

As these modifications have been discussed in [4] already, we omit the details at this point. The mRNA and protein dynamics of the core species as well as the output of the system is given

Table 3: Summary of the models and comparison of the core features.

	Desired circuit	Model 1	Model 2
Topology	Fig. 1	Fig. 9	Fig. 2
Dynamics	Eq. (1)	Eq. (20)	Eq. (21)
Features	linear	linear	nonlinear
	perfect gain	delayed gain	delayed gain
	no saturation	no saturation	saturation
	perfect difference	perfect difference	approximated difference

by

$$\dot{\overline{M}}_{tetR} = \alpha \cdot D_{tetR}{:}R{:}\sigma_{28} - \gamma \cdot M_{tetR}{:}E \tag{21a}$$

$$\dot{\overline{tetR}} = \beta \cdot M_{tetR}{:}Q - \delta \cdot \overline{tetR} - \mu \cdot tetR \cdot tetR^{\star} \tag{21b}$$

$$\dot{\overline{M}}_{tetR^{\star}} = \alpha \cdot D_{tetR^{\star}}{:}R{:}\sigma_{xx} - \gamma \cdot M_{tetR^{\star}}{:}E \tag{21c}$$

$$\dot{\overline{tetR^{\star}}} = \beta \cdot M_{tetR^{\star}}{:}Q - \delta \cdot \overline{tetR^{\star}} - \mu \cdot tetR \cdot tetR^{\star} \tag{21d}$$

$$y = \overline{M}_{tetR} - \overline{M}_{tetR^{\star}}. \tag{21e}$$

We summarized the core features of these two models and the desired circuit in Table 3. Note that Model 1 can be seen as the linearized version of Model 2.

4.3.1 Frequency domain analysis

In a first step, we compare the two models and desired behavior in the frequency domain, i.e. in terms of the Bode plot depicted in Fig. 10. This again is a classical tool from the control community and graphically shows how sinusoid input signals are modified by a certain transfer function. In the upper part, the magnitude amplification $\Lambda(\omega)$ indicates how the amplitude of the input signal is amplified for different input frequencies. In the lower part, the phase shift $\Omega(\omega)$ for these frequencies is shown. Magnitude and phase of the desired behavior (solid black) and linearized model (solid blue) are obtained trivially using Matlab.

For the nonlinear model however, we use a describing function approach as described in [33] to compare the input-output behavior of the nonlinear Model 2 with the linear ones. Therefore, the nonlinear Model 2 is excited with input

$$u(t) = A_0 + A \cdot \sin(\omega t) \tag{22}$$

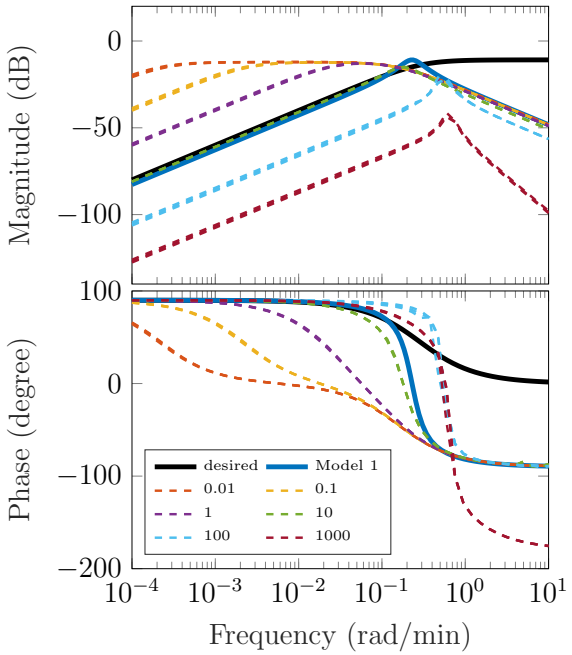

Figure 10: Bode plot of desired model (solid black) and linear Model 1 (solid cyan). In dashed lines magnitude and phase of output of nonlinear Model 2 subject to $u(t) = A_0 + A \cdot \sin(\omega t)$. Different colors indicate different values of A_0. Several values for A are plotted (lying on top of each other).

and the corresponding output $y(t)$ analyzed in terms of its Fourier coefficients. Assume that after time $t^\star = k^\star \frac{2\pi}{\omega}$, the output oscillates in a steady-state fashion, i.e. no transient dynamics occur anymore, and let

$$c_n(\omega) := \frac{1}{T} \int_{t^\star}^{t^\star + T} y(t) e^{-in\omega t} dt \quad \text{with period}$$
$$T := \frac{2\pi}{\omega}$$

be the n-th Fourier coefficient of signal $y(t)$ which corresponds to frequency ω. Then, the magnitude amplification Λ will be given as the ratio of the magnitudes of the first Fourier coefficients of output and input signal. With the input defined like in (22), the first Fourier coefficient of this signal is simply $\frac{A}{2i}$. Therefore we have

$$\Lambda(\omega) = \frac{|c_1(\omega)|}{|\frac{A}{2i}|} = \frac{2|c_1(\omega)|}{A}.$$

Further, the phase shift Ω for this frequency and particular input signal is given by

$$\Omega(\omega) = \operatorname{atan}\left(-\frac{\operatorname{Re} c_1(\omega)}{\operatorname{Im} c_1(\omega)}\right).$$

The constant part A_0 of input signal (22) is necessary to produce non-negative sinusoid functions. Due to the non-linearity of Model 2, the output $y(t)$ does not only depend on the frequency ω but the shape of the input function in general, i.e. also the variables A_0 and A. We therefore probed the system for several frequencies and values for A_0 and A.

For linear systems, an input signals with a single frequency component, like the one of (22), leads to an output with also only one frequency component, namely the same as the one of the input. In other words, higher harmonics are not existent and $|c_n(\omega)| = 0$ for $n > 1$. This is not the case for general non-linear systems where higher harmonics can also appear and in principle more than just the first Fourier coefficient should be analyzed. Thus, the way we use the describing function approach in this work relies on the assumption that higher harmonics of the output signal can be neglected. We thus analyzed the power spectrum of the output signals for different values of A, A_0 and ω and found that for most combinations, the higher harmonics contributed less than 5% to the overall power spectrum. However, in the case when A approaches A_0 and ω is close to the pole of the transfer function, it seems that the assumption does not hold, see Supplementary Data E for further details. We will see in Section 4.3.2 and Fig. 11 what this means for the output signal.

In Fig. 10, magnitudes and phases of the respective response signal are plotted as dashed lines where different colors indicate different values for A_0. The values for A are chosen as $A = kA_0$ with $k \in [0.1, 0.5, , 0.75]$ and respective responses plotted in the same color. As seen in Fig. 10, the output response does not change with varying A, however, the choice of the offset A_0 significantly influences the I/O behavior of the nonlinear signal differentiator. Very low values of A_0 (dashed red, orange and purple) lead to a very sensitive response, i.e. too high gain of the resulting closed loop and a smaller range of frequencies for which the output approximates the derivative of the input.

A value of $A_0 = 10$ (dashed green) results in the best response of the nonlinear system, matching the gain of an ideal differentiator quite well while providing almost the same frequency range as the one predicted by the linearized system ($\omega_{\max} \approx 0.03$ rad/min). For too large values

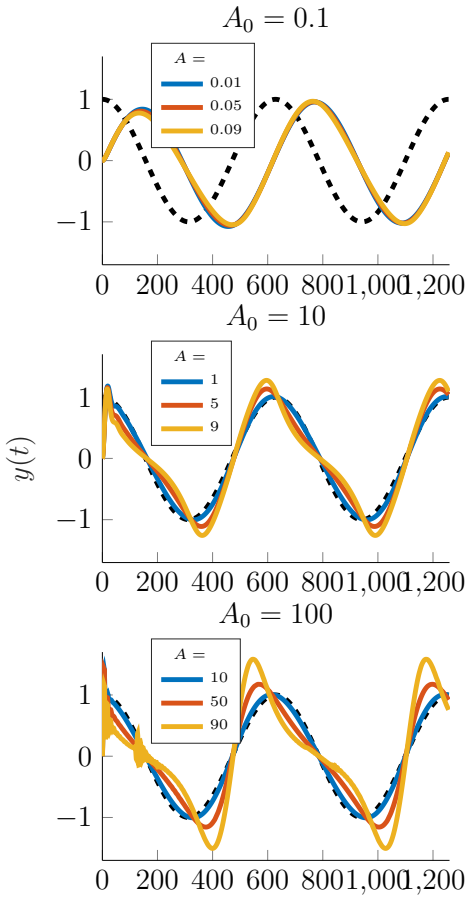

Figure 11: Normalized output $\tilde{y}(t)$ of the nonlinear closed model with input $u(t) = A_0 + A \cdot \sin(0.01t)$ for varying values of A_0 and A.

of A_0 (dashed cyan and dark red) Model 2 breaks down as expected due to the previously characterized saturation effects and the resulting loss of sensitivity towards the input signal.

4.3.2 Time domain analysis

From the previous analysis, we summarize that for the detailed Model 2, the phase of the output signal is off for too small values of A_0, the gain is very small for values of $A_0 \gg 10$ and only in case of $A_0 \approx 10$ both magnitude and phase are as desired. We now focus on the shape of the output signal of Model 2 and therefore stick to sinusoid input signals, fixing $\omega = 0.01$ but varying A_0 and amplitude A of the input signal. The normalized output

$$\tilde{y}(t) = \frac{1}{\omega A} y(t) \tag{23}$$

as response to the just described input is depicted in Fig. 11.

For a small value of A_0, as expected, the phase is off, however, the output signal still has a sinusoid shape for all amplitudes A. In contrast, for higher values of A_0, the phase is correct but with amplitude A approaching the offset A_0, the output signal becomes more and more distorted. This effect is amplified for higher offset values and is caused by a dilution of the power spectrum as discussed in the previous section.

5 Discussion

For the synthesis of genetic networks that realize arbitrary linear transfer functions, we follow a similar approach as in [3]. Therefore, it is crucial to find suitable genetic counterparts to primitive I/O functions such as gain, integration and difference. In a first attempt discussed in [4] and recapitulated in Section 2, several requirements were introduced to associate the processes of transcription and translation and combinatorial promoters with these respective I/O primitives. Now, a series of experiments and analyses was presented to verify and adapt these requirements.

By observing mRNA and protein levels as response to step inputs of varying height, it was verified in Conclusion 1 that protein degradation is almost non-existent while mRNA degradation is comparably fast. However, degradation dynamics are not as fast as desired and a quasi steady state assumption for the process of transcription would be oversimplifying. Thus, transcription should be considered as a PT1-element rather than a gain.

By fitting an ODE model to the experimental data and analyzing the corresponding parameters, it was also shown that all processes are subject to saturation due to limited amounts of resources. Using the same model and the fitted parameters, the linear operation regimes of the I/O primitives can be characterized as shown in Conclusion 2, leading to more insight into the capabilities and limitations of respective genetic circuits.

In a second series of experiments, the dependence of the performance of a combinatorial promoter on the operation regime was emphasized, realizing in Conclusion 3 that the difference of two signals can only be obtained approximately. Based on these insights, DNA concentrations for a simulation study were chosen such that the I/O behavior of the combinatorial promoter is as close as possible to the desired one. In conclusion, the use of combinatorial promoters for comparing the concentrations of two transcription factors is only possible within a limited

range of magnitudes and we suggest to use sequestration based mechanisms in future.

For the realization of a genetic signal differentiator using the studied parts, the initial goal was to realize a differentiator with high-pass filter. The corresponding transfer function is given in Eq. (1). It has a zero at the origin and one pole determined by the filter to make it a causal system. However, slow mRNA degradation leads to a behavior which, when linearized, is of relative degree one, i.e. Eq. (20) which has one zero at the origin and two poles in the left half plane. This reveals an additional delay of the transient dynamics.

If protein degradation were significantly larger than zero, this would lead to a transfer function of the form

$$G_{cl,protdeg} = \frac{K_1 s + K_1 \gamma_2}{s^2 + (\gamma_1 + \gamma_2)s + \gamma_1\gamma_2 + K_1 K_2}, \tag{24}$$

thus, shifting the zero from the origin to the right half plane and therefore leading to an additional lower frequency bound and a sign change in the output. In comparison, the differentiator introduced in [13] leads to a very similar transfer function as (24), given that all necessary assumptions introduced there hold. The main difference is that in [13], the zero of the transfer function always lies in the left half plane. On one hand, this means that a sign change is avoided. On the other hand, there inherently exists a lower bound for admissible input frequencies while for the design presented in this work, this only is be the case if protein degradation is large.

In order to conduct studies beyond the linearized model, a describing function approach is used to evaluate the response of the nonlinear model to sinusoid inputs like in Eq. (22). Therein, it can be seen that the performance of the differentiator critically depends on the constant part of the input signal, revealing again the limitations due to resource competition but also unexpectedly towards some supersensitivity at low values of A_0. With an appropriate choice of A_0, the presented network approximates the temporal derivative of an input signal for frequencies up to $\omega \approx 0.02$ rad/min. Additionally to the dependence on the absolute value of A_0, simulations in the time domain revealed a dependence on the relative amplitude $\frac{A}{A_0}$ in sense of a distortion of the output signal. When this relative amplitude approaches the value 1, the output signal looses its similarity to the sinusoid input, although phase and gain may be correct. In other words, the nonlinearities of the model lead to a dilution of the power spectrum of the output and higher harmonics are amplified.

Supplementary Data

Supplementary Data can be found after the References.

Acknowledgements

The authors are grateful for Vipul Singhal, Andrey Shur, Anandh Swaminathan and William Poole from the Murray Lab at Caltech for the thought provoking discussions and support in the laboratory.

References

[1] Purnick, P. E. M. and Weiss, R. "The second wave of synthetic biology: from modules to systems". In: *Nature Reviews Molecular Cell Biology* 10.6 (2009), pp. 410–422. DOI: 10.1038/nrm2698.

[2] Del Vecchio, D., Dy, A. J., and Qian, Y. "Control theory meets synthetic biology". In: *Journal of The Royal Society Interface* 13.120 (2016), p. 20160380. DOI: 10.1098/rsif.2016.0380.

[3] Oishi, K. and Klavins, E. "Biomolecular implementation of linear I/O systems". In: *IET Systems Biology* 5.4 (2011), pp. 252–260. DOI: 10.1049/iet-syb.2010.0056.

[4] Halter, W., Tuza, Z. A., and Allgöwer, F. "Signal differentiation with genetic networks". In: *Proceedings of the 20th IFAC World Congress*. 2017, pp. 10938–10943. DOI: 10.1016/j.ifacol.2017.08.2463.

[5] Ang, J., Bagh, S., Ingalls, B. P., and McMillen, D. R. "Considerations for using integral feedback control to construct a perfectly adapting synthetic gene network". In: *Journal of Theoretical Biology* 266.4 (2010), pp. 723–738. DOI: 10.1016/j.jtbi.2010.07.034.

[6] Yordanov, B., Kim, J., Petersen, R. L., Shudy, A., Kulkarni, V. V., and Phillips, A. "Computational design of nucleic acid feedback control circuits". In: *ACS Synthetic Biology* 3.8 (2014), pp. 600–616. DOI: 10.1021/sb400169s.

[7] Briat, C., Gupta, A., and Khammash, M. "Antithetic Integral Feedback Ensures Robust Perfect Adaptation in Noisy Bimolecular Networks". In: *Cell Systems* 2.1 (2016), pp. 15–26. DOI: 10.1016/j.cels.2016.01.004.

[8] Briat, C., Gupta, A., and Khammash, M. "Antithetic proportional-integral feedback for reduced variance and improved control performance of stochastic reaction networks". In: *Journal of the Royal Society Interface* 15.143 (2018). DOI: 10.1098/rsif.2018.0079.

[9] Briat, C. and Khammash, M. "Perfect Adaptation and Optimal Equilibrium Productivity in a Simple Microbial Biofuel Metabolic Pathway Using Dynamic Integral Control". In: *ACS Synthetic Biology* 7.2 (2018), pp. 419–431. DOI: 10.1021/acssynbio.7b00188.

[10] Qian, Y. and Del Vecchio, D. "Realizing 'integral control' in living cells: how to overcome leaky integration due to dilution?" In: *Journal of The Royal Society Interface* 15.139 (2018), p. 20170902. DOI: 10.1098/rsif.2017.0902.

[11] Harris, A. W. K., Dolan, J. A., Kelly, C. L., Anderson, J., and Papachristodoulou, A. "Designing Genetic Feedback Controllers". In: *IEEE Transactions on Biomedical Circuits and Systems* 9.4 (2015), pp. 475–484. DOI: 10.1109/TBCAS.2015.2458435.

[12] Lang, M. and Sontag, E. "Scale-invariant systems realize nonlinear differential operators". In: *Proceedings of the American Control Conference*. 1. IEEE, 2016, pp. 6676–6682. DOI: 10.1109/ACC.2016.7526722.

[13] Chevalier, M., Gomez-Schiavon, M., Ng, A., and El-Samad, H. "Design and analysis of a Proportional-Integral-Derivative controller with biological molecules". In: *bioRxiv* April (2018), p. 303545. DOI: 10.1101/303545.

[14] Halter, W., Montenbruck, J. M., Tuza, Z. A., and Allgöwer, F. "A resource dependent protein synthesis model for evaluating synthetic circuits". In: *Journal of Theoretical Biology* 420 (2017), pp. 267–278. DOI: 10.1016/j.jtbi.2017.03.004.

[15] Siegal-Gaskins, D., Tuza, Z. A., Kim, J., Noireaux, V., and Murray, R. M. "Gene Circuit Performance Characterization and Resource Usage in a Cell-Free "Breadboard"". In: *ACS Synthetic Biology* 3.6 (2014), pp. 416–425. DOI: 10.1021/sb400203p.

[16] Dolan, J., Anderson, J., and Papachristodoulou, A. "A loop shaping approach for designing biological circuits". In: *Proceedings of the 51st IEEE Conference on Decision and Control*. IEEE, 2012, pp. 3614–3619. DOI: 10.1109/CDC.2012.6426405.

[17] Gyorgy, A. and Del Vecchio, D. "Limitations and trade-offs in gene expression due to competition for shared cellular resources". In: *Proceedings of the 53th IEEE Conference on Decision and Control*. Vol. 2015-Febru. February. IEEE, 2014, pp. 5431–5436. DOI: 10.1109/CDC.2014.7040238.

[18] Del Vecchio, D. and Murray, R. M. *Biomolecular Feedback Systems*. Princeton, NJ: Princeton University Press, 2015.

[19] Takahashi, M. K. et al. "Characterizing and prototyping genetic networks with cell-free transcription–translation reactions". In: *Methods* 86 (2015), pp. 60–72. DOI: 10.1016/j.ymeth.2015.05.020.

[20] Noireaux, V., Bar-Ziv, R., and Libchaber, A. "Principles of cell-free genetic circuit assembly". In: *Proceedings of the National Academy of Sciences* 100.22 (2003), pp. 12672–12677. DOI: 10.1073/pnas.2135496100.

[21] Shin, J. and Noireaux, V. "Efficient cell-free expression with the endogenous E. Coli RNA polymerase and sigma factor 70". In: *Journal of Biological Engineering* 4.1 (2010), p. 8. DOI: 10.1186/1754-1611-4-8.

[22] Karig, D. K., Iyer, S., Simpson, M. L., and Doktycz, M. J. "Expression optimization and synthetic gene networks in cell-free systems". In: *Nucleic Acids Research* 40.8 (2012), pp. 3763–3774. DOI: 10.1093/nar/gkr1191.

[23] Sun, Z. Z., Hayes, C. A., Shin, J., Caschera, F., Murray, R. M., and Noireaux, V. "Protocols for Implementing an Escherichia coli Based TX-TL Cell-Free Expression System for Synthetic Biology". In: *Journal of Visualized Experiments* 79 (2013), pp. 1–10. DOI: 10.3791/50762.

[24] Chappell, J., Jensen, K., and Freemont, P. S. "Validation of an entirely in vitro approach for rapid prototyping of DNA regulatory elements for synthetic biology". In: *Nucleic Acids Research* 41.5 (2013), pp. 3471–3481. DOI: 10.1093/nar/gkt052.

[25] Grate, D. and Wilson, C. "Laser-Mediated , Site-Specific Inactivation of RNA Transcripts." In: *Proceedings of the National Academy of Sciences of the United States of America* 96.11 (1999), pp. 6131–6136.

[26] Sun, Z. Z., Yeung, E., Hayes, C. A., Noireaux, V., and Murray, R. M. "Linear DNA for Rapid Prototyping of Synthetic Biological Circuits in an Escherichia coli Based TX-TL Cell-Free System". In: *ACS Synthetic Biology* 3.6 (2014), pp. 387–397. DOI: 10.1021/sb400131a.

[27] Gruber, T. M. and Gross, C. A. "Multiple Sigma Subunits and the Partitioning of Bacterial Transcription Space". In: *Annual Review of Microbiology* 57.1 (2003), pp. 441–466. DOI: 10.1146/annurev.micro.57.030502.090913.

[28] Gyorgy, A. and Murray, R. M. "Quantifying resource competition and its effects in the TX-TL system". In: *Proceedings of the 55th IEEE Conference on Decision and Control*. Vol. 1. IEEE, 2016, pp. 3363–3368. DOI: 10.1109/CDC.2016.7798775.

[29] Courey, A. J. *Mechanisms in Transcriptional Regulation*. Malden, MA 02148-5020, USA: Blackwell Publishing, 2008.

[30] Shin, J. and Noireaux, V. "Study of messenger RNA inactivation and protein degradation in an Escherichia coli cell-free expression system". In: *Journal of Biological Engineering* 4 (2010), pp. 1–9. DOI: 10.1186/1754-1611-4-9.

[31] Klipp, E., Liebermeister, W., Wierling, C., Kowald, A., Lehrach, H., and Herwig, R. *Systems biology: A Textbook*. Wiley-VCH Verlag, 2009.

[32] Allgöwer, F. "Definition and Computation of a Nonlinearity Measure". In: *IFAC Nonlinear Control Systems Design* 28.14 (1995), pp. 257–262. DOI: 10.1016/S1474-6670(17)46840-6.

[33] Gelb, A. and Velde, W. E. V. *Multiple-input describing functions and nonlinear system design*. New York: McGraw-Hill, 1968.

[34] Rasmussen, C. and Williams, C. *Gaussian processes for machine learning*. The MIT Press, 2006.

Supplementary Data

A Data pre-processing

In this section we denote the data obtained in the experiments discussed in Sections 4.1 and 4.2 with $y \in \mathbb{R}$. There are mainly three issues with these data, exemplarily depicted in Fig. 12 **A** and **B** as grey crosses.

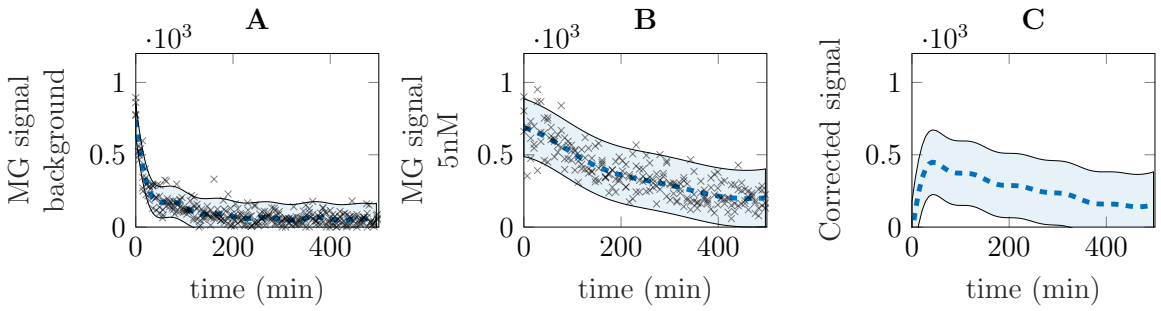

Figure 12: Data pre-processing of I/O experiments. Example: processing malachite green signal with 5nM of DNA.

First, the measurements are corrupted with noise, i.e.

$$y = f(t) + \rho, \ \rho \sim \mathcal{N}(0, \epsilon)$$

where $f(t)$ is some deterministic process generating the noise-free data and ρ the gaussian noise. This is particularly the case for the malachite green fluorescence measurements. Second, the time points at which the measurements are obtained are not uniformly spaced due to inconsistent preparation times of the experiments. This leads to a heterogeneous distribution of the measurements along the time axis. And last, for malachite green, a substantial part of the measured signal stems from some background signal caused by unbound malachite green, leading to the need of correcting the signals by subtracting the background part. However, due to the non-uniform temporal spacing of the measurements, a correction of the background requires some kind of model or interpolation scheme of the data.

We therefore assume that the measurement noise ρ is i.i.d. and model the timeseries for each

experimental condition as a gaussian process, i.e.

$$y \sim \mathcal{GP}\left(\mu, k(t, t', \theta) + \epsilon^2 \delta_{tt'}\right)$$

where $\mu \in \mathbb{R}$ is a constant mean, k is chosen as a squared exponential kernel parametrized with θ and $\delta_{tt'}$ being the Kronecker delta.

Now let $y^{(ctrl)}$ and $y^{(e)}$ be the fitted gaussian processes of a control experiment without any DNA and some other experimental condition with predicted mean $\mu_\star^{(ctrl)}$, $\mu_\star^{(e)}$ and predicted standard deviations $\sigma_\star^{(ctrl)}$, $\sigma_\star^{(e)}$ as derived in [34] and depicted in Fig. 12 as dashed blue lines (mean) and light blue shaded area (standard deviation). The background corrected signal $\tilde{y}^{(e)}$ is then determined by

$$\begin{aligned}
\tilde{\mu}_\star^{(e)} &= \mu_\star^{(e)} - \mu_\star^{(ctrl)} \\
\left(\tilde{\sigma}_\star^{(e)}\right)^2 &= \left(\sigma_\star^{(e)}\right)^2 + \left(\sigma_\star^{(ctrl)}\right)^2,
\end{aligned}$$

like depicted in Fig. 12 **C**.

Finally, the fluorescence signals are converted from the arbitrary intensity unit into a concentration unit, using the previously obtained calibration relations

$$\begin{aligned}
1723\, a.u. &= 1\mu M \text{ GFP} \\
775\, a.u. &= 1\mu M \text{ mRNA}.
\end{aligned}$$

B Genetic constructs

Gene	functional contents	sequence information
D_1	pBest-deGFP-MGapt	addgene.org/67734/
D_2	pTar-tetO-deGFP	see * for sequence
D_{s28}	pBest-$\sigma_2$8	addgene.org/45779/
D_{tetR}	pBest-tetR	addgene.org/45778/

*GGCATGCCAAGCTTCAATAAAGTTTCCCCCCTCCTTGCCGATAATCCCTATC
AGTGATAGAGAGCTAGCAATAATTTTGTTTAACTTTAAGAAGGAGATATACCA
TGGAGCTTTTCACTGGCGTTGTTCCCATCCTGGTCGAGCTGGACGGCGACGTA
AACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGG
CAAGCTGACCCTGAAGTTCATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGC
CCACCCTCGTGACCACCCTGACCTACGGCGTGCAGTGCTTCAGCCGCTACCCC
GACCACATGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGT
CCAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCG
AGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATC
GACTTCAAGGAGGACGGCAACATCCTGGGGCACAAGCTGGAGTACAACTACAA
CAGCCACAACGTCTATATCATGGCCGACAAGCAGAAGAACGGCATCAAGGTGA
ACTTCAAGATCCGCCACAACATCGAGGACGGCAGCGTGCAGCTCGCCGACCAC
TACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCTGCTGCCCGACAACCA
CTACCTGAGCACCCAGTCCGCCCTGAGCAAAGACCCCAACGAGAAGCGCGATC
ACATGGTCCTGCTGGAGTTCGTGACCGCCGCCGGGATCAGAAGGGAAGAAAGA
GCAAAGAAGGTAGCATAA

C Models

This section extends the results presented in Section 3.2.

Proof of Proposition 1. By setting $\frac{d}{dt}A = 0$, one arrives at the quadratic equation

$$(A{:}B)^2 - A{:}B(K + \overline{A} + \overline{B}) + \overline{A}\overline{B} = 0$$

which in general can have either none, exactly one or two real solutions, determined by the discriminant

$$\Delta = (K + \overline{A} + \overline{B})^2 - 4\overline{A}\overline{B}.$$

Taking into account that only $K > 0$, $\overline{A} > 0$ and $\overline{B} > 0$ are biologically meaningful, one finds

$$\Delta = K^2 + K\overline{A} + K\overline{B} + (\overline{A} - \overline{B})^2 \geq 0$$

thus at least one real solution exists. For the existence of exactly one solution, one would need $K = 0$ which was excluded previously. Otherwise, the quadratic formula yields

$$A{:}B_{1,2} = \\ \frac{1}{2}\left(K + \overline{A} + \overline{B} \pm \sqrt{(K + \overline{A} + \overline{B})^2 - 4\overline{A}\overline{B}}\right)$$

where we assign $A{:}B_1$ to the solution with the negative and $A{:}B_2$ to the one with the positive sign. Due to mass conservation, we are interested in the solution for which

$$0 \leq A{:}B_i \leq \min(\{\overline{A}, \overline{B}\}) \tag{25}$$

holds. Now, $0 \leq A{:}B_i$ for both $i = [1, 2]$ follows directly from

$$\left(K + \overline{A} + \overline{B}\right)^2 \geq \left(K + \overline{A} + \overline{B}\right)^2 - 4\overline{A}\overline{B}.$$

and with

$$\begin{aligned}\frac{1}{2}\left(K+\overline{A}+\overline{B}\right) &\geq \frac{1}{2}\left(K+2\min(\{\overline{A},\overline{B}\})\right)\\ &> \min(\{\overline{A},\overline{B}\})\end{aligned}$$

it can be seen that $A{:}B_2$ violates (25). It remains to realize that

$$\begin{aligned}K+\overline{A}+\overline{B}-2\min(\{\overline{A},\overline{B}\}) \leq\\ \sqrt{(K+\overline{A}+\overline{B})^2-4\overline{A}\overline{B}}\end{aligned}$$

to conclude that $A{:}B_1$ is the only biologically meaningful solution. □

D pTar promoter characterization

The time series data of the pTar characterization experiment is depicted in Fig. 13.

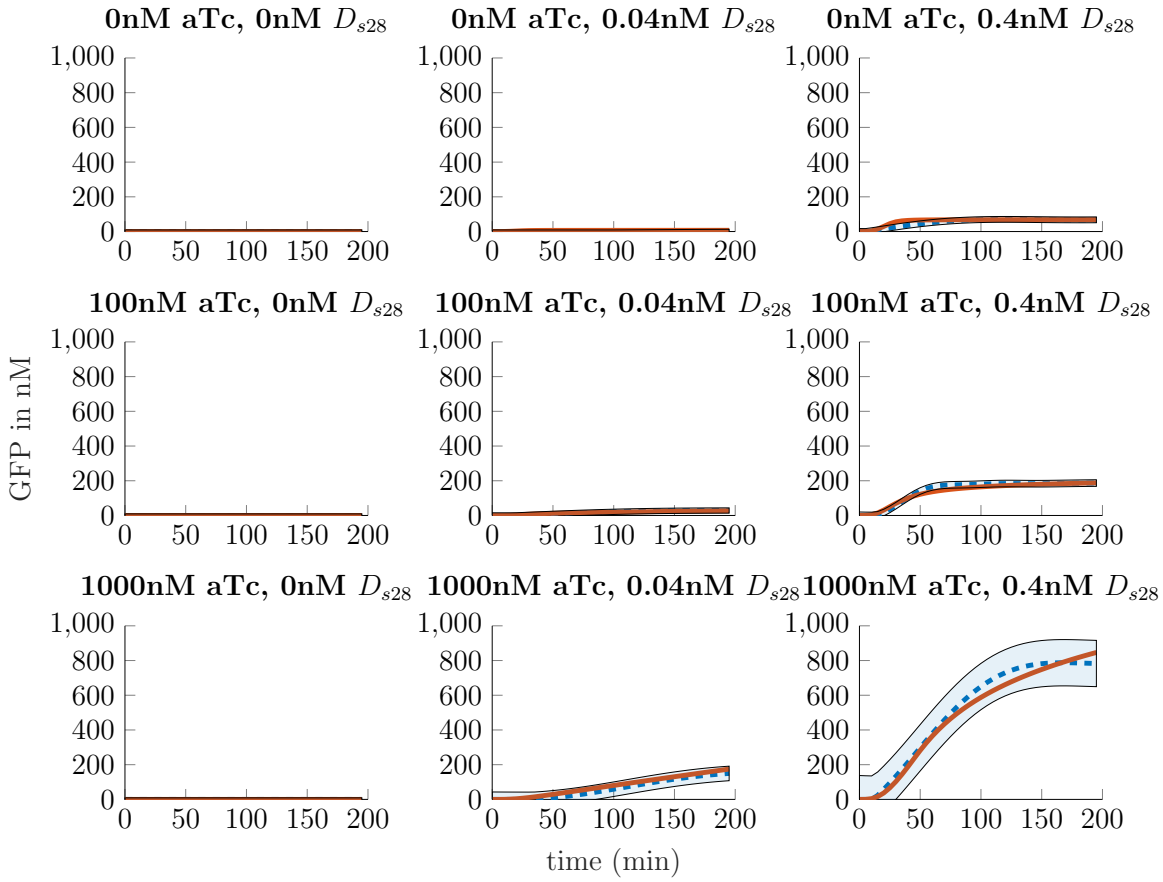

Figure 13: Concentration of GFP over time. pTar concentration at 5 nM, varying sigma factor DNA $\overline{D}_3$(increasing from left to right) and inhibitor concentrations (increasing from top to bottom).

E Limitations of the Describing Function approach

The way the Describing Function approach has been used in Section 4.3.1, we assume that higher harmonics can be neglected in the output signal. This, however, may not the be case for every combination of parameters A, A_0 and ω of the input signal given in (22). Therefore, we analyzed the output signal of the nonlinear Model 2 in terms of its 10 first Fourier coefficients

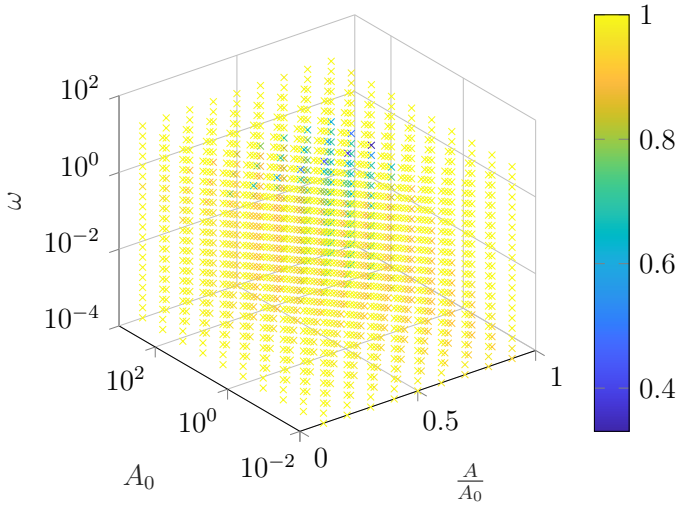

Figure 14: Proportion of basis frequency in the power spectrum of the output signal generated by the nonlinear system subject to input (22) over different parameters of the input signal.

and calculated the proportion of the basis frequency in the power spectrum, i.e.

$$p_{\text{rel}} = \frac{|c_1(\omega)|^2}{\sum_{n=1}^{10} |c_n(\omega)|^2}. \tag{26}$$

If $p_{\text{rel}} \approx 1$, this indicates that higher harmonics can be neglected. As shown in Fig. 14, this is not always the case. For large values of $\frac{A}{A_0}$ and input frequencies in the range $\omega \in [10^{-2}, 10^0]$, the value of p_{rel} drops below 0.8, suggesting that the output signal will significantly be influenced by frequency components other than the basis frequency ω. This means that the output signal will have a distorted shape.