Completeness of transcriptional repressor networks operating in the unsaturated regime

Ophelia S. Venturelli[1] and David Soloveichik[2,*]
[1]Department of Bioengineering, University of California Berkeley, Berkeley, CA, USA
[2]Center for Synthetic & Systems Biology, University of California San Francisco, San Francisco, CA, USA
[*]To whom correspondence should be addressed: David.Soloveichik@ucsf.edu

Abstract

While the range of boolean behaviors realized by transcriptional networks is well understood, the range of dynamic behaviors that can be realized by transcriptional networks operating in the graded regime is not yet known. We show that networks of appropriately connected and tuned repressor transcription factors can approximate the large class of mass-action dynamical systems. This completeness result generalizes the logical completeness of repressors in the saturated regime for arbitrary boolean functions. Our argument relies on showing how combinatorial control over the rate of transcription can mimic combinatorial control over degradation that is outside of the transcriptional scope. Our result yields a systematic method to construct novel analog dynamics using synthetic transcriptional networks, and contributes to the understanding of the range of system behaviors generated by transcriptional regulation.

Keywords: transcriptional networks; analog computation; dynamical systems; mass-action

Introduction

Synthetic biology aims to apply engineering principles to design and control the behavior of living cells. The target applications are inaccessible to traditional engineering approaches and include re-programming cells to target pathogens, detoxify the environment, and produce biofuels. All these applications require synthetic regulatory networks to manage how cells navigate their environment and regulate their internal state. The principles of analog computation can provide crucial insights into building complex information processing functions in living cells. Synthetic circuits that use the analog computation framework can require significantly fewer components and use less energy than digital alternatives, and are particularly suited for applications that do not rely on high precision computation [1, 2]. Indeed, operating under the same constraints, there are many examples of analog behavior in natural biological regulatory networks [3, 4].

An important conceptual step in engineering computational systems has been the development of the notion of *completeness*. Informally, a component type is complete for a class of systems if any instance of that class can be constructed by combining and recombining this component type in various ways. In the boolean context, completeness provides a way to systematically assemble a set of simple components into circuits to compute complex logic functions. As a notable example, Claude Shannon showed how boolean algebra can systematize combining switches and relays — a connection that revolutionized our ability to create electronic circuits at the time [5]. The completeness of minimal functional elements like NOR gates was crucial to the development of computers (e.g. the computational core of the Apollo navigation computer constituted entirely of NOR gates) [6]. In synthetic biology, transcription factors that repress specific sets of genes (repressors) can similarly constitute a NOR gate for biological computation [7]. Together, the transcriptional implementation of the NOR operation and the ability to compose such systems enables the construction of arbitrary logic functions. As in the boolean case, history is not without precedent for theories of analog completeness. For example, Shannon again provides inspiration: his work on the differential analyzer put this analog computer on solid mathematical grounds [8]. More

recent work explored a systematic way to implement dynamical systems in analog electronics [9].

Motivated by the broad applications of systematic design using complete basis elements in the Boolean context, we investigate a complete basis for biological analog dynamics. We develop a method showing how a simple form of regulation – repression of a promoter by a transcription factor – can be complete for a well-studied and computationally powerful class of mass-action dynamical systems.

Transcriptional networks can naturally exhibit both digital and analog regimes. Depending on parameters such as the Hill-coefficient and affinity, the transcription factor binding site in the promoter can be saturated (always unoccupied or occupied) or unsaturated. Whereas the transition between the two saturated regimes represents a boolean state (ON or OFF), the unsaturated regime exhibits a continuous dependence of the transcriptional activity on the concentration of the transcription factor. Our focus on transcriptional regulation is also motivated by its programmable and modular characteristics. Diverse libraries of mined transcription factors and new technologies such as CRISPRi have recently emerged that will significantly advance the rational design of orthogonal transcriptional circuits [10, 11, 11, 12]. We rely on repressors since activators are functionally redundant: two repressors composed serially can mimic activation. Further, repressors are mechanistically simple: these proteins function by sterically occluding RNA polymerase thus facilitating rigorous mathematical modeling [13, 14].

We show a novel route to a transcriptional implementation of mass-action behaviors, which encompass a diverse repertoire of analog behaviors including temporal patterning, multi-stability and memory, logic, signal processing, control systems and distributed algorithms [15–19]. The completeness of unsaturated repressor networks for mass-action kinetics also implies that repressor networks can approximate the kinetic behavior of any other chemical system albeit at a different time scale — and in that sense they are "maximally powerful". We target mass-action kinetics rather than the more general class of polynomial ODEs because of mathematical convenience, and not because of the ostensible concurrence between the chemical nature of transcription and the central role of mass-action in chemical kinetics. In particular, it turns out that approximating complex mass-action degradation terms is easier than arbitrary negative monomials. An important limitation of our comprehensive approach is that implementing some systems would require precise tuning of certain parameters, which may be difficult to achieve in the messy environment of the cell. The precise enumeration of systems "robust" to imperfect tuning of the parameters in our model remains open.

Related work

Biological networks for implementing desired analog behaviors have been designed using two general types of approaches: enumeration strategies and dynamical systems theory. Systematic enumeration and *in silico* evolution approaches have been used to search exhaustively for circuit topologies that implement specific behaviors such as adaptive dynamics and programmable spatial localization [20–24]. However, in systems with greater than three or four nodes, the computational power required to search exhaustively for the function of interest becomes prohibitive. Alternatively, rational approaches based on dynamical systems theory have been used to construct specific classes of behaviors such as bistability and limit-cycle oscillations. For example, nullcline analysis is a well-established method for constructing a bistable switch from first principles [25]. However, what constitutes the set of "functionally complete" behaviors has not yet been formulated for transcriptional regulation. A unified framework for synthesizing analog behaviors could significantly expand the range of achievable functions and provide important insights into the limitations of synthetic biology.

A number of papers have developed frameworks for doing analog computation with transcriptional machinery. Perkins and Cory proposed a systematic framework for building arithmetic functions with transcriptional networks [26]. However, they did not apply their approach to dynamical systems. Further, their method requires more complex basis elements including protein-level interactions between transcription factors. Recent experimental work showed synthetic analog computation in living cells [2]. While their use of the log-linear regime simplifies some computations (eg multiplication) and increases

robustness to noise, it is not clear how to combine such computations into larger dynamical systems with feedback.

Models

Cells use complex regulatory networks that consist of many levels of regulation, using distinct physical mechanisms to perform computation. Considering different physical mechanisms in isolation could clarify the contribution of the different layers of regulation to the system as a whole. Here we focus on a single layer of regulation by transcriptional repression. Even at the transcriptional level, Nature provides many knobs to tune, including the transcription and translation rates, cooperativity, binding affinity of the transcription factor to its binding site and degradation rate of the transcription factor. Among these parameters, transcription rates can be most easily modified by the choice of well-characterized promoters [27]. Indeed, our completeness result relies on setting the transcription rates of the genes for the various repressors in the network, without needing to vary the remaining parameters. Such restrictions provide important insights into which parameters are essential for generating a large diversity of behaviors and constitutes an important step towards making the model "implementable."

For simplicity, in our model we lump transcriptional and translational dynamics into a single production law. This assumption is justified in the regime in which transcription is the rate limiting step (e.g. due to high levels of transcriptional repression or low basal levels of transcription). Furthermore, there are novel and versatile technologies for implementing RNA-based repressors that do not require translation for regulation [10, 28]. Finally, we focus on a deterministic model of transcriptional networks as opposed to a stochastic framework, which is valid in high concentration regimes. However, a stochastic implementation of a network demonstrates that stochastic effects do not significantly alter the designed functionality of the transcriptional networks (see Discussion).

In our model, we assume that the probability that a recognition site for repressor X is unoccupied is a Hill function

$$\frac{1}{1+(X/k_X)^{nx}} \tag{1}$$

with parameters: order n_X and binding constant k_X. *A promoter is active only if none of the repressor sites are occupied.* When a promoter p is active, expression occurs with some rate α_p. (Recall that we do not model separate RNA and protein level dynamics.) Each repressor undergoes a linear decay with rate β_X due to a combination of dilution (due to cell division) and degradation. The rate β_X is independent of the repressor concentrations in our network because we do not allow direct physical interactions between repressors. The same repressor gene could be present in multiple copies, under the control of different promoters. Adding up the contributions of multiple promoters for the same repressor, and subtracting a linear degradation rate, yields the following dynamics of each repressor X_i:

$$\dot{X}_i = \sum_{p\in\mathcal{P}(X_i)} \left(\alpha_p \prod_{j\in\mathcal{R}(p)} \frac{1}{1+(X_j/k_{X_j})^{n_{X_j}}} \right) - \beta_i X_i. \tag{2}$$

where $\mathcal{P}(X_i)$ is the set of promoters for repressor X_i.

To convert to dimensionless form, we use $x_i = X_i/k_{X_i}$, $a_p = \alpha_p\tau/k_{X_i}$ (for $p \in \mathcal{P}(X_i)$), $b_i = \beta_i\tau$, where τ sets the time scale. The dimensionless form then becomes as follows, which we term the *Hill-function model* of unsaturated repressor networks:

$$\dot{x_i} = \sum_{p\in\mathcal{P}(X_i)} \left(a_p \prod_{j\in\mathcal{R}(p)} \frac{1}{1+x_j^{n_{X_j}}} \right) - b_i x_i. \tag{3}$$

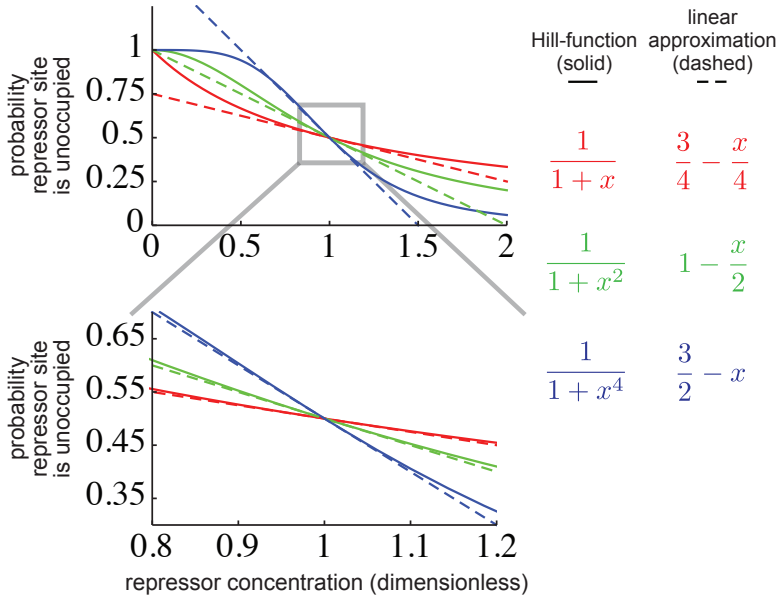


Figure 1. Unsaturated regime of Hill-functions. When the concentration of a transcription factor X is close to its half-max value (i.e. $X \approx k$, the binding constant), we can approximate the probability that its binding site remains unoccupied by $\frac{1}{2} - \frac{n(X-k)}{4k}$, where n is the Hill-coefficient. We plot the dimensionless quantity $x = X/k$ on the abscissa.

Networks of repressors operate in the unsaturated regime when X_i is close to k_{X_i}. In dimensionless form the equivalent condition is that all x_i remain close to 1. When x_j is close to 1, then we can approximate the Hill-function 1 by the linear function

$$\frac{1}{2} + \frac{n_{X_j}}{4}(1 - x_j). \tag{4}$$

(See Fig. 1). This yields the following approximation to the Hill-function model (eq 3), which we call the *component-wise linear model*:

$$\dot{x_i} = \sum_{p \in \mathcal{P}(X_i)} \left(a_p \prod_{r \in \mathcal{R}(p)} \left(\frac{1}{2} + \frac{n_{X_j}}{4}(1 - x_j) \right) \right) - b_i x_i. \tag{5}$$

We will use the simpler component-wise linear model to facilitate analysis and enable systematic design. Note that the component-wise linear model of a promoter with multiple repressors is overall non-linear since the linear terms due to each repressor are multiplied. A larger linear unsaturated regime enables a faithful generation of desired dynamics over a greater concentration range. In the Discussion, we describe a few strategies for expanding the linear range in these genetic circuits (Fig. 5).

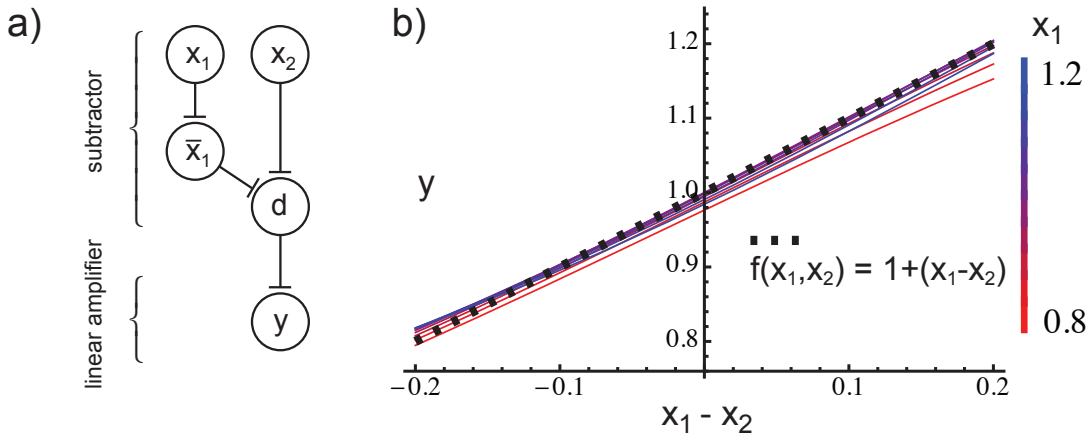
Deterministic ODE simulations were performed in Mathematica (Wolfram Research) and MATLAB (Mathworks) and stochastic simulations were performed using StochKit2 [29].

Results

Motivating Examples

To highlight the diversity of behavior of repressor networks operating in unsaturated regime, we start with three examples. For each example, the component-wise linear model (eq. 5) provides crucial insights into their behavior.

In control theory, a key task is to compute the difference between the output and the desired reference (error signal). We show how an unsaturated repressor network can perform this function by computing the difference of inputs x_1 and x_2 (Fig. 2). A challenge in our implementation is to use strictly positive concentrations to represent both positive and negative values. Our solution is to shift the signals to operate around 1 instead of 0. As a result, we compute the function $f(x_1, x_2) = 1 + x_1 - x_2$, where the (positive or negative) difference of x_1 and x_2 is represented as the (positive or negative) deviation from 1. The component-wise linearization enables a simple interpretation of the subtraction function computation (Fig. 2d,e). Numerical simulation of the full unsaturated repressor network confirms subtraction computation (Fig. 2b).



c)

$$\dot{\bar{x}}_1 = 2\left(\frac{1}{1+x_1^2}\right) - \bar{x}_1$$
$$\dot{d} = \left(\frac{1}{1+x_2^2}\right) + \left(\frac{1}{1+\bar{x}_1^2}\right) - d$$
$$\dot{y} = 2\left(\frac{1}{1+d^4}\right) - y$$

d) linear operating regime

$$\dot{\bar{x}}_1 = (2 - x_1) - \bar{x}_1$$
$$\dot{d} = \frac{1}{2}(2 - x_2) + \frac{1}{2}(2 - \bar{x}_1) - d$$
$$\dot{y} = (3 - 2d) - y$$

e) feed-forward equilibrium

$$y = 1 + (x_1 - x_2)$$

Figure 2. Subtraction example: Computation of the function $f(x_1, x_2) = 1 + x_1 - x_2$. **(a)** Repressor network topology and **(c)** the dimensionless Hill-function model (in the form of eq. 3). **(b)** Plot of the equilibrium output y as a function of the difference $x_1 - x_2$ in the Hill-function model of panel (c). Multiple colored lines correspond to difference values of x_1. The dotted line shows the expected ideal behavior. **(d)** Component-wise linear model (in the form of eq. 5).**(e)** Equilibrium solution of panel (d) showing that the correct function is computed.

A key feature of biological sensory networks is adaptation – a persistent change in the input signal results in a transient output response which then relaxes to its original level. Previous results have

shown that an incoherent feedforward loop can perform adaptation [30]. Similarly, unsaturated repressors composed in a feedforward loop can compute adaptation (Fig. 3). Although adaptation encompasses a broad range of transient responses, the output of our network can be shown to approximate specifically the differentiation function $f(x) = dx/dt$. Differentiation is also crucial for feedback control systems: many control algorithms compute the derivative to determine the direction and magnitude of desired actuation [31].

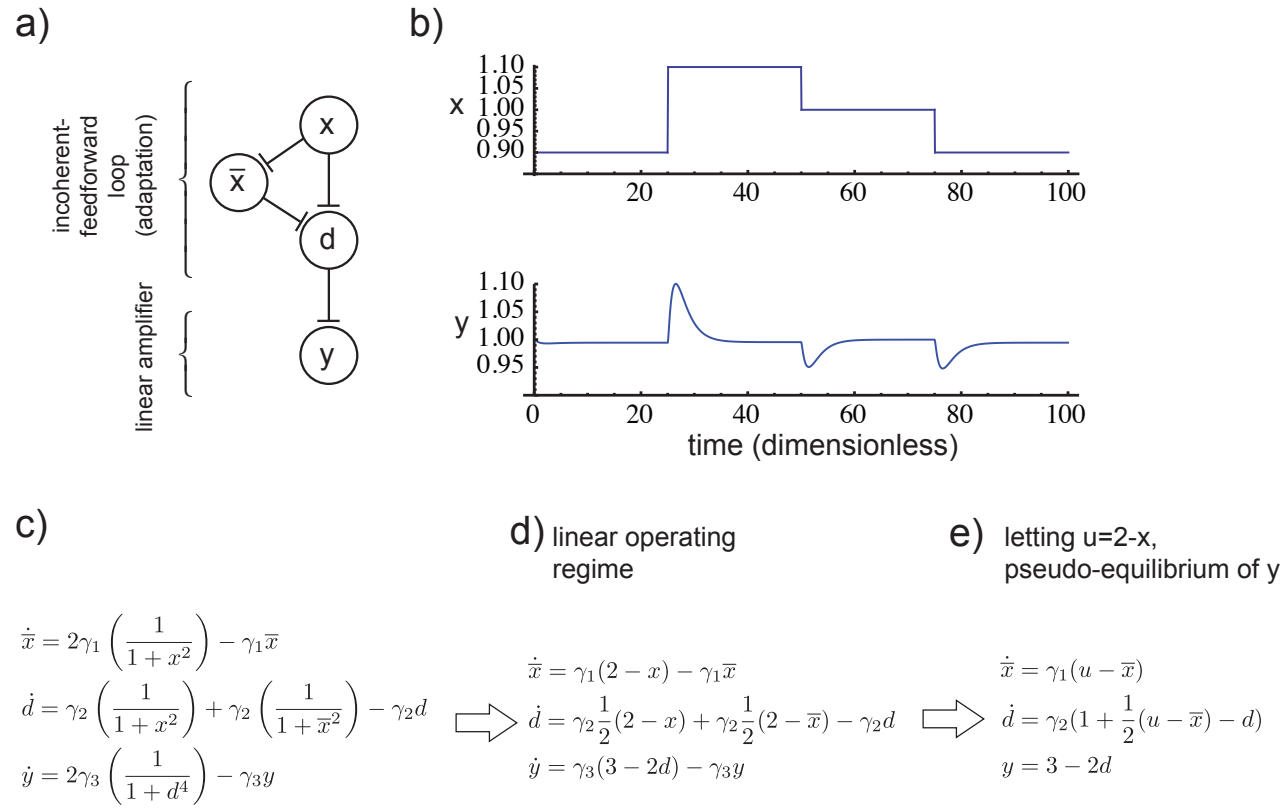


Figure 3. Adaptation example: $f(x) = 1 + dx/dt$ **(a)** Repressor network topology and **(c)** the dimensionless Hill-function model (in the form of eq. 3). **(b)** Numerical simulation of the model in panel (c) showing the response of y (bottom) to a step function input on x (top). **(d)** Component-wise linear model (in the form of eq. 5). **(e)** If we let $u = 2 - x$, then the ODEs for $\bar{x}$ and d can be understood in terms of the linear approximation differentiator (eqs. 6–7). The linear amplifier y multiplies and inverts the d which results in a (shifted) derivative.

Again, the component-wise linear model illuminates how our network achieves adaptation: at a constant input value of x, the equilibrium of $\bar{x}$ is $2 - x$, which results in equilibrium value $d = 1$, and thus $y = 1$, independent of x. Further analysis of the linearization shows how the output y computes the derivative of the input signal x. Setting $u = 2 - x$, the ODE for $\bar{x}$ is $\dot{\bar{x}} = \gamma_1(u - \bar{x})$. Note that the system

$$\dot{\bar{x}} = \gamma_1(u - \bar{x}) \quad (6)$$

$$w = \gamma_1(u - \bar{x}) \quad (7)$$

is the classic linear approximation differentiator with input u and output w, and transfer function $\frac{s}{s/\gamma_1+1}$, where $w \approx du/dt$ in the limit of large γ_1. (Intuitively, by eq. 6, $\bar{x}$ follows u with the fidelity determined

by γ_1. At the same time, since the right hand side of eq. 6 is the derivative of $\overline{x}$ and $\overline{x} \approx u$, we can infer that $w = \gamma_1(u - \overline{x}) \approx \frac{du}{dt}$.) Finally, returning to our original system, the ODE $\dot{d} = \gamma_2(1 + \frac{1}{2}(u - \overline{x}) - d)$ implies that d follows $1 + \frac{1}{2}(u - \overline{x}) \approx 1 + \frac{1}{2\gamma_1}\frac{du}{dt}$, which is a scaled and shifted version of the derivative of u. As before, the upward shift is necessary since repressor concentrations cannot be negative. Finally, y inverts d restoring the correct sign of the derivative with respect to x rather than u, and amplifies the signal. The accuracy of the approximation will depend on the parameters γ_1, γ_2, γ_3, and the time scale over which the input changes. Matching the adaptation response, numerical simulation of the full unsaturated repressor network confirms that the output signal responds only to changes in the input signal, and relaxes to its original value on constant input (Fig. 3b).

Due to their complex non-equilibrium behavior and their utility as biological clocks, oscillators have played a key role in the history of synthetic biology [32]. While a number of nonlinear oscillatory systems have been constructed, the class of oscillators that exhibit neutral stability has received less attention. Such oscillators do not approach a limit cycle, but rather maintain the initial oscillation amplitude over time and therefore the system's initial conditions can modulate amplitude and frequency. This unique property of neutral cycle oscillators can be used as a form of *analog memory*, in contrast to multi-stable systems which preserve one of a few states (digital memory).

One of the most well-studied neutral cycle oscillators is the Lotka-Volterra system (predator-prey) which has a rich history in mathematical ecology. Recently, a synthetic microbial ecosystem was engineered to exhibit this behavior at the community level [33]. Here we show that unsaturated repressor networks can realize the Lotka-Volterra oscillator by a genetic circuit within a cell (Fig. 4). Again, component-wise linearization reveals that our network approximates the desired behavior. In the linearization, the pseudo-equilibrium values of $\overline{x_1}$ and $\overline{x_2}$ are $2 - x_1$ and $2 - x_2$ for large γ (Fig. 4d). Plugging these into the ODEs for x_1 and x_2 and simplifying yields the Lotka-Volterra system (Fig. 4e). Simulations of the repressor network confirm that we approximate the dynamics of the Lotka-Volterra system and that the initial conditions govern the oscillation amplitude (Fig. 4b).

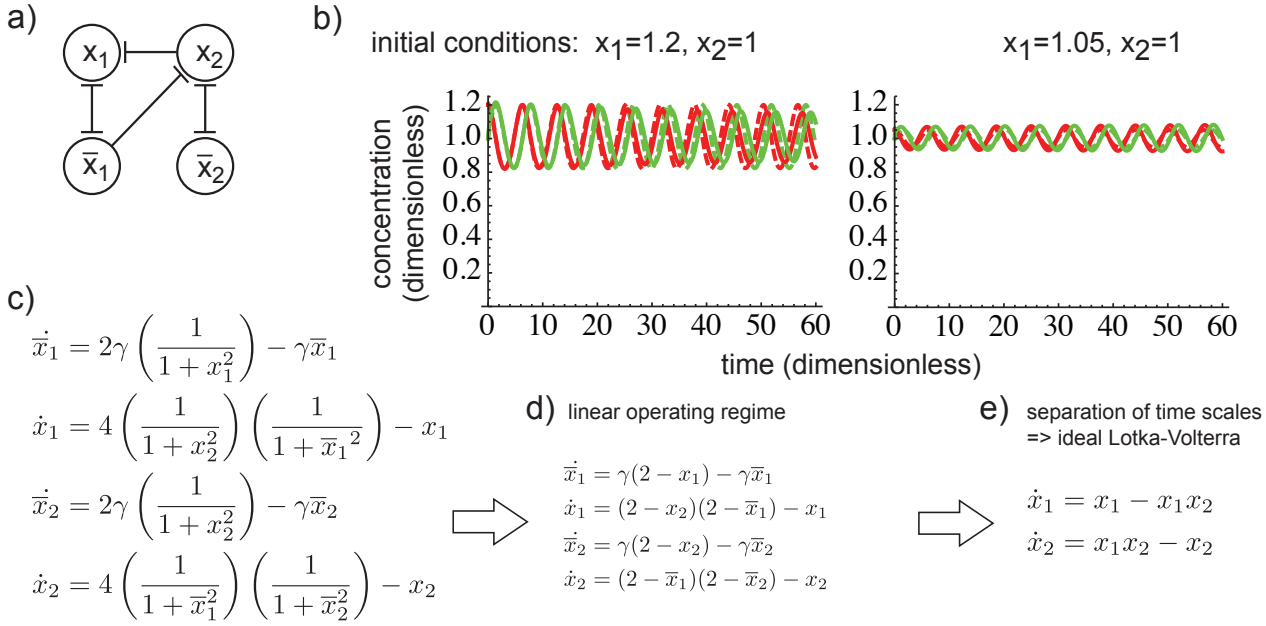
Repressor networks are complete for mass-action kinetics

The above examples motivate the question: How diverse is the set of behaviors of unsaturated repressor networks? Is there a systematic way to construct unsaturated repressor networks to implement desired functionalities?

In this section we provide a systematic method for constructing an unsaturated repressor network given any mass-action system such that the unsaturated repressor network mimics the behavior of the mass-action system (as long as the evolution of the mass-action systems stays close enough to $x_i \approx 1$). Mass-action systems encompass a very broad class of behaviors including all of classical chemical kinetics [16]. Given their ability to implement a wide variety of behaviors, and because they have already been extensively studied, mass-action systems constitute a natural design target. Moreover, we will see that the particular form of mass-action systems makes this class directly amenable to implementation with unsaturated repressor networks, compared with, say, systems of arbitrary (autonomous) polynomial differential equations. While there are systematic ways that mass-action systems can approximate arbitrary polynomial differential equations [34, 35], the resulting approximation occurs at a higher level.

We note that among the examples shown in the previous section, the subtraction example (Fig. 2) and adaptation example (Fig. 3) are not directly approximating mass-action systems (i.e. systems shown in parts (e) are not mass-action). It is still an open question to delineate the class of non-mass-action systems implementable with unsaturated repressor networks in our manner.

We first clarify what we mean by mass-action dynamical systems, and then describe an example illustrating the basic challenges and the corresponding solutions that arise when attempting to approximate such systems with repressor networks. Finally, we describe a general solution.



$$\dot{\bar{x}}_1 = 2\gamma\left(\frac{1}{1+x_1^2}\right) - \gamma\bar{x}_1$$
$$\dot{x}_1 = 4\left(\frac{1}{1+x_2^2}\right)\left(\frac{1}{1+\bar{x}_1{}^2}\right) - x_1$$
$$\dot{\bar{x}}_2 = 2\gamma\left(\frac{1}{1+x_2^2}\right) - \gamma\bar{x}_2$$
$$\dot{x}_2 = 4\left(\frac{1}{1+\bar{x}_1^2}\right)\left(\frac{1}{1+\bar{x}_2^2}\right) - x_2$$

$$\dot{\bar{x}}_1 = \gamma(2-x_1) - \gamma\bar{x}_1$$
$$\dot{x}_1 = (2-x_2)(2-\bar{x}_1) - x_1$$
$$\dot{\bar{x}}_2 = \gamma(2-x_2) - \gamma\bar{x}_2$$
$$\dot{x}_2 = (2-\bar{x}_1)(2-\bar{x}_2) - x_2$$

$$\dot{x}_1 = x_1 - x_1x_2$$
$$\dot{x}_2 = x_1x_2 - x_2$$

Figure 4. Lotka-Volterra (aka Predator-Prey) oscillator example. **(a)** Repressor network topology and **(c)** the dimensionless Hill-function model (in the form of eq. 3). **(b)** Numerical simulations of the ideal mass action Lotka-Volterra system (dashed) and the Hill-function repressor network from panel (c) (solid). Note that different initial conditions lead to oscillations of different amplitude. **(d)** Component-wise linear model (in the form of eq. 5).**(e)** For large γ, we can assume pseudo-equilibrium on $\bar{x}_1$ and $\bar{x}_2$, which yields the ideal Lotka-Volterra system. In simulations, the separation of time scales $\gamma = 100$ parameter was used.

Mass-action dynamical systems

Any mass-action dynamical system over N variables $x_1, \ldots, x_N$ (dimensionless) can be described by the following differential equations:

$$\dot{x}_i = p_i(x_1, \ldots, x_N) - d_i(x_1, \ldots, x_N)x_i \tag{8}$$

where p_i and d_i are polynomials with non-negative terms. Note that unlike the general class of autonomous polynomial ODEs, mass-action systems are restricted in the following way: Every negative term in the ODE for $\dot{x}_i$ must contain x_i as a factor. Mass-action systems naturally arise in chemical kinetics where the rate of the reactions consuming a species depend on the amount of that species. Mass-action systems are exactly the class of autonomous polynomial ODEs in which no variable x_i can become negative starting with arbitrary non-negative values of all variables. Since chemical concentrations are non-negative, this property provides an a priori justification for the appropriateness of mass-action models in chemistry, at least among the class of autonomous polynomial ODEs.

Example: Toward a systematic construction

While the examples in Section demonstrated some of the power of unsaturated repressor networks, they did not elucidate a general design process. In this section we consider the challenges encountered when trying to realize the mass-action Lotka-Volterra system (Fig. 4) as an unsaturated repressor network, and

show the reasoning that leads to the system in Fig. 4c. In the next section this approach is generalized to other mass-action systems.

Recall that the Lotka-Volterra oscillator is a mass-action system given by the following differential equations (where we take all rate constants to be 1):

$$\dot{x}_1 = x_1 - x_1 x_2 \tag{9}$$
$$\dot{x}_2 = x_1 x_2 - x_2 \tag{10}$$

Our goal is to construct a system in the component-wise linear form shown in eq. (5) that approximates the Lotka-Volterra system. We must also ensure that all variables (including new ones we will introduce) remain close to 1 when the oscillator operates in the desired domain (x_1, x_2 close to 1). The resulting system is then straightforward to implement using the Hill-function model of eq. (3), and thus can be considered realizable with an unsaturated repressor network.

The first step is to implement activation with repressors: we need, for example, x_2 to be positively affected by x_1. We can use a convenient property of order-2 Hill-functions that allows two repressors (say x_i and $\bar{x}_i$) wired serially to approximate linear activation (i.e. transcription rate proportional to x_i). Specifically, for each x_i we introduce a variable $\bar{x}_i$ with dynamics

$$\dot{\bar{x}}_i = \gamma \cdot (2 - x_i) - \gamma \bar{x}_i. \tag{11}$$

This equation is in the form of the component-wise linear model (5) with $n_{X_j} = 2$, $a_p = 2\gamma$, $b_i = \gamma$. For large γ, $\bar{x}_i$ converges to its pseudo-equilibrium value $2 - x_i$ on a faster time-scale than the dynamics of x_i. Thus we can use $(2 - \bar{x}_i)$ as a replacement for x_i. Observe that if $x_i \approx 1$ then $\bar{x}_i \approx 1$ as well, as required for the component-wise linear model to faithfully approximate the Hill-function model.

For the Lotka-Volterra system, with these new variables we can rewrite the second equation (eq. (10)) as:

$$\dot{x}_2 = (2 - \bar{x}_1)(2 - \bar{x}_2) - x_2 \tag{12}$$

which is now in the desired form (5), with $n_{X_j} = 2$, $a_p = 4\gamma$, $b_i = 1$. In the limit of large γ, the pseudo-equilibrium on the dynamics of $\bar{x}_1$ and $\bar{x}_2$ yields $\bar{x}_1 = 2 - x_1$ and $\bar{x}_2 = 2 - x_2$, and eq. (12) reduces to $\dot{x}_2 = x_1 x_2 - x_2$ as desired. (Note that the $(2 - x)$ terms produced from the component-wise linearization of order-2 Hill-functions are unique to satisfy two properties: Composition twice over yields the original x (as shown in (11) and (12)), and $(2 - x)$ is close to 1 when x is close to 1 (allowing $\bar{x}_i$ to be close to 1). This motivates our use of order-2 Hill-functions.)

The second challenge is to approximate the non-linear negative term $-x_1 x_2$ in eq. (9) of the Lotka-Volterra system. We define *complex degradation terms* to be the negative terms in the ODE for variable x_i that are non-linear in x_i (i.e. not of the form $-c \cdot x_i$ for some constant $c \geq 0$), and thus cannot be accounted for by linear degradation. For example, in the Lotka-Volterra system the term $-x_1 x_2$ of $\dot{x}_1$ is a complex degradation term. Complex degradation terms present a problem since our gene regulatory network model allows only a linear decay. Nonetheless, complex degradation terms can be indirectly accounted for in unsaturated repressor networks. Consider the following dynamics for x_i, which are in component-wise linear form:

$$\dot{x}_1 = (2 - x_2)(2 - \bar{x}_1) - x_1 \tag{13}$$
$$\approx (2 - x_2)x_1 - x_1 \tag{14}$$
$$= 2x_1 - x_2 x_1 - x_1 \tag{15}$$
$$= x_1 - x_2 x_1 \tag{16}$$

The approximation (14) follows by the pseudo-equilibrium of $\bar{x}_1$ as described above.

Note that the complex negative term arises in (15) as a result of applying the distributive law. Finally, note that to obtain (16) we cancel the positive term produced by the distributive law ($2x_1$) with the negative linear degradation x_1 term. The distributive law followed by cancelation with the linear degradation term can be universally applied to obtain arbitrary complex degradation terms in mass-action ODEs (as shown in the next section).

The restriction to mass-action systems seems natural. Suppose we wanted to implement the non-mass-action ODE $\dot{x}_1 = \cdots - x_2x_3 + \ldots$, where the complex degradation term does not contain x_1. To get the negative x_2x_3 term using the above technique would require using the distributive law on $(2 - x_2)x_3$ or $(2 - x_3)x_2$. But then, in addition to the desired negative x_2x_3 term, we would also get either a positive $2x_3$ or $2x_2$ term. If this extra term is not wanted, we cannot cancel it with the negative linear degradation term x_1.

General case: Design of unsaturated repressor networks for mass-action kinetics

In this section we show how given a mass-action system we can construct a component-wise linear mode that approximates it. This implies that in the regime where the component-wise linear model is a good approximation for the Hill-function repressor dynamics, the repressor network will approximate the target mass-action system. As before, the suitable regime is a sufficiently small neighborhood around $x_i = 1$ in dimensionless quantities. We suppose the goal mass-action system (8) and the initial conditions are such that $\forall x_i$, $1 - \delta \leq x_i \leq 1 + \delta$ for some constant $\delta < 1$ (eg. $\delta = 0.1$) at times of interest. We follow the strategy generalized from the Lotka-Volterra example in the previous section.

We describe a "minimal" way that unsaturated repressor networks are complete. While having control over all parameters of our model (eq. (3)) may enable easier construction of desired functions, we show that control over only one type of parameter is sufficient. In particular, among the parameters of the model (eq. (3)), we allow only specific control over production rate constant $a_p > 0$, but fix all degradation rate constants $b_i = b > 0$. Further, we use a universal Hill-coefficient $n_{X_j} = n = 2$. Then the component-wise linear model (eq. 5) simplifies to the following form, which we term the *standard* component-wise linear model:

$$\dot{x_i} = \sum_{p \in \mathcal{P}(X_i)} \frac{a_p}{2} \prod_{j \in \mathcal{R}(p)} (2 - x_j) - bx_i. \tag{17}$$

As before, the linear approximation holds when $x_i \approx 1$ using dimensionless quantities.

For polynomial p of positive monomials, let $\mathcal{M}(p)$ be the set of its monomials. For a monomial m, let $\mathcal{V}(m)$ be the multiset of variables (ie to account for x^n, x appears n times), and $c(m)$ be the (positive) multiplicative constant of the monomial.

Suppose we are given a target mass-action system described by eq. (8). We claim that for large γ, the following system converges to it, where we let $c(p) = \sum_{m \in \mathcal{M}(p)} c(m)$ be the sum of the coefficients of polynomial p:

$$\dot{\overline{x}}_i = \gamma \cdot (2 - x_i) - \gamma \overline{x}_i \tag{18}$$

$$\dot{r}_i = \sum_{m \in \mathcal{M}(d_i)} \frac{\gamma c(m)}{c(d_i)} \prod_{j \in \mathcal{V}(m)} (2 - \overline{x}_j) - \gamma r_i \tag{19}$$

$$\dot{x}_i = \sum_{m \in \mathcal{M}(p_i)} c(m) \prod_{j \in \mathcal{V}(m)} (2 - \overline{x}_j) + c(d_i)(2 - r_i)(2 - \overline{x}_i) + (\gamma - 2c(d_i))(2 - \overline{x}_i) - \gamma x_i \tag{20}$$

The above system has the property that, for $\gamma > 2c(d_i)$, it is of the form (17) and thus can be implemented as a gene regulatory network.

The following convergence argument is informal; for a more formal treatment singular perturbation theory may be used. First, we use a time separation argument with $\bar{x}_i$ and r_i as the fast variables, and x_i

as the slow variables. In the limit $\gamma \to \infty$, a pseudo-equilibrium is established for $\bar{x}_i$ and r_i as functions of x_i:

$$\bar{x}_i = 2 - x_i$$
$$r_i = \sum_{m \in \mathcal{M}(d_i)} \frac{c(m)}{c(d_i)} \prod_{j \in \mathcal{V}(m)} (2 - \overline{x}_j) = \frac{1}{c(d_i)} \sum_{m \in \mathcal{M}(d_i)} c(m) \prod_{j \in \mathcal{V}(m)} x_j$$

Plugging this into (20) yields:

$$\begin{aligned}
\dot{x}_i &= \sum_{m \in \mathcal{M}(p_i)} c(m) \prod_{j \in \mathcal{V}(m)} x_j + c(d_i) \left(2 - \frac{1}{c(d_i)} \sum_{m \in \mathcal{M}(d_i)} c(m) \prod_{j \in \mathcal{V}(m)} x_j \right) x_i + (\gamma - 2c(d_i)) x_i - \gamma x_i \\
&= \sum_{m \in \mathcal{M}(p_i)} c(m) \prod_{j \in \mathcal{V}(m)} x_j - \sum_{m \in \mathcal{M}(d_i)} c(m) \prod_{j \in \mathcal{V}(m)} x_j x_i \\
&= p_i(x_1, \ldots, x_N) - d_i(x_1, \ldots, x_N) x_i
\end{aligned}$$

This is exactly the form of the original mass-action system (8).

A few properties bear mentioning. If x_i in equation 18 is 1 then the pseudo-equilibrium of $\bar{x}_i$ is $\bar{x}_i = 1$. Similarly, if all x_j in equation 19 are 1, then the pseudo-equilibrium of r_i is $r_i = 1$, no matter what the polynomial coefficients are. This means that $\bar{x}_i$ and r_i are in the linear regime (near 1) if the variables they depend on are as well. Finally, if all the variables of the mass-action system (eq 8) remain close to 1 then x_i will remain near 1 in equation 20, satisfying the condition necessary for the standard component-wise linear model to correspond to the Hill-function model (3).

Discussion

In this work, we provide a proof of principle that gene regulatory networks are a natural substrate for analog computation. Thus engineering the key building blocks for our construction — a set of independent repressors that exhibit a large linear regime of activity — could advance analog design in synthetic biology.

The proposed systematic construction has a few important limitations. The promoter production rates have to be tuned carefully to ensure that the system operates in the unsaturated regime. Luckily, however, there are ways to expand the linear regime. If the intermediate transcription factor binding states are stable, a complex rational function is a more accurate representation of the rate of transcription. In this case, the unsaturated regime can be expanded thus broadening the functional parameter regime (Fig. 5a). In addition, the size of the linear regime can increase with the number of non-specific transcription factor binding sites as shown in Fig. 5b [2].

To achieve the full generality of mass-action kinetics requires the precise tuning of transcription rate constants α_p. In particular, the argument of section relies on the exact cancelation of certain terms in the component-wise linearization. Nonetheless, the behavior of some systems is robust to imperfect cancellation. For example, imperfect cancellation in the Lotka-Volterra system changes the rate constants of the ideal system, while preserving the overall neutrally stable behavior. It remains to demarcate the class of systems that can be implemented with unsaturated repressor in a manner that is robust to model parameters. Our construction also relies on a time-scale separation approximation leading to necessarily slow time scales of the dynamics of the system.

Since stochastic fluctuations in molecular concentrations are pervasive in the cell, parameter tuning required by our model could be difficult to achieve due to this molecular noise. Nevertheless, stochastic simulations of the unsaturated repressor network Lotka-Volterra system demonstrates that realistic parameters ranges can generate neutral cycle oscillatory dynamics (Fig. 6). Matching the deterministic

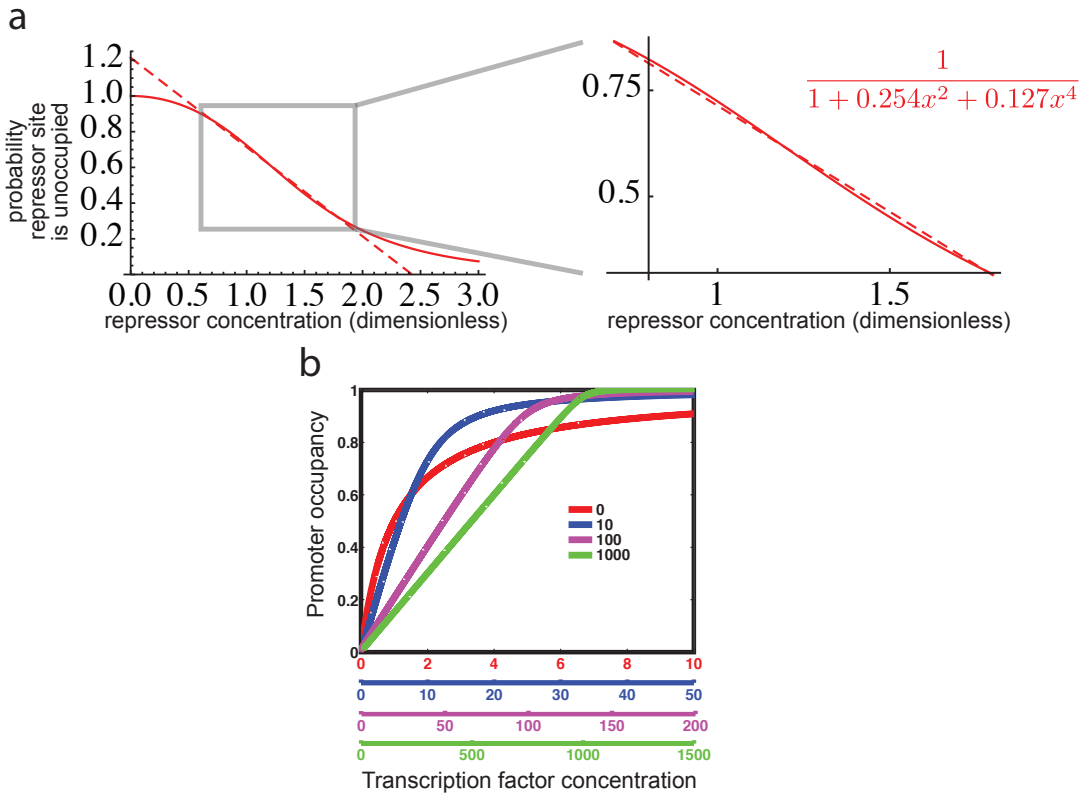


Figure 5. Expanding the unsaturated operating regime. **(a)** More complex rational functions can have a larger linear unsaturated regime than Hill-functions. The gray box illustrates the region of good linear approximation. We plot the dimensionless quantity $x = X/k$, where k is the binding constant k. **(b)** Increasing the number of non-specific transcription factor binding sites expands the linear regime. Total concentration of transcriptional activator is plotted as a function of the promoter activity (as described in [36]). Parameters include Kn = 1 and Kp = 1.

model, this stochastic implementation behaves as a neutral cycle oscillator since variation in the initial conditions changes the amplitude and frequency.

It is likely that there are many ways to implement mass-action systems with transcriptional networks. For example, the same Lotka-Volterra oscillator can be constructed in a different way using a combination of repressors and activators (Fig. 7). In contrast to unsaturated repressor networks, the repressors in Fig. 7 operate in the saturated regime where promoter activity scales as the reciprocal of the repressor concentration. A more complete theory would utilize both the saturated and unsaturated regimes of the Hill function for analog computation. Further, as the first two examples in Section suggest, a certain class of non-mass-action systems can be directly implemented by unsaturated repressor networks. However, more work needs to be done to ascertain exactly the boundaries of this class as well as to elucidate the distinguishing properties of the dynamical systems in it.

The power of post-transcriptional networks such as phosphorylation circuits has not yet been systematically explored for analog computation. Beyond the relatively faster timescale of post-transcriptional networks, it is not clear whether transcriptional or post-transcriptional regulation is more suitable for analog computation.

Our choice of the model is dictated in part by the desire to understand the computational power

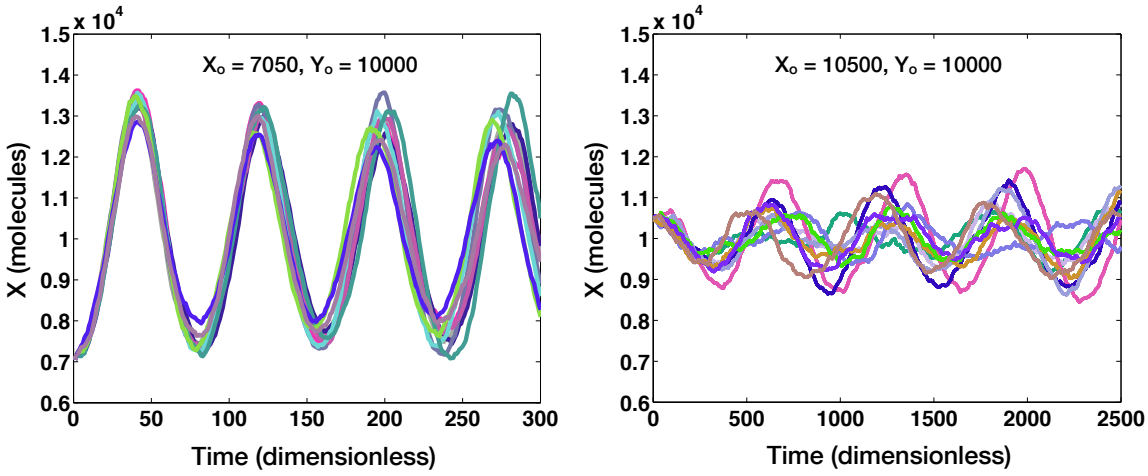


Figure 6. Stochastic simulations the predator prey unsaturated repressor network (equations in Fig. 4C) for two different initial conditions. Note that different initial conditions lead to oscillations of different amplitudes matching the deterministic model. Numerical simulations were performed using StochKit [29].

of transcriptional regulation in biology. In demarcating the range of system behaviors obtainable with transcriptional regulation, our work provides insight into the types of regulation that natural biological systems use for implementing specific functions.

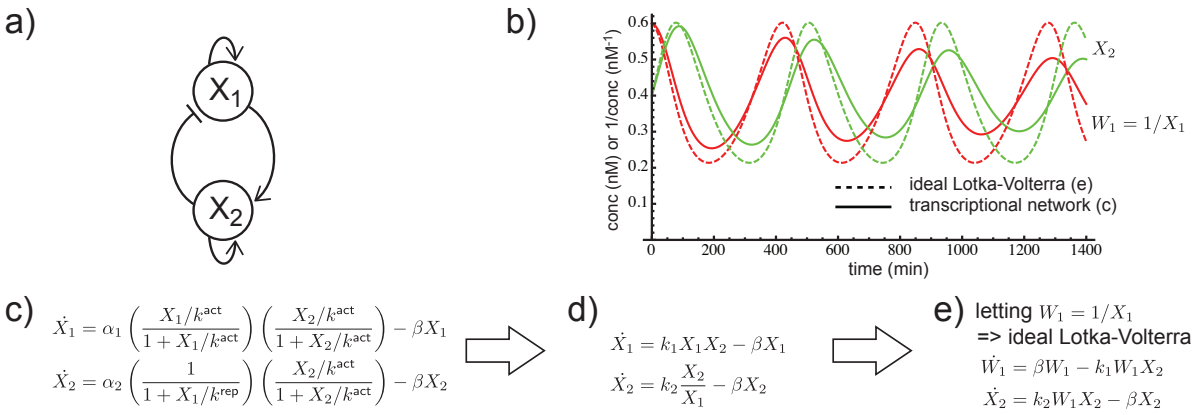


Figure 7. Implementation of the Lotka-Volterra oscillator by recasting the reciprocal of a transcription factor as a variable of the target dynamical system. **(a)** Transcriptional network topology. **(b)** Numerical simulation of the ideal Lotka-Volterra model (dashed line) and the Hill-function transcriptional network from panel (c). **(c)** Transcriptional network model. Unlike in the previous examples, both repression and activation is used, and the Hill-coefficient is 1. **(d)** Simplification of the model in panel (c) using the following approximations. First, the order-1 activation Hill-function $\frac{X/k^{\text{act}}}{1+X/k^{\text{act}}} \approx X/k^{\text{act}}$ when X is small compared to k^{act}. Second, the order-1 repression Hill-function $\frac{1}{1+X/k^{\text{rep}}} \approx k^{\text{rep}}/X$ when X is large compared with k^{rep}. **(e)** The differential equations in panel (d) are exactly equivalent to the Lotka-Volterra system of equations with the substitution $W_1 = 1/X_1$. This can be confirmed with the reciprocal rule of differentiation. A downstream transcriptional network can use X_1 as a repressor in an order-1 Hill function to "read-out" the reciprocal value W_1. Parameters are: maximum transcription rate $\alpha_1 = 100$ nM min^{-1}, $\alpha_2 = 20$ nM min^{-1}, binding constants $k^{\text{act}} = 50$ nM, $k^{\text{rep}} = 0.1$ nM, and degradation rate $\beta = 0.015$ min^{-1}. The concentration range of transcription factors was $0.2 - 5$ nM throughout.

Acknowledgments

We would like to thank Wendell Lim, Hana El-Samad and Jaijeet Roychowdhury for helpful discussions.

References

1. Rahul Sarpeshkar, "Analog versus digital: extrapolating from electronics to neurobiology", *Neural Computation* **10**(7), pp. 1601–1638 (1998).

2. Ramiz Daniel, Jacob R Rubens, Rahul Sarpeshkar, and Timothy K Lu, "Synthetic analog computation in living cells", *Nature* **497**, pp. 619–623 (2013).

3. Y Setty, AE Mayo, MG Surette, and U Alon, "Detailed map of a cis-regulatory input function", *Proceedings of the National Academy of Sciences* **100**(13), pp. 7702–7 (2003).

4. Jacob Stewart-Ornstein, Christopher Nelson, Joe DeRisi, Jonathan S Weissman, and Hana El-Samad, "Msn2 coordinates a stoichiometric gene expression program", *Current Biology* **23**, pp. 2336–2345 (2013).

5. Claude E Shannon, "A symbolic analysis of relay and switching circuits", *Electrical Engineering* **57**(12), pp. 713–723 (1938).

6. David G Hoag, "The history of Apollo onboard guidance, navigation, and control", *Journal of Guidance, Control, and Dynamics* **6**(1), pp. 4–13 (1983).

7. Alvin Tamsir, Jeffrey J Tabor, and Christopher A Voigt, "Robust multicellular computing using genetically encoded NOR gates and chemical 'wires"', *Nature* **469**(7329), pp. 212–215 (2010).

8. Claude E Shannon, "Mathematical theory of the differential analyzer", *J. Math. Phys. MIT* **20**, pp. 337–354 (1941).

9. Makoto Itoh, "Synthesis of electronic circuits for simulating nonlinear dynamics", *International Journal of Bifurcation and Chaos* **11**(03), pp. 605–653 (2001).

10. Lei S Qi, Matthew H Larson, Luke A Gilbert, Jennifer A Doudna, Jonathan S Weissman, Adam P Arkin, and Wendell A Lim, "Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression", *Cell* **152**(5), pp. 1173–1183 (2013).

11. Virgil A Rhodius, Thomas H Segall-Shapiro, Brian D Sharon, Amar Ghodasara, Ekaterina Orlova, Hannah Tabakh, David H Burkhardt, Kevin Clancy, Todd C Peterson, Carol A Gross, et al., "Design of orthogonal genetic switches based on a crosstalk map of σs, anti-σs, and promoters", *Molecular Systems Biology* **9**, pp. 702 (2013).

12. Brynne C Stanton, Alec AK Nielsen, Alvin Tamsir, Kevin Clancy, Todd Peterson, and Christopher A Voigt, "Genomic mining of prokaryotic repressors for orthogonal logic gates", *Nature Chemical Biology* **10**, pp. 99–105 (2013).

13. Lacramioara Bintu, Nicolas E Buchler, Hernan G Garcia, Ulrich Gerland, Terence Hwa, Jané Kondev, and Rob Phillips, "Transcriptional regulation by the numbers: models", *Current Opinion in Genetics & Development* **15**(2), pp. 116–124 (2005).

14. Hernan G Garcia and Rob Phillips, "Quantitative dissection of the simple repression input–output function", *Proceedings of the National Academy of Sciences* **108**(29), pp. 12173–12178 (2011).

15. Adam Arkin and John Ross, "Computational functions in biochemical reaction networks", *Biophysical Journal* **67**(2), pp. 560–578 (1994).

16. Irving R Epstein and John A Pojman, *An Introduction to Nonlinear Chemical Dynamics: Oscillations, Waves, Patterns, and Chaos (Topics in Physical Chemistry)*, Oxford University Press (1998).

17. David Soloveichik, Georg Seelig, and Erik Winfree, "DNA as a universal substrate for chemical kinetics", *Proceedings of the National Academy of Sciences* **107**(12), pp. 5393–5398 (2010).

18. Kevin Oishi and Eric Klavins, "Biomolecular implementation of linear I/O systems", *Systems Biology, IET* **5**(4), pp. 252–260 (2011).

19. Yuan-Jyue Chen, Neil Dalchau, Niranjan Srinivas, Andrew Phillips, Luca Cardelli, David Soloveichik, and Georg Seelig, "Programmable chemical controllers made from DNA", *Nature nanotechnology* **8**(10), pp. 755–762 (2013).

20. SR Paladugu, V Chickarmane, A Deckard, JP Frumkin, M McCormack, and HM Sauro, "In silico evolution of functional modules in biochemical networks", *Systems Biology IEE Proceedings* **153**(4), pp. 223–35 (2006).

21. W Ma, A Trusina, H El-Samad, WA Lim, and C Tang, "Defining network topologies that can achieve biochemical adaptation", *Cell* **138**(4), pp. 760–73 (2009).

22. P Francois and V Hakim, "Design of genetic networks with specified functions by evolution in silico", *Proceedings of the National Academy of Sciences* **101**(2), pp. 580–5 (2004).

23. WA Lim, CM Lee, and C Tang, "Design principles of regulatory networks: searching for the molecular algorithms of the cell", *Molecular Cell* **49**(2), pp. 202–12 (2013).

24. AH Chau, JM Walter, J Gerardin, C Tang, and WA Lim, "Designing synthetic regulatory networks capable of self-organizing cell polarization", *Cell* **151**(2), pp. 320–32 (2012).

25. Timothy S Gardner, Charles R Cantor, and James J Collins, "Construction of a genetic toggle switch in *Escherichia coli*", *Nature* **403**(6767), pp. 339–342 (2000).

26. SM Cory and TJ Perkins, "Implementing arithmetic and other analytic operations by transcriptional regulation", *PLoS Computational Biology* **4**(4), pp. e1000064 (2008).

27. Robert Sidney Cox, Michael G Surette, and Michael B Elowitz, "Programming gene expression with combinatorial promoters", *Molecular Systems Biology* **3**, pp. 145 (2007).

28. Chang C Liu, Lei Qi, Julius B Lucks, Thomas H Segall-Shapiro, Denise Wang, Vivek K Mutalik, and Adam P Arkin, "An adaptor from translational to transcriptional control enables predictable assembly of complex regulation", *Nature Methods* **9**(11), pp. 1088–1094 (2012).

29. Kevin R Sanft, Sheng Wu, Min Roh, Jin Fu, Rone Kwei Lim, and Linda R Petzold, "StochKit2: software for discrete stochastic simulation of biochemical systems with events", *Bioinformatics* **27**(17), pp. 2457–2458 (2011).

30. Lea Goentoro, Oren Shoval, Marc W Kirschner, and Uri Alon, "The incoherent feedforward loop can provide fold-change detection in gene regulation", *Molecular Cell* **36**(5), pp. 894–899 (2009).

31. Karl Johan Aström and Richard M Murray, *Feedback systems: an introduction for scientists and engineers*, Princeton University Press (2010).

32. MB Elowitz and S Leibler, "A synthetic oscillatory network of transcriptional regulators", *Nature* **403**(6767), pp. 335–8 (2000).

33. Frederick K Balagaddé, Hao Song, Jun Ozaki, Cynthia H Collins, Matthew Barnet, Frances H Arnold, Stephen R Quake, and Lingchong You, "A synthetic *Escherichia coli* predator–prey ecosystem", *Molecular Systems Biology* **4**, pp. 187 (2008).

34. M.D. Korzuhin, "Mathematical modeling of the kinetics of homogenous chemical systems, ii-iii (in Russian)", In *Oscillatory Processes in Biological and Chemical Systems (in Russian)* , pp. 231–251. Nauka (1967).

35. W. Klonowski, "Simplifying principles for chemical and enzyme reaction kinetics", *Biophys Chem* **18**(2), pp. 73–87 (1983).

36. Tek-Hyung Lee and Narendra Maheshri, "A regulatory role for repeated decoy transcription factor binding sites in target gene expression", *Molecular Systems Biology* **8**, pp. 576 (2012).