Bayesian and Algebraic Strategies to Design in Synthetic Biology

This article provides an overview of two complementary approaches (algebraic and Bayesian) for rational design in synthetic biology.

By ROBYN P. ARAUJO , SEAN T. VITTADELLO , AND MICHAEL P. H. STUMPF

ABSTRACT | Innovation in synthetic biology often still depends on large-scale experimental trial and error, domain expertise, and ingenuity. The application of rational design engineering methods promises to make this more efficient, faster, cheaper, and safer. However, this requires mathematical models of cellular systems. For these models, we then have to determine if they can meet our intended target behavior. Here, we develop two complementary approaches that allow us to determine whether a given molecular circuit, represented by a mathematical model, is capable of fulfilling our design objectives. We discuss algebraic methods that are capable of identifying general principles guaranteeing desired behavior; we provide an overview of Bayesian design approaches that allow us to choose a model that has the highest probability of fulfilling our design objectives from a set of models. We discuss their uses in the context of biochemical adaptation and, then, consider how robustness can and should affect our design approach.

KEYWORDS | Adaptation; approximate Bayesian computation (ABC); robustness; synthetic biology.

I. INTRODUCTION

Trial and error coupled with ingenuity have arguably been responsible for many important innovations in biotechnology. Unfortunately, however, these do not scale. Instead, recent progress in synthetic biology is generally based on rational design engineering approaches [1], [2]. Design in cellular and molecular biology cannot be based directly on the same physical and chemical principles at work in technological systems. Naturally evolved systems are just too complex; our knowledge of their structure and dynamics too primitive, and our ability to interfere with and control their behavior is still very limited.

Here, we discuss a set of complementary approaches that can be used in order to design synthetic biological systems that exhibit certain types of behavior or fulfill certain types of biological functions. We need to acknowledge the differences between biological and technical systems and our ignorance of the details of the molecular mechanisms driving important cellular functions and phenotypes. In light of this, we need better ways of developing meaningful, realistic, reliable, and validated mathematical models. In the context of synthetic biology, we need quick and robust ways of picking those models that can meet our design targets and objectives. The latter is the main purpose of this article.

In the following, we first outline the problem of design in synthetic biology: we are interested in dynamical systems, implemented in synthetically engineered molecular circuits that fulfill certain design objectives. One overarching problem is that they need to function robustly and reliably on top of cellular physiology. We use modeling or *in silico* analysis as a way of triaging the vast potential design space in order to identify the best designs, and we will attempt to outline this problem and the confounding mathematical factors in the necessary detail, before illustrating two complementary strategies of exploring this design space. One way of assessing the promise of designs is to assess their level of robustness to uncertainties in the models that

Robyn P. Araujo is with the School of Mathematical Sciences, Queensland University of Technology, Brisbane, QLD 4000, Australia.
Sean T. Vittadello and **Michael P. H. Stumpf** are with the School of BioSciences and the School of Mathematics and Statistics, The University of Melbourne, Parkville, VIC 3010, Australia (e-mail: mstumpf@unimelb.edu.au).

we investigate, and we will make more precise what we mean by robustness.

We will provide an overview of how algebraic approaches can be used to distill design principles for, at least, certain types of behavior. In particular, we illustrate how such an approach helps to identify absolute requirements for a dynamical system to exhibit a fundamentally important network response known as robust perfect adaptation (RPA) [3]. Algebraic approaches are capable of yielding qualitative insights; Bayesian design methods provide an alternative approach [4]. Unlike the algebraic approaches, they are computationally costly but provide rational criteria for choosing from among competing models. We illustrate, in general terms, how this can be used in practice, again in the context of adaptive responses.

We conclude by a discussion of how the robustness of mathematical models can reassure us that a design may function in practice. Robustness takes on different meanings in this article: RPA in a feature of biological systems; Bayesian and—where feasible—algebraic approaches allow us to design systems that exhibit the certain desired behavior, robustly. That is, even though our knowledge of the system is incomplete, we expect the system to behave in an intended way. Section V aims to make these concepts more precise. We argue that, taken together, algebraic methods, Bayesian design, and robustness offer enormous scope for enabling rational design in synthetic biology.

II. DESIGN IN SYNTHETIC BIOLOGY: OBJECTIVES AND STRATEGIES

In synthetic biology, we are interested in identifying a (dynamical) system that reliably produces a certain type of target behavior. For example, for a given input, x, our system state y should produce a target output or behavior, T. Both x and T can be vector-valued though often they may not be, but y, the system state, will be a vector, including all the relevant molecular abundances of mRNA, proteins, posttranslationally modified proteins, metabolites, and so on, involved in a given process.

We consider systems of the following form:

$$\frac{dy}{dt} = f(y; x, t, \theta) \tag{1}$$

where θ denotes the system parameters (which are typically unknown). We want to determine if the system given by $f(y; x, t, \theta)$ is capable of producing the desired output, T, and meet the design objectives.

In our applications, we may be interested in systems that can exhibit switch-like behavior, filter out high-frequency noise, and exhibit perfect adaptation, or more complex phenotypes, such as the ability of a cellular system to produce Turing pattern behavior. Specifying the design objectives, T, can involve qualitative characteristics or quantitative aspects of system behavior. Often, we may resort to surrogate data that capture the desired behavior; this is especially the case if precise mathematical or algebraic definitions of system behavior are difficult

(see [4]) or if there are no simple qualitative features (such as algebraic identities).

Model-based design in synthetic biology can be done in different ways: 1) we can use domain expertise to formulate mathematical models, $f(y; x, t, \theta)$, determine if they fulfill our design objectives, and optimize the performance by varying θ or by making small alterations to the model; 2) choose from a set of models $\mathcal{F} = \{f_1, f_2, \ldots, f_N\}$ using, for example, statistical model selection in order to determine which models are best able to deliver the design objectives; 3) explore exhaustively a class of relevant models, that is, now, $\mathcal{F}$ might contain all models, f, with two, three, or four species or nodes (the dimension of the vector y) to determine if they meet the design objectives; and 4) identify universal characteristics, guaranteeing that the design objectives are met and construct models (and, as the end-target, synthetic biological systems) embodying these principles. The first approach relies a lot on luck in getting the initial model right. The second and third approaches can become computationally prohibitively expensive even for small networks; they are, however, flexible and widely applicable. The final approach is less *ad hoc* than the others and can provide guarantees on the behavior of designs. However, it requires a more grounded theoretical understanding of the design objectives than we often have.

Of course, there are also ways in which we can combine the last three approaches to gain confidence in our analyses and designs and to streamline the design process. If, for example, design principles are known, we can use any such theoretical principles to reduce the set of potential candidate models, $\mathcal{F}$, potentially quite considerably.

III. ALGEBRAIC APPROACHES TO DISCOVERING DESIGN PRINCIPLES

In this section, we consider algebraic approaches that may be able to identify, within the vast potential design space, the exceedingly limited regions that contain the "best" designs—those that can perform an essential biological function in a robust and reliable manner. We emphasize that such analytical methods may not be applicable to the study of all possible design objectives and may not be well suited to the elucidation of complex biological behaviors, such as multistationarity [5] or Turing patterning [6]. However, algebraic (and ultimately topological) methods are beginning to show tremendous promise in the context of robust asymptotic tracking problems and have recently delineated the full solution space for network topologies capable of RPA [3]. RPA is of fundamental significance in biological systems, both natural and synthetic, and has been widely observed in cellular signaling networks [3], [7], [8], sensory systems [9], [10], in the regulation of bacterial chemotaxis [11], [12], bacterial stress response [13], [14], intracellular osmolarity [15], sigma-70 activity [16], and homeostatic control of plasma mineral concentrations [17], [18], and is thought to play a role in robust patterning during organism development [19], [20].

A. Control Theory Viewpoint on RPA

RPA is the ability of a system to generate an output that returns to a fixed reference level (its "setpoint") following a persistent change in the system input, with no need for parameter fine-tuning [3]. RPA is, thus, equivalent to the well-established control-theoretic notion of *robust steady-state tracking* or *constant disturbance rejection*, while absolute concentration robustness (ACR, a special type of RPA [21]) is equivalent to the control-theoretic concept of *robustness to disturbances in initial conditions* [22]. This fundamental problem of robust asymptotic tracking of the desired trajectory, while rejecting unwanted disturbances, is one of the defining problems of classical automatic control [23]. In the 1970s, the landmark studies of Francis and Wonham [24], [25] investigated the necessary controller structures to achieve robust regulation of a system, with internal stability, and established what is now known as the internal model principle (IMP). By this principle, a control system is able to reject exogenous stimuli or disturbances by incorporating within itself a model of the dynamic structure of the network stimuli. For RPA, then, the internal model must produce constant signals and is equivalent to the requirement for integral (feedback) control [26]. Thus, in all RPA-capable biochemical reaction networks, we expect an integrator—an internal model of a system that produces constant inputs or disturbances—to be embedded or concealed within the biochemical reaction rates that govern the dynamical response of the network.

Suppose that we have a n-dimensional system of interacting molecules, with state vector $x(t) = [x_1, \ldots, x_n]^{\mathrm{T}}$ representing the concentrations of the molecules. In general, we seek some change of coordinates that identifies a subsystem, z, of the form

$$\frac{dz}{dt} = p(x)(\hat{x} - c) \qquad (2)$$

where z is an *output-driven internal model* [26] of persistent disturbances to the system. The *output* in this case is $\hat{x} \in \{x_1, \ldots, x_n\}$, which represents the concentration of the RPA-capable molecule, while c (the *setpoint* of the system) is some function of system parameters. The function p may be zero-order in its arguments, in which case RPA is "unconstrained" by any other system properties. In all other cases, (2) encodes a concept known as *constrained integral control* [21], [22], in which case RPA obtains only if $p(x) \neq 0$.

In simple biochemical network designs, a linear change of coordinates may be sufficient to identify the internal model such that an integrator of the form in (2) lies in the rowspan of the system. Antithetic integral control [16] represents a well-known example of this type and is, topologically speaking, a feedback-regulated system. A simple representation of this control motif is depicted in Fig. 1(a), with four interacting molecules with concentrations $A, B, C,$ and D. From the reaction forms shown, it
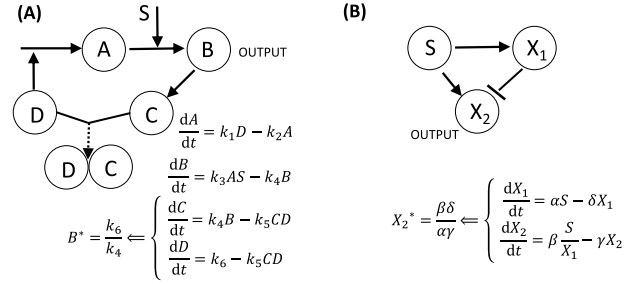

Fig. 1. *Two simple motifs from which an internal model may be identified through a simple change of coordinates. (a) Antithetic integral control motif [16], as discussed by Ferrell, Jr, [28]. (b) Incoherent feedforward motif [26].*

is clear that

$$\frac{dc}{dt} - \frac{dD}{dt} = k_4 \left(B - \frac{k_6}{k_4} \right)$$

which captures the internal model explicitly since $B^* = k_6/k_4$ represents the setpoint of the molecule B. New methods have recently been developed to detect an internal model in the rowspan of an RPA (or ACR)-capable system, if such exists, in more complex networks of biochemical interactions [22], [27].

Feedforward structured networks can also be recast as integral feedback systems through a judicious (generally nonlinear) change of coordinates. Such a nonlinear mapping may be difficult to identify for complex feedforward-structured networks but may be readily obtained for particularly simple feedforward motifs. For instance, Shoval *et al.* [29] consider a minimal incoherent feedforward motif of the form

$$\frac{dX_1}{dt} = \alpha S - \delta X_1$$
$$\frac{dX_2}{dt} = \beta \frac{S}{X_1} - \gamma X_2$$

as depicted in Fig. 1(b), and show that the map

$$(X_1, X_2) \to (z_1, z_2) = (X_2, \alpha X_2 - \beta \log X_1)$$

is a diffeomorphism with inverse $X_1 = e^{(\alpha z_1 - z_2)/\beta}$ and $X_2 = z_1$, which transforms the reaction equations into the partitioned form

$$\frac{dz_1}{dt} = \beta S e^{(z_2 - \alpha z_1)/\beta} - \gamma z_1$$
$$\frac{dz_2}{dt} = \beta \delta - \alpha \gamma z_1.$$

The second of these equations captures the internal model since the solutions to $dz_2/dt = 0$ represent all possible

constant signals and determine the setpoint of the system to be $z_1 = X_2 = \delta\beta/\alpha\gamma$.

The recognition that RPA may be achieved even for very complex networks by adding an integrator to the network within a feedback structure has been of tremendous utility for synthetic implementations [8]. On the other hand, this recognition alone does not provide a straightforward tool for exploring a more comprehensive design space of RPA-capable networks—particularly in view of the fact that the connection to integral feedback regulation may be very complex for networks without a clear (physical) feedback structure (such as feedforward structured networks) [26]. Moreover, there is now strong experimental evidence that RPA, in natural (endogenous) cellular chemical reaction networks, frequently requires multiple adaptation mechanisms working together in concert [30]. Even in the relatively simple chemotaxis-regulating cascades in organisms, such as Dictyostelium, for example, it appears that multiple distinct RPA modules are at work [30], making it difficult to propose synthetic implementations of these kinds of network structures relying only on the concept of integral control.

In Section III-B, we consider a complementary approach to the integral control viewpoint, by considering how an algebraic representation of the RPA problem may be analyzed in such a way as to reveal the complete, comprehensive design space of RPA-capable network topologies. These methods identify RPA-capable network structures directly and retain a representation of the RPA problem in terms of the original coordinates (i.e., the concentrations of the interacting molecules). Rather than extracting a single internal model for the constant input stimuli to an RPA-capable system, this approach reveals how such an internal model could be *distributed* topologically over potentially vast and intricate networks of interacting molecules.

B. Topological RPA Problem

Many algebraic formulations for the RPA problem have been proposed [3], [31], [32] and have been shown to be equivalent to the requirement for integral control [3], [31]. Commonly, these formulations allow the network to have multiple "input nodes," i.e., molecules that are subjected to stimulation or perturbation by an external source, as well as multiple "output nodes," which are the molecular endpoints of interest, which are expected to exhibit the RPA property. However, to delineate the full solution space of RPA-capable network topologies from such an algebraic condition, it is expedient to start from a simpler assumption: a network of interacting nodes that are perturbed at a *single* input node, and for which a *single* output node is monitored for the RPA property, thereby allowing network topologies to be *induced* relative to the designated input–output node pair [3]. For an RPA-capable network topology identified by this method, there may exist additional nodes to which a stimulus may be applied

while preserving the RPA property at the designated output node. Likewise, an RPA-capable network design may also exhibit RPA at nodes other than the designated output node. In this way, a complete picture of RPA-capable multi-input/multi-output network designs emerges.

The topological RPA problem has now been solved in complete generality [3] and is the fulfillment of a long-standing conjecture in biological network theory [33]–[37] that held that robust biosystems are likely to require very specific network architectures—a conjecture based largely on analogies with engineering control systems and the well-defined design features that are known to be required, or at least beneficial, in those contexts. The essential ideas underlying this general solution are given as follows.

Suppose that we have a network consisting of n interacting molecules, or *nodes*, denoted $x_1, x_2, \ldots, x_n$. One of these nodes (x_i, say) will be selected as the designated input node, and one node (x_j, not necessarily distinct from x_i) will be selected as the output. To each noninput node, we associate a reaction rate $dx_k/dt = f_k(x_1, \ldots, x_n)$, which encodes both the strength and the nature (stimulating or inhibitory) of the intermolecular interactions upon the node x_k. The input node is also regulated by an external stimulus, S, and has a reaction rate $dx_i/dt = f_i(x_1, \ldots, x_n, S)$. From this simple mathematical representation of the interactions in an n-node network, and its encounter with an external stimulus, it is straightforward to show that the network is capable of RPA at node x_j only if

$$\frac{\mathrm{Det}(M_{ij})}{\mathrm{Det}(J_n)} = 0 \tag{3}$$

where $J_n = (\partial(f_1, \ldots, f_n))/(\partial(x_1, \ldots, x_n))$ denotes the system Jacobian and M_{ij} is the Jacobian minor obtained by eliminating the input row and the output column from J_n. Thus, a necessary condition for RPA in *any* network, regardless of size or complexity, is

$$\mathrm{Det}(M_{ij}) = 0 \tag{4}$$

with

$$\mathrm{Det}(J_n) \neq 0. \tag{5}$$

Equation (4) is known as the RPA equation [3], which must be satisfied in any RPA network, subject to the RPA constraint [see (5)]. The RPA equation, thereby, provides an algebraic condition that allows us to *derive* structural conditions on RPA-capable networks, such that the RPA property will obtain at the designated output node (x_j) for any possible steady-state of the system and for any possible parameter choices, at least in principle, assuming that the system steady state can remain in the positive orthant for the range of values of S being considered.

The cornerstone of the methodology developed in [3] is the interpretation of the determinant of M_{ij} not simply as a function $\mathrm{Det}: \mathbb{R}^{(n-1)\mathrm{x}(n-1)} \to \mathbb{R}$ but as a *set* of τ terms, $1 \leq \tau \leq (n-1)!$, with each term carrying either a positive or a negative sign arising from its relationship
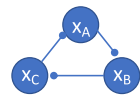
Topological Feature	FACTOR appearing in a TERM of the RPA Equation	Network Schematic representation of FACTOR
ROUTE SEGMENT	$\dfrac{\partial f_B}{\partial x_A}\dfrac{\partial f_C}{\partial x_B}\dfrac{\partial f_D}{\partial x_C}$	$x_A \bullet\!\!-\!\!\bullet x_B \quad x_C \bullet\!\!-\!\!\bullet x_D$
FEEDBACK LOOP	$\dfrac{\partial f_B}{\partial x_A}\dfrac{\partial f_C}{\partial x_B}\dfrac{\partial f_A}{\partial x_C}$	$x_A,\ x_C,\ x_B$
BIOCHEMICAL REACTION MECHANISM	$\dfrac{\partial f_A}{\partial x_A}\dfrac{\partial f_B}{\partial x_B}\dfrac{\partial f_C}{\partial x_C}$	no contribution to network interconnections

Fig. 2. *Relationship of the factors within individual terms of the RPA equation to the topology of the underlying network. Here, we present three examples of a triplet of factors (partial derivatives) that could appear within the total of (n − 1) factors in a term of the RPA equation of an n-node network. The presence of a partial derivative of the form (∂f_B)/(∂x_A) in terms of the RPA equation means that node x_A regulates x_B in the underlying network. If x_B is not regulated by x_A, (∂f_B)/(∂x_A) will be identically zero, and the corresponding term will not appear in the RPA equation. Factors of the form (∂f_A)/(∂x_A) encode the dependence of the reaction rate of x_A (f_A = (dx_A)/(dt)) on its own molecular concentration, x_A, and, thus, encode information on reaction kinetics rather than network interconnections.*

to the determinant expansion and from the nature (either inhibitory or activating) of the internodal interactions encoded by the partial derivatives in each term (see [3] for full details). Indeed, each of the $(n-1)!$ signed terms of the RPA equation is a product of $(n-1)$ partial derivatives, each encoding information on either the strength of the interactions between molecules or the reaction kinetics of the molecules. In this way, each term constitutes a representation of a "route" of signal flow through the network from the input node to the output node. The contents of each term, thereby, correspond, in a very precise way, to topological characteristics of the underlying network, representing either structural features of the network or constraints on reaction kinetics (see Fig. 2).

We note that the RPA equation will, in general, contain a very large number of terms, depending on the number of nodes and the interconnectedness of the network, highlighting the fact that the topological RPA problem quickly eludes computational screening approaches as n grows. Indeed, for a fully connected network, where every node regulates every other node, there will be $(n-1)!$ terms in the RPA equation.

It is clear from this set-theoretic interpretation of the RPA equation that partitions of the set may exist, for which the contents of every subset can sum to zero independently of the contents of every other subset [3]. By identifying general conditions under which such "independently adapting subsets" can exist, we can determine, definitively, the topological structure that *any* RPA-capable network must possess, no matter how large or complex the network. In particular, it has been shown [3] that there exist *two and only two* possible types of "independently adapting

subset" of the RPA equation, known as S-sets and M-sets. S-sets contain the terms that are all identically zero, for all possible values of input stimulus, and all possible parameter values, and arise when a feedback structure exists within the associated network, and where one or more nodes embedded into that feedback structure satisfy a special constraint on reaction kinetics known as *Opposer kinetics*. These S-sets correspond to the presence of a special subnetwork structure known as an Opposer module [see Fig. 3(a)]. By contrast, M-sets contain terms that are all strictly nonzero but, together, sum to zero for all possible values of input stimulus and all possible parameter values. M-sets arise when a collection of parallel pathways exists in the network, and where the nodes embedded into those parallel pathways satisfy a special constraint on reaction kinetics known as *Balancer kinetics*. In addition, the convergence point for these parallel pathways must also satisfy a constraint on reaction kinetics known as *Connector kinetics* in order to complete the "balancing act" of the M-set. M-sets correspond to the presence of a special subnetwork structure known as a Balancer module [see Fig. 3(b)].

Opposer and Balancer modules are generalizations of the simple three-node RPA-capable motifs identified by Ma *et al.* [38] through computational screening methods. However, these more general RPA-promoting modules may be vastly more complex and can contain previously unknown topological features, such as *Opposing sets*. We refer interested readers to [3] and [39] for a comprehensive description of Opposer and Balancer modules.

These two RPA modules, thus, represent a *topological basis* for all RPA-capable networks. In other words,
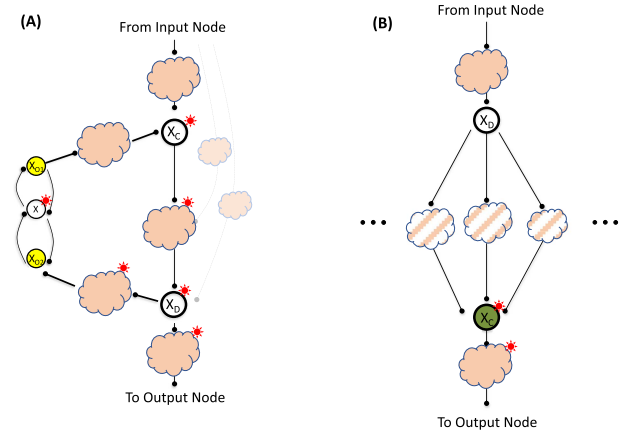
Fig. 3. *Two module classes represent a topological basis for all RPA-capable networks. RPA-capable nodes are indicated by a red asterisk; all other nodes are unable to exhibit RPA. An orange cloud indicates that an arbitrary subnetwork may be embedded at the location indicated without affecting the RPA capacity of the network's output node. A complete description of RPA basis modules is given in [3]. (a) Opposer module, which involves a feedback structure. At least one Opposer node, which satisfies a special constraint known as Opposer kinetics, must be embedded into the feedback portion of an Opposer module. In the representation depicted here, two Opposer nodes (indicated in yellow) collaborate to form a special architecture known as a two-node opposing set. (b) Balancer module is characterized by a collection of parallel pathways, with a diverter node (X_D) at the source point of the parallel pathways, and a connector node (X_C) at the downstream convergence point. All nodes between X_D and X_C, known as balancer nodes (indicated by orange shading), must satisfy a special constraint known as balancer kinetics. The connector node (indicated in green) must also satisfy a special constraint known as connector kinetics. (Figure adapted from [3, Fig. 1].)*

RPA-capable networks are necessarily *modular* and must be decomposable into these well-defined classes of network substructures—an intriguing feature of immense practical utility in the context of synthetic network design. In other words, modularity is not just a convenient conceptual tool, or design strategy, or analytical simplification but a strict *necessity* if a network is required to exhibit RPA.

Diagrammatic methods have recently been developed to assist in the analysis of RPA capacity for specific network architectures [39]. For the network structure given in Fig. 4, for instance, it can easily be shown that the RPA equation for this particular network contains just three terms and admits two (and only two) decompositions into RPA basis modules: a single Opposer module [see Fig. 4(a)], corresponding to an RPA equation whose terms are all identically zero at steady state and, thereby, form a single independently adapting subset $\{\{0,0,0\}\}$, or a Balancer module in parallel with a (smaller) Opposer module [see Fig. 4(b)], corresponding to an RPA equation whose terms partition into *two* independently adapting subsets $\{\{0\}, \{-\alpha, +\alpha\}\}$. We refer interested readers to [3] and [39] for further details on the general principles of partitioning the RPA equation into independently

adapting subsets. The rules of interconnectivity for RPA basis modules to create larger, multimodular networks, and the concept of series and parallel connections of basis modules are also known in complete generality and are explained, in full, in [3] and [39].

As we close this section, we note that the topological view of the RPA problem brings into sharp relief the concept that biochemical signaling networks do not simply transmit biochemical signals but are required to perform precise and coordinated *computations* on those signals in order to implement the robust function. Although the traditional view of integral control has been to consider a (generally nonlinear) change of coordinates that identifies a single "internal model," which computes a single integral, the topological RPA problem suggests that the internal model may be viewed in a more distributed way for complex biological networks. Indeed, topological analysis of robust network designs highlights the presence of a *constellation* of integral control problems, which are distributed in very specific topological arrangements across a potentially large and intricate signaling network. Fig. 5 illustrates this key idea via a schematic representation of integral distributions in two minimal RPA modules—a three-node
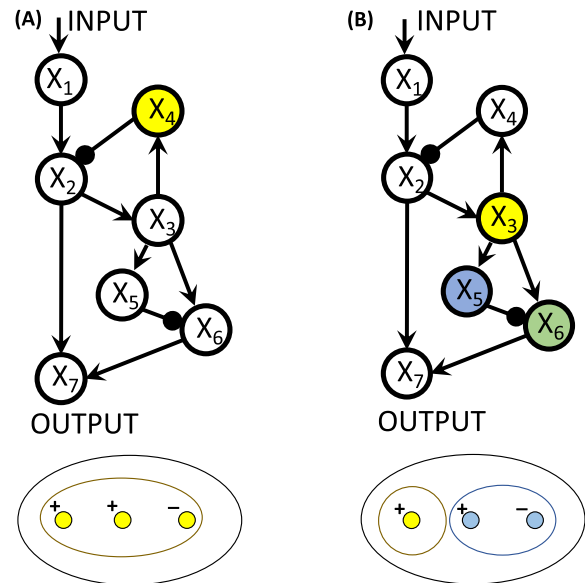
Fig. 4. *Simple network architecture that supports two different decompositions into RPA basis modules. Opposer nodes are indicated in yellow; balancer nodes are indicated in blue; and connector nodes are indicated in green. It can readily be shown via diagrammatic methods [39] that the RPA equation for this network contains three terms, of which two are positive and one is negative. These terms admit two different partitions into "independently adapting subsets," as indicated schematically beneath each network. (a) Opposer module emerges as the three terms of the RPA equation are assigned to a single S-set (yellow). (b) Opposer module in series with a Balancer module emerges as the negative term and one of the positive terms are assigned to an M-set (blue), while the remaining term is assigned to an S-set (yellow).*

(A)

(B)

Fig. 5. *RPA network topologies compute a collection of invariants that, together, confer the RPA property on the output node. (a) Minimal Balancer m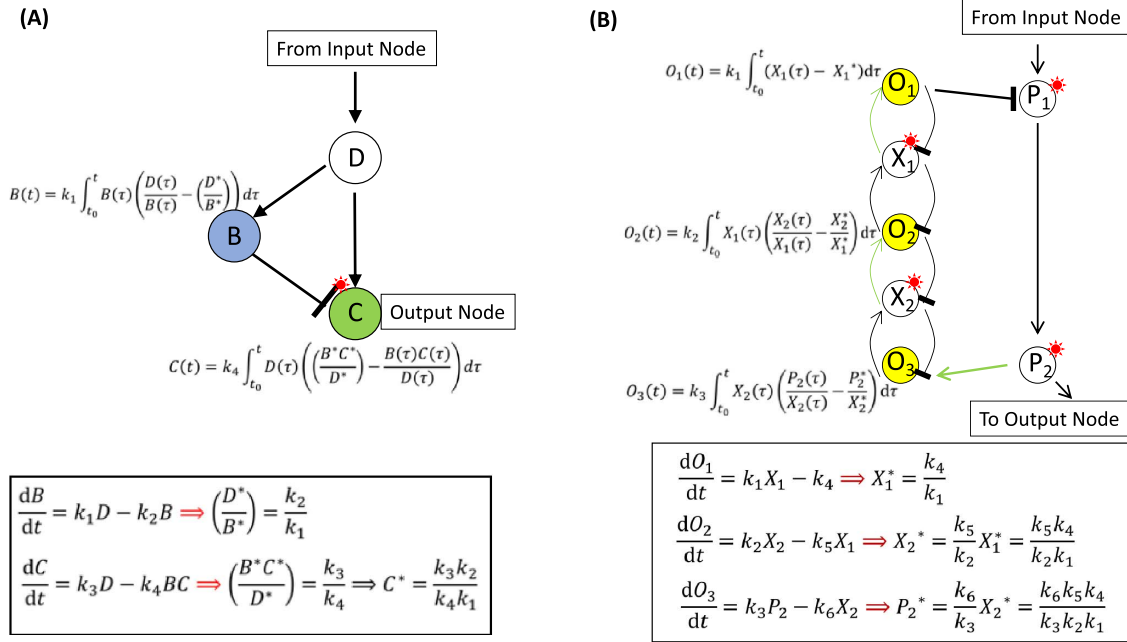odule, in which balancer kinetics at node B compute a balancer invariant, which is then transferred to node C, which then computes the setpoint. (b) Three-node opposing set. Each of the three independent Opposer nodes (yellow) calculates its own Opposer invariant, which is transferred from O_1 to O_2 to O_3, thereby conferring RPA on P_2.*

Balancer module [see Fig. 5(a)] and an Opposer module containing a topological feature known as a three-node opposing set [see Fig. 5(b)]. Each node that is subjected to a constraint on its reaction kinetics—a balancer node, an Opposer node, or a connector node—thereby, computes an RPA-promoting invariant in order to confer the RPA property on the output node of the module and, ultimately, the entire network.

IV. BAYESIAN PERSPECTIVES ON DESIGN

Bayesian inference procedures [40] arguably provide a very natural framework for expressing and reporting uncertainty. Because of this and its ability to accommodate background information and knowledge, it has become increasingly popular in systems biology [41]. One of the main advantages is that it provides a powerful and internally consistent framework for model selection—that is, choosing from a set of candidate models, those that have a high probability of producing the observed data [42]. This can also be used in the context of design for synthetic biology.

A. Bayesian Inference for Design

Bayesian design approaches take a very different, but we feel, complementary approach. Here, we consider sets of models $\mathcal{F} = \{f_1, f_2, \ldots, f_N\}$ [see Fig. 6(a) and (b)] and try to identify which of these models is best capable of meeting the design objectives, T. Here, we adopt an inferential perspective, but, instead of calibrating a model against observed data, we calibrate the models against the data that we would like to see, that is, the design objective.

For each model $f_i \in \mathcal{F}$, the aim is to obtain an estimate for the probability that f_i is the best model capable of producing the target behavior, given T and the set of candidate models, $\mathcal{F}$, $\Pr(f_i|T)$. This *posterior model probability* is given by

$$\Pr(f_i|T) = \frac{\Pr(T|f_i)\pi(f_i)}{\Pr(T)} = \frac{\Pr(T|f_i)\pi(f_i)}{\sum_i \Pr(T|f_i)\pi(f_i)} \quad (6)$$

where $\pi(f_i)$ is the *prior probability* of model f_i, $\Pr(f_i)$, denotes the *likelihood*, and $\Pr(T) = \sum_i \Pr(T|f_i)\pi(f_i)$, the *evidence*, which can be interpreted as a normalizing factor (akin to the partition function in statistical physics).

The terms in (6) refer to the models. We have subsumed the parameter dependence implicitly into these terms, but we need to consider them, of course, as a model f_i will typically only exhibit the target behavior, T, for certain regions of the parameter spaces, Ω_{θ_i}. We let the i subscript refer to the model, f_i, and we allow different parameter spaces for the different models. The *marginal likelihood*, $\Pr(T|f_i)$, is given by

$$\Pr(T|f_i) = \int_{\Omega_i} d\theta' \Pr(T|f_i, \theta') \pi_i(\theta') \quad (7)$$
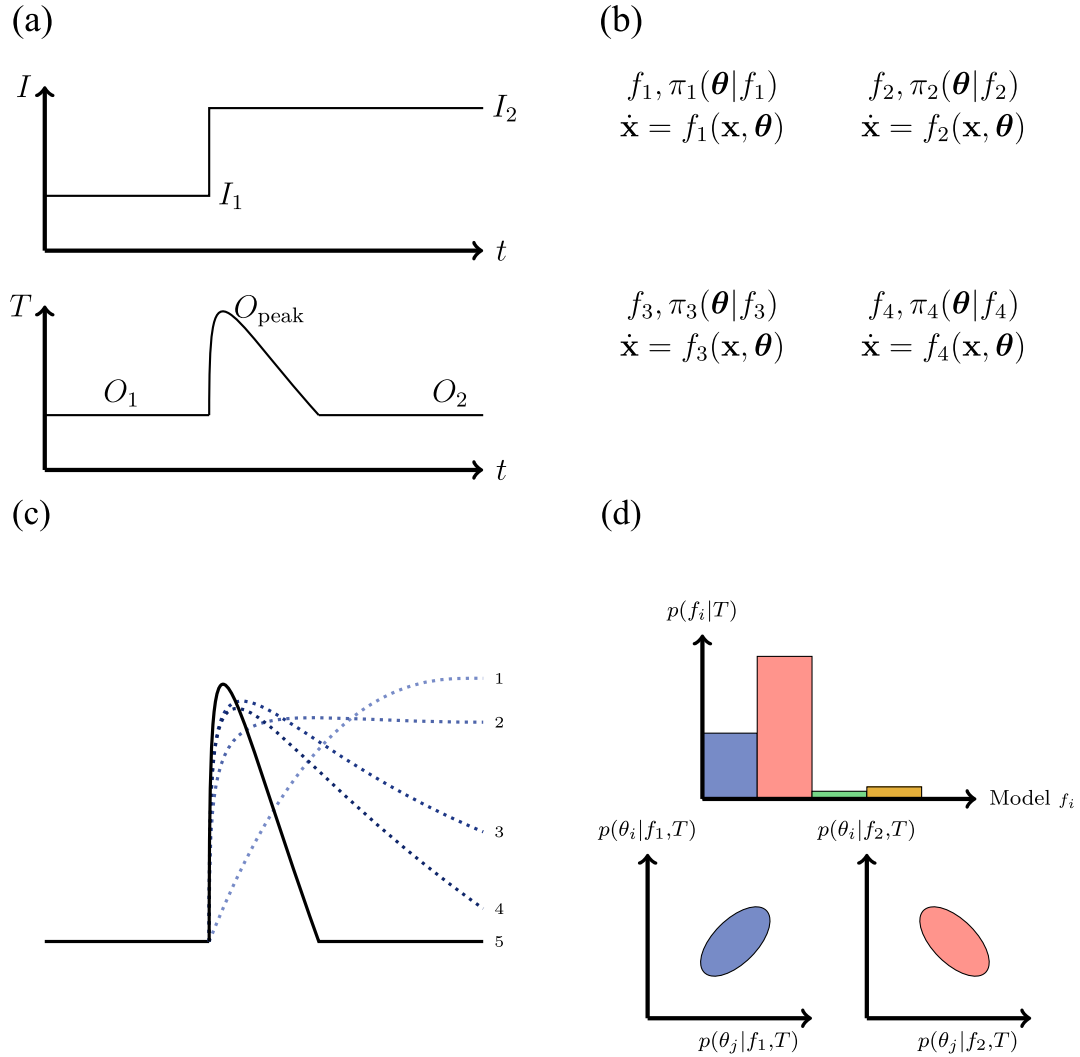
(a)

(b)

$$f_1, \pi_1(\boldsymbol{\theta}|f_1) \qquad f_2, \pi_2(\boldsymbol{\theta}|f_2)$$
$$\dot{\mathbf{x}} = f_1(\mathbf{x}, \boldsymbol{\theta}) \qquad \dot{\mathbf{x}} = f_2(\mathbf{x}, \boldsymbol{\theta})$$

$$f_3, \pi_3(\boldsymbol{\theta}|f_3) \qquad f_4, \pi_4(\boldsymbol{\theta}|f_4)$$
$$\dot{\mathbf{x}} = f_3(\mathbf{x}, \boldsymbol{\theta}) \qquad \dot{\mathbf{x}} = f_4(\mathbf{x}, \boldsymbol{\theta})$$

(c)

(d)

Fig. 6. *Illustration of the Bayesian design framework for synthetic biology. (a) Input and desired system output, T for a system exhibiting adaptation (design objectives). (b) Four alternative models, $f_i(x, \theta)$, are considered, each with a parameter prior, $\pi_i(\theta)$ (model definitions). (c) ABC-SMC approach begins by considering responses that increasingly reflect the target objective, T (system evolution). (d) Outputs of the ABC-SMC procedure are model posterior probabilities and for each model parameter posteriors (posterior distribution). In the example, only the first two models, f_1 and f_2, have sufficiently high posterior model probabilities and need to be considered further.*

where $\Pr(T|f_i, \theta)$ is the likelihood of the parameter and $\pi_i(\theta)$ denotes the prior of the parameters for model i (accounting for the fact that parameter vectors can differ between models).

The rationale for the Bayesian design is that it allows us to estimate the probability that a given model, f_i, will be able to meet the design objectives, T. This is dependent not only on T but also on the set of alternative models considered. The model posterior probability, $\Pr(f_i|T)$, can, thus, be used to choose from the available alternative designs [see Fig. 6(c) and (d)].

This approach has an intrinsic robustness criterion embedded at its heart: in (7), the integral is taken over the parameter space, which means that the model, f_i, with the maximum *a posteriori* probability will generally show good performance against the design objectives over a range of parameters. With this information, it is possible to assess the likely robustness—a concept that we will revisit in the following—of a design and compare competing designs in terms of expected performance and robustness.

B. Approximate Bayesian Computation for Design

This likelihood is also difficult to evaluate for the most important problems of scientific and technological interest. As a result, a range of approximations has been developed, which forgo the explicit evaluation of the likelihood in favor of using simulations. Of these, approximate Bayesian computation (ABC) [43], [44] is perhaps the best-known alternative to evaluating the full posterior via the likelihood function. It has been used in the context of design with some success [4], and we outline it, in the following, before considering applications and generalizations.

In brief, we sample a parameter value, θ', from the prior, $\pi_i(\theta)$, and we then simulate the model for that parameter value to obtain simulated data $\mathcal{D} = \{d_1, d_2, \ldots, d_n\}$. This is then compared to the real data—or in the present context to the design objectives, perhaps via surrogate data $\mathcal{D}_T$—via a distance function, $\Delta(X, Y)$, such that θ' is regarded as a draw from the *approximate posterior*, if

$$\Delta(\mathcal{D}, T) \leq \epsilon \tag{8}$$

where ϵ has to be sufficiently small. The full posterior for the parameter of a given model is, thus, replaced by

$$\Pr(\theta|T) \approx \Pr_{\text{ABC}}(\theta|\Delta(\mathcal{D}, T) \leq \epsilon) \tag{9}$$

and analogously for the model posterior. (Here, we can think of sampling models and then model parameters simultaneously.) Note that, here, we do not, as is frequently the case [45], take a summary statistic of real and simulated data but compare the simulation output directly to the data (or the design objectives).

In ABC inference, the choice of the distance function is of pivotal importance and can make profound differences in the quality and, indeed, the validity of the inference. Classical rejection-based ABC (where only a single threshold is chosen) tends to be computationally inefficient and wasteful. Instead, most inference approaches in the literature adopt a sequential Monte Carlo (SMC) approach, where a sequence of thresholds, $\epsilon_1 \geq \epsilon_2 \geq \cdots \geq \epsilon_n$, is used (either predefined or constructed "online" over the course of the inference procedure—thus, either n or ϵ_n can be a random variable). The choice of this threshold schedule and other aspects of the ABC-SMC approach are still requiring a lot of attention.

In the context of design, the distance function also needs to do a lot of heavy lifting, both in terms of performance and, more fundamentally, for the conceptual setup of the ABC design approach.

In the simplest case, we could take an indicator function

$$\Delta(\mathcal{D}, T) = \begin{cases} 1, & \text{iff } \mathcal{D} \text{ has property } T \\ 0, & \text{otherwise.} \end{cases} \tag{10}$$

However, if we want to put ABC design to work in the context of the vastly more efficient SMC samplers, we have to rethink such indicators. This may not always be possible, of course. The apparent simplicity of ABC approaches masks that there is a great deal of care required to make sure that model selection can be performed in this framework (discussed, e.g., in [45]–[47]).

C. ABC Design for Perfect Adaptive Behavior

We can illustrate the ABC for design approach in the context of adaptive behavior. In addition to the algebraic

methods [3], which can be used to distill the design characteristics, large-scale computational searches [38] have been used. These identified two models are able to generate adaptive behavior. Either the algebraic approach or brute-force computational surveys can be used to produce sets of model designs, $\mathcal{F} = \{f_1, \ldots, f_n\}$, which can then be used as the starting point for model selection for design.

The output characteristics in [4] were chosen to be the adaptation efficiency and sensitivity [38], given by

$$E = \frac{(O_2 - O_1)/O_1}{(I_2 - I_1)/I_1} \tag{11}$$

and

$$S = \frac{(O_{\text{peak}} - O_1)/O_1}{(I_2 - I_1)/I_1} \tag{12}$$

where I_1 and I_2 are the baseline and activated input levels; O_2 and O_1 are the steady-state outputs before and after the signal is applied; and O_{peak} is the maximal transient output level of the response. The ABC inference is based on the two-component distance metric, $\rho = \{E, S^{-1}\}$ (see [4] for details).

In Fig. 7, we show two models [see Fig. 7(a) and (b)] that have been argued to exhibit adaptive behavior in previous studies [4], [38] and their respective posterior probabilities (assuming uniform model priors [48]) to produce adaptive behavior if cooperativity is allowed [see Fig. 7(c)] or not allowed [see Fig. 7(d)]. The two posterior probabilities are sufficiently different in the cooperative scenario [see Fig. 7(c)] to warrant choice of model 1 $\Pr(f_1|T) \approx 0.96$, whereas, for noncooperative dynamics, the choice is less clear with $\Pr(f_2|T) \approx 0.64$. This is not decisive enough, and further analysis, and consideration of further design candidates, may become necessary.

V. ROBUSTNESS IN SYNTHETIC BIOLOGY DESIGN

In the following, we develop two notions of robustness in synthetic biology design.

Biological robustness refers to the ability of a system to perform its function in the presence of perturbations.

Model robustness is related to mismatches between the mathematical model and the real biological systems that it sets out to describe.

Both are relevant in the context of synthetic and engineering biology and design approaches and are discussed in the following.

A. Robustness of Biological Systems and Models

Robustness as a property of both biological systems and their representations as mathematical models have been the subject of much discussion [11], [12], [49]–[58]. For biological systems, robustness is generally defined as the property that ensures that the system maintains its
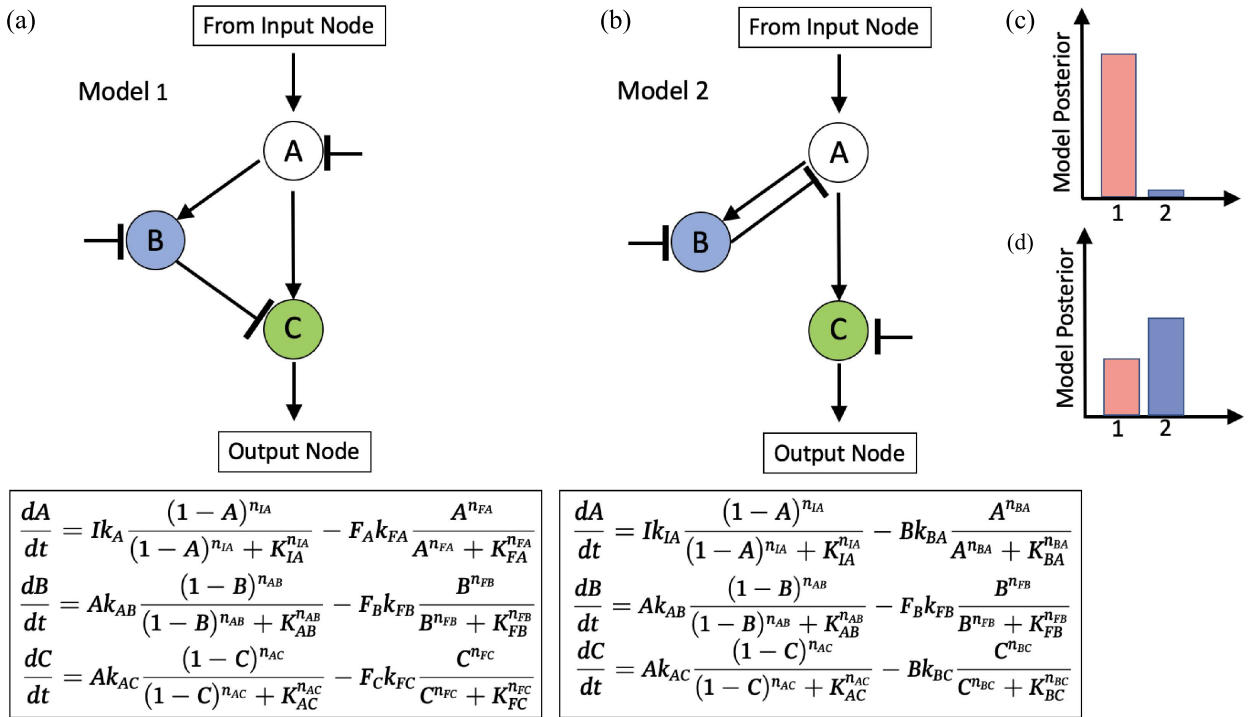
(a)

Model 1

From Input Node

A

B

C

Output Node

(b)

Model 2

From Input Node

A

B

C

Output Node

(c)

Model Posterior

1   2

(d)

Model Posterior

1   2

$$\frac{dA}{dt} = Ik_A \frac{(1-A)^{n_{IA}}}{(1-A)^{n_{IA}} + K_{IA}^{n_{IA}}} - F_A k_{FA} \frac{A^{n_{FA}}}{A^{n_{FA}} + K_{FA}^{n_{FA}}}$$

$$\frac{dB}{dt} = Ak_{AB} \frac{(1-B)^{n_{AB}}}{(1-B)^{n_{AB}} + K_{AB}^{n_{AB}}} - F_B k_{FB} \frac{B^{n_{FB}}}{B^{n_{FB}} + K_{FB}^{n_{FB}}}$$

$$\frac{dC}{dt} = Ak_{AC} \frac{(1-C)^{n_{AC}}}{(1-C)^{n_{AC}} + K_{AC}^{n_{AC}}} - F_C k_{FC} \frac{C^{n_{FC}}}{C^{n_{FC}} + K_{FC}^{n_{FC}}}$$

$$\frac{dA}{dt} = Ik_{IA} \frac{(1-A)^{n_{IA}}}{(1-A)^{n_{IA}} + K_{IA}^{n_{IA}}} - Bk_{BA} \frac{A^{n_{BA}}}{A^{n_{BA}} + K_{BA}^{n_{BA}}}$$

$$\frac{dB}{dt} = Ak_{AB} \frac{(1-B)^{n_{AB}}}{(1-B)^{n_{AB}} + K_{AB}^{n_{AB}}} - F_B k_{FB} \frac{B^{n_{FB}}}{B^{n_{FB}} + K_{FB}^{n_{FB}}}$$

$$\frac{dC}{dt} = Ak_{AC} \frac{(1-C)^{n_{AC}}}{(1-C)^{n_{AC}} + K_{AC}^{n_{AC}}} - Bk_{BC} \frac{C^{n_{BC}}}{C^{n_{BC}} + K_{BC}^{n_{BC}}}$$

Fig. 7. *Bayesian design of an adaptive system. (a) and (b) Two models capable of exhibiting adaptive behavior, including their representations as ordinary differential equations. The horizontal inhibitory edges (⊣) indicate autorepression on the nodes. (c) Posterior model probabilities for achieving adaptive behavior by models 1 and 2 if there is cooperativity (Hill coefficient, n > 1). (d) Posterior model probabilities for achieving adaptive behavior by models 1 and 2 when there is no cooperativity. The posterior probabilities in (c) and (d) sum up to one. Note that these systems differ from the ones found using the algebraic approach; they are designed to show adaptive behavior and for different, Michaelis–Menten kinetics, which the Bayesian approach is able to deal with.*

functions in the presence of perturbations from internal and external sources [33], [34], [59]. In the context of design, we take a pragmatic point of view of robustness. We are interested in identifying designs that are robust to: 1) inherent stochasticity; 2) environmental and physiological signals and fluctuations; and 3) uncertainty in our knowledge of the precise system architecture and kinetic reaction rates.

For mathematical models, many working definitions of robustness have been proposed, which are based on single- and multi-parameter sensitivity analyses [51], [58]. They include, for example, invariance under external fluctuations affecting the initial conditions [49] or structural stability [60]. There is, however, no clear and consistent definition that allows for universal and consistent quantification of robustness, which is necessary if the robustness characteristics of different systems are to be compared. Perhaps, the primary importance of quantifying model robustness is in model validation, as a valid model must exhibit equivalent robustness to the corresponding biological system that it represents. In fact, a lack of robustness may indicate a weak model that provides a poor representation of the biological system [52], so robustness may be used as a selection tool for promising models that correctly represent the functions of the biological system.

The determination of weak models through robustness quantification is also useful for providing insight into the robustness of the biological system in a synthetic biology design context.

Importantly, robustness relates to the maintenance of the functions of the system and not necessarily the states of the system, so it is more general than stability and homeostasis and may involve transitions to new steady states or even to instability if required to maintain function [34]. For example, the system function, where a morphogen concentration gradient determines the position in a biological system, appears to be robust despite temporal perturbations of the gradient to different steady states [61]. Despite their complexity, evolved biological systems apparently exhibit robustness as a ubiquitous and fundamental property [33], [62], [63]. Synthetic cellular systems have to work in the context of such naturally evolved contexts.

B. Mathematical Definitions of Robustness

Unfortunately, and as noted some time ago [34], there is still no adequate mathematical theory of robustness. There are a number of analytical issues to be addressed, in particular, the difficulty involved with high-dimensional parameter spaces. For example, control-theoretic techniques have

been employed to study the robustness of biochemical network models [51]; however, these approaches are limited to low-dimensional parameter spaces [33].

A robust model must maintain the represented biological function with respect to some degree of perturbation of the inputs. In other words, we require small changes in the inputs—model parameters, initial conditions, boundary conditions, and external conditions—to produce sufficiently small changes in the output so that the integrity of the biological function remains. Note that this is different from requiring invariance of the output: the system output can vary as long as the biological function is maintained. We, therefore, only require invariance of biological function: the design objectives, T, still have to be met.

We require a way of assessing the robustness and focus here, for didactic reasons, on robustness to parametric uncertainty or variation. In order to compare distances between input values and output values (including any design objective, T), we require the sets of input and output values to form metric spaces.

We, furthermore, require a functional relationship between the input and output metric spaces, which would generally be a continuous function in order to ensure that the output values change smoothly with changes in the inputs. It is also possible that an appropriate function has discontinuities, such as a step discontinuity, so maybe only piecewise continuous. In this case, we just need to ensure that the function output still corresponds to the invariance of biological function by accounting for the behavior near the discontinuities. This makes it possible to accommodate qualitative design criteria. We would also like to be able to specify the robustness of a system by a single parameter; we, therefore, need a methodology that can be applied to the whole parameter space to give an overall measure of robustness.

From a global perspective, we require that close values of the input produce output values that are consistently close. Note that local continuity of a function, or global uniform continuity, is not sufficient for defining robustness as they allow for arbitrarily large derivatives of the function and, thus, a lack of robustness. The Lipschitz function [64, p. 154, Sec. 9.4] immediately lends itself to a rigorous definition of robustness for mathematical models. However, inexplicably, it has not been employed within the mathematics literature, to the best of our knowledge; it has, however, appeared in other fields, such as software engineering [65] and cyber–physical systems [66].

We denote the input and outputs by X and Y, respectively; we define suitable metrics, d_X and d_Y on the state spaces, for X and Y; and we can, therefore, consider (X, d_X) and (Y, d_Y) to be metric spaces. A function $f: X \rightarrow Y$ is a *Lipschitz function* on X with a *Lipschitz constant* $K \geq 0$ if f satisfies

$$d_Y(f(x_1), f(x_2)) \leq K\, d_X(x_1, x_2) \qquad (13)$$

for all $x_1, x_2 \in X$. We also say that f is *Lipschitz continuous*. The smallest such constant K is *the* Lipschitz constant. The Lipschitz constant provides an upper bound on the change of the function output given a change in the input. Ideally, we would like an analytical value for the Lipschitz constant; however, there may be Lipschitz functions for which this is not easily obtained, so numerical estimates may be helpful. Numerical methods for estimating Lipschitz constants exist [67]; however, there is no guarantee that the estimated constant is valid globally. In control theory, *transfer functions* are employed to provide a functional relationship between the inputs and outputs of a system, often within the frequency domain, by applying the Laplace transform to both the inputs and outputs, thereby allowing direct comparison of functions [68]. While the overall approach using transfer functions overlaps with our requirements, it is perhaps too specific and limited for characterizing the robustness of complex biological systems.

C. Parametric Robustness as a Design Criterion in Synthetic Biology

The algebraic approaches outlined above allow us to detect the design principles underlying a given desired behavior or objective, T. This allows us to design models that meet the objectives of construction.

The Bayesian approach contains an assessment of robustness via the posterior [69], which integrates over the parameters [see (7)]. Rather than the design objective, T, we may use the posterior parameter density, $\Pr(\theta|T)$, or the ABC posterior, (7), as the output and then compare $d_\theta(\theta, \theta')$ and $d_P(\Pr(\theta|T), \Pr(\theta'|T))$ using the Lipschitz criterion (13). We can focus this analysis on, for example, the 90% credible interval [40].

If our design, which is the mathematical model, is robust against variation in parameters, if it is structurally stable, and if it is even robust against changes to the model structure [70], then we can view this as reassuring that the design will also be robust to the discrepancies between the mathematical model and biological, reality [6], [41].

VI. CONCLUSION AND DISCUSSION

The models in Section III and Fig. 5 are structurally very similar to the models considered in Section V and Fig. 7 but differ in practical terms as we first use mass action kinetics and then Michaelis–Menten kinetics. Algebraic methods are more powerful wherever they are applicable, as they result in definitive criteria that must be satisfied for a given behavior to emerge. From this, we can often gain profound insights into the dynamics of cellular systems [3], [5].

For many classes of design objectives, other methods need to be applied, and the Bayesian design framework is one method that is widely applicable for different models. Basing the design on Bayesian model selection [4], [48], [71] has two intrinsic advantages: 1) we are able to balance the performance and the robustness of different

designs and 3) for ABC methods, we can, at least in principle, apply Bayesian design to arbitrary models as long as we are able to simulate them. The first point is potentially an advantage over large-scale or exhaustive analyses of models [6], [38], [70], [72], where robustness can be assessed *post hoc*. While global analyses and Bayesian methods are both computationally expensive, the former is better suited for the comprehensive analysis of a large number of systems: the objective there is to scan model and parameter spaces to find if there are any parameters for which the design objectives, T, are realized. In the Bayesian approach, by contrast, we also obtain estimates of the parameters, θ_i, for each model $f_i \in \mathcal{F}$.

Thus, we can view the algebraic and Bayesian frameworks as complementary, and given the complexities of rationally designing synthetic circuits [1], [73], a flexible and diverse toolbox of alternative design methods is clearly beneficial. However, algebraic methods can also support and work in conjunction with Bayesian methods (or exhaustive scans of large sets of models): we can determine initial sets of models, $\mathcal{F} = \{f_1, \ldots, f_n\}$, to fulfill the design criteria based on algebraic rules or design principles (even if under perhaps restricted kinetic regimes, such as mass-action kinetics); a natural choice for the prior model probabilities would then be $\pi(f_i) = 1/n$ for all $i = 1, \ldots, n$. These models can then be subjected to the Bayesian model selection/design process to choose those models that have the highest probability of delivering the design objective, T (and do so robustly).

The challenge of design in synthetic biology is considerable. Unlike in physics and many engineering disciplines, we cannot draw on first principles. Reliable mathematical models for cellular systems are still, despite much recent progress, rare, but we need them to apply rational engineering design approaches in synthetic biology. The methods described here—algebraic ways to identify design principles, Bayesian design, and the concept of robustness—can help to identify synthetic circuits capable of delivering a target objective, T. In combination, they allow us to take some of the guesswork and trial and error out of synthetic biology design. ∎

REFERENCES

[1] W. A. Lim, "Designing customized cell signalling circuits," *Nature Rev. Mol. Cell Biol.*, vol. 11, no. 6, pp. 393–403, Jun. 2010.

[2] J. Santos-Moreno, E. Tasiudi, J. Stelling, and Y. Schaerli, "Multistable and dynamic CRISPRi-based synthetic circuits," *Nature Commun.*, vol. 11, no. 1, pp. 2746–2748, Dec. 2020.

[3] R. P. Araujo and L. A. Liotta, "The topological requirements for robust perfect adaptation in networks of any size," *Nature Commun.*, vol. 9, no. 1, p. 1757, Dec. 2018.

[4] C. P. Barnes, D. Silk, X. Sheng, and M. P. H. Stumpf, "Bayesian design of synthetic biological systems," *Proc. Nat. Acad. Sci. USA*, vol. 108, no. 37, pp. 15190–15195, Sep. 2011.

[5] H. A. Harrington, E. Feliu, C. Wiuf, and M. P. H. Stumpf, "Cellular compartments cause multistability and allow cells to process more information," *Biophys. J.*, vol. 104, no. 8, pp. 1824–1831, Apr. 2013.

[6] N. S. Scholes, D. Schnoerr, M. Isalan, and M. P. H. Stumpf, "A comprehensive network atlas reveals that Turing patterns are common but not robust," *Cell Syst.*, vol. 9, no. 3, pp. 243–257, 2019.

[7] D. Muzzey, C. A. Gómez-Uribe, J. T. Mettetal, and A. van Oudenaarden, "A systems-level analysis of perfect adaptation in yeast osmoregulation," *Cell*, vol. 138, no. 1, pp. 160–171, Jul. 2009.

[8] M. H. Khammash, "Perfect adaptation in biology," *Cell Syst.*, vol. 12, no. 6, pp. 509–521, Jun. 2021.

[9] U. B. Kaupp, "Olfactory signalling in vertebrates and insects: Differences and commonalities," *Nature Rev. Neurosci.*, vol. 11, no. 3, pp. 188–200, Mar. 2010.

[10] K.-W. Yau and R. C. Hardie, "Phototransduction motifs and variations," *Cell*, vol. 139, no. 2, pp. 246–264, 2009.

[11] N. Barkai and S. Leibler, "Robustness in simple biochemical networks," *Nature*, vol. 387, no. 6636, pp. 913–917, Jun. 1997.

[12] U. Alon, M. G. Surette, N. Barkai, and S. Leibler, "Robustness in bacterial chemotaxis," *Nature*, vol. 397, no. 6715, pp. 168–171, 1999.

[13] T. Toni, G. Jovanovic, M. Huvet, M. Buck, and M. P. Stumpf, "From qualitative data to quantitative models: Analysis of the phage shock protein stress response in Escherichia coli," *BMC Syst. Biol.*, vol. 5, no. 1, p. 69, 2011.

[14] N. Joly *et al.*, "Managing membrane stress: The phage shock protein (Psp) response, from molecular mechanisms to physiology," *FEMS Microbiol. Rev.*, vol. 34, no. 5, pp. 797–827, Sep. 2010.

[15] T. You *et al.*, "A systems biology analysis of long and short-term memories of osmotic stress adaptation in fungi," *BMC Res. Notes*, vol. 5, no. 1, pp. 1–16, Dec. 2012.

[16] C. Briat, A. Gupta, and M. Khammash, "Antithetic integral feedback ensures robust perfect adaptation in noisy biomolecular networks," *Cell Syst.*, vol. 2, pp. 15–26, Jan. 2016.

[17] F. Mairet, "A biomolecular proportional integral controller based on feedback regulations of protein level and activity," *Roy. Soc. Open Sci.*, vol. 5, no. 2, Feb. 2018, Art. no. 171966.

[18] X. Y. Ni, T. Drengstig, and P. Ruoff, "The control of the controller: Molecular mechanisms for robust perfect adaptation and temperature compensation," *Biophysical J.*, vol. 97, no. 5, pp. 1244–1253, Sep. 2009.

[19] A. Eldar, R. Dorfman, D. Weiss, H. Ashe, B.-Z. Shilo, and N. Barkai, "Robustness of the BMP morphogen gradient in Drosophila embryonic patterning," *Nature*, vol. 419, no. 6904, pp. 304–308, Sep. 2002.

[20] D. Ben-Zvi and N. Barkai, "Scaling of morphogen gradients by an expansion-repression integral feedback control," *Proc. Nat. Acad. Sci. USA*, vol. 107, no. 15, pp. 6924–6929, Apr. 2010.

[21] F. Xiao and J. C. Doyle, "Robust perfect adaptation in biomolecular reaction networks," in *Proc. IEEE Conf. Decis. Control (CDC)*, Dec. 2018, pp. 4345–4352.

[22] D. Cappelletti, A. Gupta, and M. Khammash, "A hidden integral structure endows absolute concentration robust systems with resilience to dynamical concentration disturbances," *J. Roy. Soc. Interface*, vol. 17, no. 171, Oct. 2020, Art. no. 20200437.

[23] C. I. Byrnes and A. Isidori, "Limit sets, zero dynamics, and internal models in the problem of nonlinear output regulation," *IEEE Trans. Autom. Control*, vol. 48, no. 10, pp. 1712–1723, Oct. 2003.

[24] B. A. Francis and W. M. Wonham, "The internal model principle for linear multivariable regulators," *Appl. Math. Optim.*, vol. 2, no. 2, pp. 170–194, Jun. 1975.

[25] B. A. Francis and W. M. Wonham, "The internal model principle of control theory," *Automatica*, vol. 12, no. 5, pp. 457–465, Sep. 1976.

[26] E. D. Sontag, "Adaptation and regulation with signal detection implies internal model," *Syst. Control Lett.*, vol. 50, no. 2, pp. 119–126, Oct. 2003.

[27] R. L. Karp, M. P. Millán, T. Dasgupta, A. Dickenstein, and J. Gunawardena, "Complex-linear invariants of biochemical networks," *J. Theor. Biol.*, vol. 311, pp. 130–138, Oct. 2012.

[28] J. E. Ferrell, "Perfect and near-perfect adaptation in cell signaling," *Cell Syst.*, vol. 2, no. 2, pp. 62–67, Feb. 2016.

[29] O. Shoval, U. Alon, and E. Sontag, "Symmetry invariance for adapting biological systems," *SIAM J. Appl. Dyn. Syst.*, vol. 10, no. 3, pp. 857–886, Jan. 2011.

[30] O. Hoeller, D. Gong, and O. D. Weiner, "How to understand and outwit adaptation," *Develop. Cell*, vol. 28, no. 6, pp. 607–616, Mar. 2014.

[31] T.-M. Yi, Y. Huang, M. I. Simon, and J. Doyle, "Robust perfect adaptation in bacterial chemotaxis through integral feedback control," *Proc. Nat. Acad. Sci. USA*, vol. 97, no. 9, pp. 4649–4653. 2000.

[32] Z. F. Tang and D. R. Mcmillen, "Design principles for the analysis and construction of robustly homeostatic biological networks," *J. Theor. Biol.*, vol. 408, pp. 274–289, Nov. 2016.

[33] H. Kitano, "Biological robustness," *Nature Rev. Genet.*, vol. 5, pp. 826–837, Nov. 2004.

[34] H. Kitano, "Towards a theory of biological robustness," *Mol. Syst. Biol.*, vol. 3, no. 1, p. 137, Jan. 2007.

[35] R. P. Araujo, L. A. Liotta, and E. F. Petricoin, "Proteins, drug targets and the mechanisms they control: The simple truth about complex networks," *Nature Rev. Drug Discovery*, vol. 6, no. 11, pp. 871–880, Nov. 2007.

[36] R. Araujo and L. Liotta, "A control theoretic paradigm for cell signaling networks: A simple complexity for a sensitive robustness," *Current Opinion Chem. Biol.*, vol. 10, no. 1, pp. 81–87, Feb. 2006.

[37] J. J. Tyson, K. C. Chen, and B. Novak, "Sniffers, buzzers, toggles and blinkers: Dynamics of regulatory and signaling pathways in the cell," *Current Opinion Cell Biol.*, vol. 15, no. 2, pp. 221–231, Apr. 2003.

[38] W. Ma, A. Trusina, H. El-Samad, W. A. Lim, and C. Tang, "Defining network topologies that can achieve biochemical adaptation," *Cell*, vol. 138, no. 4, pp. 760–773, Aug. 2009.

[39] R. P. Araujo and L. A. Liotta, "Design principles underlying robust perfect adaptation of complex

biochemical networks," *Methods Mol. Biol.*, to be published.

[40] C. P. Robert, *The Bayesian Choice*. New York, NY, USA: Springer, 2007.

[41] P. D. W. Kirk, A. C. Babtie, and M. P. H. Stumpf, "Systems biology (un)certainties," *Science*, vol. 350, no. 6259, pp. 386–388, Oct. 2015.

[42] P. Kirk, T. Thorne, and M. P. Stumpf, "Model selection in systems and synthetic biology," *Current Opinion Biotechnol.*, vol. 24, no. 4, pp. 767–774, Aug. 2013.

[43] T. Toni, D. Welch, N. Strelkowa, A. Ipsen, and M. P. H. Stumpf, "Approximate Bayesian computation scheme for parameter inference and model selection in dynamical systems," *J. Roy. Soc. Interface*, vol. 6, no. 31, pp. 187–202, 2009.

[44] R. D. Wilkinson, "Approximate Bayesian computation (ABC) gives exact results under the assumption of model error," *Stat. Appl. Genet. Mol. Biol.*, vol. 12, no. 2, pp. 129–141, Jan. 2013.

[45] C. P. Robert, J.-M. Cornuet, J.-M. Marin, and N. S. Pillai, "Lack of confidence in approximate Bayesian computation model choice," *Proc. Nat. Acad. Sci. USA*, vol. 108, no. 37, pp. 15112–15117, Sep. 2011.

[46] C. P. Barnes, S. Filippi, M. P. H. Stumpf, and T. Thorne, "Considerate approaches to constructing summary statistics for ABC model selection," *Statist. Comput.*, vol. 22, no. 6, pp. 1181–1197, Nov. 2012.

[47] D. Prangle, P. Fearnhead, M. P. Cox, P. J. Biggs, and N. P. French, "Semi-automatic selection of summary statistics for ABC model choice," *Stat. Appl. Genet. Mol. Biol.*, vol. 13, no. 1, pp. 67–82, Jan. 2014.

[48] T. Toni and M. P. H. Stumpf, "Simulation-based model selection for dynamical systems in systems and population biology," *Bioinformatics*, vol. 26, no. 1, pp. 104–110, Jan. 2010.

[49] E. Crampin, "Reaction and diffusion on growing domains: Scenarios for robust pattern formation," *Bull. Math. Biol.*, vol. 61, no. 6, pp. 1093–1120, Oct. 1999.

[50] J. W. Little, "Robustness of a gene regulatory circuit," *EMBO J.*, vol. 18, no. 15, pp. 4299–4307, Aug. 1999.

[51] L. Ma and P. A. Iglesias, "Quantifying robustness of biochemical network models," *BMC Bioinf.*, vol. 3, p. 38, Dec. 2002.

[52] M. Morohashi, A. E. Winn, M. T. Borisuk, H. Bolouri, J. Doyle, and H. Kitano, "Robustness as a measure of plausibility in models of biochemical networks," *J. Theor. Biol.*, vol. 216, no. 1, pp. 19–30, May 2002.

[53] H. Kitano, "Cancer as a robust system: Implications for anticancer therapy," *Nature Rev. Cancer*, vol. 4, no. 3, pp. 227–235, Mar. 2004.

[54] T. Tian, "Robustness of mathematical models for biological systems," in *Proc. 11th Comput. Techn. Appl. Conf. (CTAC)*, J. Crawford and A. J. Roberts, Eds., vol. 45, 2004, pp. C565–C577.

[55] D. A. Rand, "Mapping global sensitivity of cellular network dynamics: Sensitivity heat maps and a global summation law," *J. Roy. Soc. Interface*, vol. 5, no. 1, pp. S59–S69, Aug. 2008.

[56] J. Raerinne, "Robustness and sensitivity of biological models," *Phil. Stud.*, vol. 166, no. 2, pp. 285–303, Nov. 2013.

[57] L. Li, M. Han, and Y. Liu, "Existence and uniqueness of traveling wave front of a nonlinear singularly perturbed system of reaction–diffusion equations with a heaviside step function," *J. Math. Anal. Appl.*, vol. 410, no. 1, pp. 202–212, Feb. 2014.

[58] Y. Matsuoka, N. Jahan, and H. Kurata, "S-system-based analysis of the robust properties common to many biochemical network models," *Bioprocess Biosyst. Eng.*, vol. 39, no. 5, pp. 735–746, May 2016.

[59] J. Stelling, U. Sauer, Z. Szallasi, F. J. Doyle, and J. Doyle, "Robustness of cellular functions," *Cell*, vol. 118, no. 6, pp. 675–685, Sep. 2004.

[60] F. Blanchini and E. Franco, "Structurally robust biological networks," *BMC Syst. Biol.*, vol. 5, no. 1, p. 74, 2011.

[61] F. Tostevin, P. R. ten Wolde, and M. Howard, "Fundamental limits to position determination by concentration gradients," *PLoS Comput. Biol.*, vol. 3, no. 4, p. e78, Apr. 2007.

[62] A. Wagner, "Robustness and evolvability: A paradox resolved," *Proc. Roy. Soc. B, Biol. Sci.*, vol. 275, no. 1630, pp. 91–100, Jan. 2008.

[63] J. T. Young, T. S. Hatakeyama, and K. Kaneko, "Dynamics robustness of cascading systems," *PLOS Comput. Biol.*, vol. 13, no. 3, Mar. 2017, Art. no. e1005434.

[64] M. Ó Searcóid, *Metric Spaces*. London, U.K.: Springer, 2007.

[65] S. Chaudhuri, S. Gulwani, and R. Lublinerman, "Continuity and robustness of programs," *Commun. ACM*, vol. 55, no. 8, pp. 107–115, Aug. 2012.

[66] R. Majumdar, E. Render, and P. Tabuada, "Robust discrete synthesis against unspecified disturbances," in *Proc. 14th Int. Conf. Hybrid Systems, Comput. Control (HSCC)*, 2011, pp. 211–220.

[67] G. R. Wood and B. P. Zhang, "Estimation of the Lipschitz constant of a function," *J. Global Optim.*, vol. 8, no. 1, pp. 91–103, Jan. 1996.

[68] J. Zabczyk, *Mathematical Control Theory*, 2nd ed. Basel, Switzerland: Birkhäuser, 2020.

[69] . Secrier, T. Toni, M. P. H. Stumpf, and M. P. H. Stumpf, "The ABC of reverse engineering biological signalling systems," *Mol. Biosyst.*, vol. 5, no. 12, pp. 1925–1935, 2009.

[70] A. C. Babtie, P. Kirk, and M. P. H. Stumpf, "Topological sensitivity analysis for systems biology," *Proc. Nat. Acad. Sci. USA*, vol. 111, no. 52, pp. 18507–18512, Dec. 2014.

[71] C. P. Barnes, D. Silk, and M. P. H. Stumpf, "Bayesian design strategies for synthetic biology," *Interface Focus*, vol. 1, no. 6, pp. 895–908, Dec. 2011.

[72] Y. Schaerli, A. Munteanu, M. Gili, J. Cotterell, J. Sharpe, and M. Isalan, "A unified design space of synthetic stripe-forming networks," *Nature Commun.*, vol. 5, no. 1, p. 4905, Dec. 2014.

[73] O. Purcell, B. Jain, J. R. Karr, M. W. Covert, and T. K. Lu, "Towards a whole-cell modeling approach for synthetic biology," *Chaos, Interdiscipl. J. Nonlinear Sci.*, vol. 23, no. 2, Jun. 2013, Art. no. 025112.