Optimization-based synthesis of stochastic biocircuits with statistical specifications

Yuta Sakurai*, Yutaka Hori*†

Abstract

Model-guided design has become a standard approach to engineering biomolecular circuits in current synthetic biology. However, the stochastic nature of biomolecular reactions is often overlooked in the design process. As a result, cell-cell heterogeneity causes unexpected deviation of biocircuit behaviors from model predictions and requires additional iterations of design-build-test cycles. To enhance the design process of stochastic biocircuits, this paper presents a computational framework to systematically specify the level of intrinsic noise using well-defined metrics of statistics and design highly heterogeneous biocircuits based on the specifications. Specifically, we use descriptive statistics of population distributions as an intuitive specification language of stochastic biocircuits and develop an optimization based computational tool that explores parameter configurations satisfying design requirements. Sensitivity analysis methods are also developed to ensure the robustness of a biocircuit design. These design tools are formulated using convex optimization programs to enable efficient and rigorous quantification of the statistics without approximation, and thus, they are amenable to the synthesis of stochastic biocircuits that require high reliability. We demonstrate these features by designing a stochastic negative feedback biocircuit that satisfies multiple statistical constraints. In particular, we use a rigorously quantified parameter map of feasible design space to perform in-depth study of noise propagation and regulation in negative feedback pathways.

The authors contributed equally to this work. *Department of Applied Physics and Physico-Informatics, Keio University, Kanagawa, 223-8522 Japan. †To whom correspondence should be addressed. yhori@appi.keio.ac.jp.

Introduction

The last two decades of intense efforts in synthetic biology have greatly expanded our ability to build synthetic biomolecular circuits by adopting many concepts and techniques from engineering disciplines. Model-guided design is one of such examples that have been routinely used to create safe and robust control systems in traditional engineering [1] and have been adopted in the design process of biocircuits [2]. To date, many biocircuit modules were engineered with the help of model-based simulations, including logic gates [3, 4, 5], oscillators [6, 7, 8, 9] and genetic memory [10, 11] to name a few. A current challenge of biocircuit engineering is to integrate these circuit modules and build systems for complex operations in the real-world environments, which requires more stringent reliability of each circuit module. As is the case with any engineering systems, a first key step to the robust design of such complex systems is to set appropriate and well-defined performance norms that can specify all the necessary features of systems' behavior. Mathematical and computational tools then facilitate design space exploration to find parameter configurations that achieve pre-specified performance requirements. Compared with this ideal, the current design process of biocircuits is still far immature in that models are used mostly for simulations to gain only qualitative insights rather than for quantitatively guaranteeing the performance of biocircuits by fully benefitting from advanced theory and algorithms. This motivates us to develop computational frameworks that streamline the design process by systematically certifying and optimizing the performance levels of biocircuits.

One of the important features that should be carefully considered in the design process of biocircuits is cellular heterogeneity. In biological cells, the low copy nature of molecules induces randomness of molecular collision events that fire chemical reactions, resulting in the large variation of biocircuit states across cell populations even if the cells are genetically identical and grown in the same condition [12, 13, 14]. In many cases, the signal-to-noise ratio of biocircuits is much lower than that of mechanically and electronically engineered systems. Although noise attenuation has been a rule of thumb in engineering, recent studies demonstrated opposite strategies to take advantage of the highly stochastic nature of biomolecular reactions and design biocircuits that operate collectively at a population level [15, 16, 17, 18]. For example, collections of binary outputs from stochastic biocircuits can form a graded response that enables analog decision making in highly stochastic environments [18, 19]. These examples illustrate that we can design novel mechanisms that are different from those in traditional systems to control biocircuits by actively leveraging the heterogeneous responses of cell populations.

To enhance the design process of stochastic biocircuits, the first key step is to produce specifications that can capture all the necessary design features using simple but well-defined performance metrics. For this purpose, useful criteria would be descriptive statistics such as covariance, correlation, the coefficient of variation (CV) and Fano factor in addition to the population mean values. In current synthetic biology, the majority of studies uses Monte Carlo based stimulations of single cell trajectories to approximately evaluate these statistics of population distributions[20]. To complement the time-consuming nature of the Monte Carlo approach, other computational tools are available to directly quantify population distributions [21, 22] and raw moments [23, 24, 25, 26] without running simulations. However, these methods are not designed to directly compute descriptive statistics of biocircuits, which makes it difficult to further develop systematic design tools that can handle statistical biocircuit specifications.

In this paper, we present a design-oriented computational framework that directly calculates steady state statistics of stochastic biocircuits and their sensitivity to parameter perturbations (Fig. 1). Building upon a moment computation approach [26, 27, 28], we formulate the biocircuit design problems in the form of convex optimization programs [29], which enable efficient evaluation of the statistics and its sensitivity without running time-consuming simulations of single-cell trajectories. Our optimization based synthesis approach is capable of characterizing feasible design space that satisfies multiple and possibly incompatible performance specifications with mathematical rigor. Thus, it greatly facilitates rational engineering process of noisy biomolecular reactions. In addition, the sensitivity analysis ensures robustness against parameter uncertainty by compensating for errors due to model misidentification and perturbations to the host cell environments.

We use the proposed algorithms to explore the design parameter space of a self-negative feedback biocircuit, where a repressor protein regulates its own expression. Specifically, we run the convex optimization programs and obtain a parameter map of the feasible design space with which the negative feedback biocircuit satisfies pre-specified performance requirements. Interestingly, the parameter maps indicate the existence of an optimal translation rate that minimizes the CV of the repressor copy numbers, implying that increasing the copy number of the protein by strong translation does not necessarily attenuate the noise due to some effects of negative feedback pathways. To better understand the mechanisms, we perform in-depth study of the propagation of the noise and identify two sources

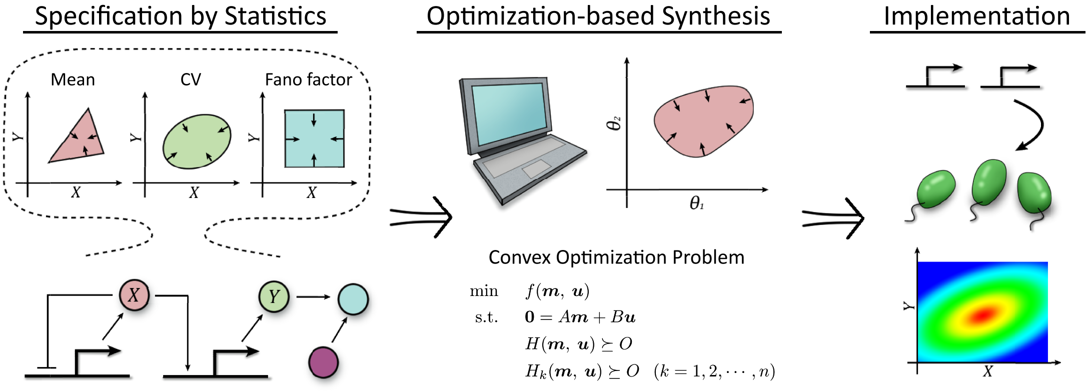

Specification by Statistics
Mean
CV
Fano factor
Y
X
X
Y
Optimization-based Synthesis
θ_2
θ_1
Convex Optimization Problem
min $f(\boldsymbol{m},\,\boldsymbol{u})$
s.t. $\boldsymbol{0} = A\boldsymbol{m} + B\boldsymbol{u}$
$H(\boldsymbol{m},\,\boldsymbol{u}) \succeq O$
$H_k(\boldsymbol{m},\,\boldsymbol{u}) \succeq O \quad (k = 1, 2, \cdots, n)$
Implementation

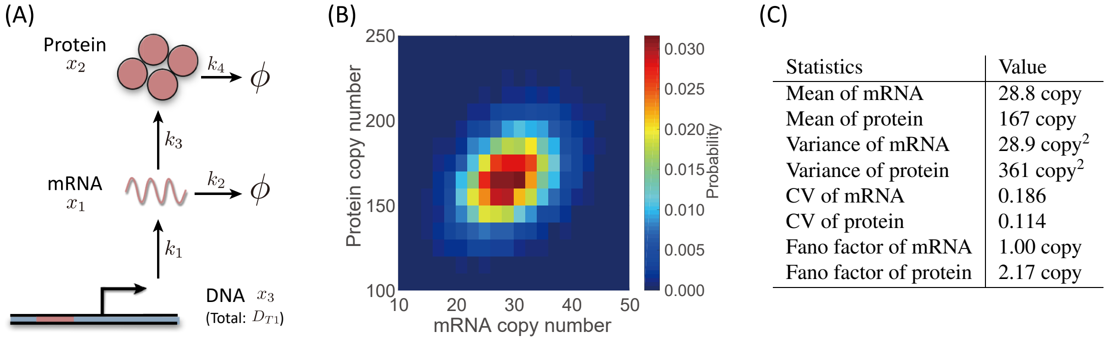
Figure 1: Overview of the optimization based synthesis approach. The optimization program allows for rigorous characterization of parameter space that satisfies given statistical specifications.

of noise in a trade-off relation, which produces an optimal configuration that minimizes the CV of the repressor protein. Informed by these analyses, we determine a design strategy of the negative feedback biocircuits. Sensitivity analysis is further performed to assess the robustness of our design against parameter perturbations.

Results and Discussion

Mathematical model of stochastic biocircuits

We start with a general model of stochastic biomolecular reactions and introduce an ordinary differential equation (ODE) model that describes the evolution of stochastic moments of biocircuits. Suppose a biocircuit consists of n species of molecules that vary in time and r types of chemical reactions. The copy numbers of the n molecules, or the state of biocircuits, fluctuate randomly in time and become heterogeneous between cells due to the stochastic chemical reactions. To model the stochastic dynamics, we denote the copy number of the i-th molecule by x_i $(i = 1, 2, \cdots, n)$ and define the probability that there are $\boldsymbol{x} = [x_1, x_2, \cdots, x_n]^T$ molecules in a cell at time t by $P_{\boldsymbol{x}}(t)$. As an illustration example, we consider a simple transcription-translation process in Fig. 2A. In this example, mRNA and protein copy numbers are the state variables of the biocircuit $(n = 2)$, and there are four reactions $(r = 4)$, transcription, translation and degradation of mRNA and protein. The heterogeneity of a cell population is then captured by the joint distribution of mRNA and protein copy numbers $P_{[x_1,x_2]^T}(t)$, where x_1 and x_2 denote the copy numbers of mRNA



Statistics	Value
Mean of mRNA	28.8 copy
Mean of protein	167 copy
Variance of mRNA	28.9 copy^2
Variance of protein	361 copy^2
CV of mRNA	0.186
CV of protein	0.114
Fano factor of mRNA	1.00 copy
Fano factor of protein	2.17 copy

Figure 2: (A) Schematic diagram of a simple transcription-translation biocircuit. (B) Joint population distribution of mRNA and protein copy numbers at $t = 1440$ min. (C) Summary statistics of the population distribution.

and protein, respectively.

In general, the dynamics of the probability distribution $P_{\boldsymbol{x}}(t)$ are modeled by a set of ODEs called Chemical Master Equation (CME) [30].

$$\frac{dP_{\boldsymbol{x}}(t)}{dt} = \sum_{i=1}^{r} \{w_i(\boldsymbol{x}-\boldsymbol{s}_i)P_{\boldsymbol{x}-\boldsymbol{s}_i}(t) - w_i(\boldsymbol{x})P_{\boldsymbol{x}}(t)\} \quad (\boldsymbol{x} \in \mathbb{N}_0^n), \tag{1}$$

where $w_i(\cdot)$ is a propensity function (reaction rate) associated with the i-th chemical reaction ($i = 1, 2, \cdots, r$), and $\boldsymbol{s}_i$ is a n-dimensional row vector representing the stoichiometry of the i-th reaction. We assume that the reactions are elementary, and thus $w_i(\cdot)$ is a polynomial of x_i ($i = 1, 2, \cdots, n$) [31]. Specific forms of $w_i(\cdot)$ and $\boldsymbol{s}_i$ for the transcription-translation process in Fig. 2A are summarized in Supporting Information S.2. It should be noted that the entries of the vector $\boldsymbol{x}$ in (1) take all combinations of nonnegative numbers, and thus, the CME is composed of infinitely many coupled equations. Although it is hard to analytically solve the equation in terms of $P_{\boldsymbol{x}}(t)$, it is possible to simulate many numbers of single-cell trajectories using a Monte Carlo approach [20] and obtain approximate distributions of $P_{\boldsymbol{x}}(t)$ as illustrated in Fig. 2B. To quantitatively capture the important features of population distributions, widely used statistics are the mean $\mathbb{E}[\boldsymbol{x}]$ and the covariance $\mathbb{E}[(\boldsymbol{x}-\mathbb{E}[\boldsymbol{x}])(\boldsymbol{x}-\mathbb{E}[\boldsymbol{x}])^T] = \mathbb{E}[\boldsymbol{x}\boldsymbol{x}^T]-\mathbb{E}[\boldsymbol{x}]\mathbb{E}[\boldsymbol{x}]^T$, which are the first two central moments of the distribution. Other examples of useful descriptive statistics are the coefficient of variation (CV) $\sqrt{\mathbb{E}[x_i^2] - (\mathbb{E}[x_i])^2}/\mathbb{E}[x_i]$ and Fano factor $(\mathbb{E}[x_i^2]-(\mathbb{E}[x_i])^2)/\mathbb{E}[x_i]$, which quantify the dispersion of distributions. In particular, Fano factor become exactly one if the distribution $P_{\boldsymbol{x}}(t)$ is Poisson. Correlation $\mathbb{E}[x_i x_j]/\sqrt{(\mathbb{E}[x_i^2] - (\mathbb{E}[x_i])^2)(\mathbb{E}[x_j^2] - (\mathbb{E}[x_j])^2)}$, is also a useful measure when we are interested in the relation between two molecules.

The design-oriented computational framework presented in this paper complements the Monte Carlo approach by allowing for rigorous evaluation of descriptive statistics without approximation. For this purpose, we first derive an ODE model that describes the dynamics of raw moments, or a moment equation for short, based on the CME (1). To elucidate the following mathematical development, we first consider a specific model for the transcription-translation process in Fig. 2A. Let $\boldsymbol{m}$ denote a vector of raw moments $\boldsymbol{m} := \big[\mathbb{E}[1],\ \mathbb{E}[x_1],\ \mathbb{E}[x_2],\ \mathbb{E}[x_1^2],\ \mathbb{E}[x_1 x_2],\ \mathbb{E}[x_2^2]\big]$, and consider to derive an ODE model for $\boldsymbol{m}$. The basic idea for the derivation is to multiply x_i's to both sides of the CME (1) and take the sum of x_i's for all nonnegative numbers (Supporting Information S.2 for details). Using this approach, we obtain the moment dynamics as

$$\frac{d}{dt}\boldsymbol{m}(t) = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ k_1 D_{T_1} & -k_2 & 0 & 0 & 0 & 0 \\ 0 & k_3 & -k_4 & 0 & 0 & 0 \\ k_1 D_{T_1} & 2k_1 D_{T_1} + k_2 & 0 & -2k_2 & 0 & 0 \\ 0 & 0 & k_1 D_{T_1} & k_3 & -k_2 - k_4 & 0 \\ 0 & k_3 & k_4 & 0 & 2k_3 & -2k_4 \end{bmatrix} \boldsymbol{m}(t), \tag{2}$$

where the first entry of $\boldsymbol{m}(t) = \mathbb{E}[1]$ represents the sum of the zero-th order moments of x_1 and x_2, guarantees the sum of the probability $P_{\boldsymbol{x}}(t)$ to be one. Note that no approximation is used in the derivation. In particular, the equation (2) is a linear ODE, and thus, we can rigorously calculate the raw moments $\boldsymbol{m}$ by solving (2).

When the reactions reach steady state, the left-hand side of (2), which is the time derivative of $\boldsymbol{m}(t)$, goes to zero, leading to a set of linear equations. Thus, we solve the linear equations to obtain the first and second order steady state raw moments of the protein copy number, x_2 as

$$\mathbb{E}[x_2] = \frac{k_1 k_3}{k_2 k_4} D_{T_1}, \tag{3}$$

$$\mathbb{E}[x_2^2] = \frac{k_1 k_3}{k_2 k_4} D_{T_1} \left(1 + \frac{k_3}{k_2 + k_4} + \frac{k_1 k_3}{k_2 k_4} D_{T_1}\right). \tag{4}$$

Using these solutions, we can further compute the variance of the protein copy number as

$$\mathbb{E}[x_2^2] - \mathbb{E}[x_2]^2 = \frac{k_1 k_3}{k_2 k_4} D_{T_1} \left(1 + \frac{k_3}{k_2 + k_4}\right). \tag{5}$$

Substituting parameter values, we confirm that the analytic solution indeed agrees with the simulated statistics (Fig. 2C). In the design process of stochastic biocircuits, these analytic solutions are useful for characterizing the parameter space that satisfies design requirements and narrowing possible combinations of genetic parts of biocircuits.

Computing descriptive statistics using semi-algebraic optimization

Unfortunately, analytic solutions are not necessarily available when a biocircuit of interest is slightly more complicated since, in general, a moment is dependent on other (higher order) moments, and the order of a moment equation is infinite. More formally, we define raw moments of a distribution $P_{\boldsymbol{x}}(t)$ by

$$m_{\boldsymbol{\alpha}}(t) := \mathbb{E}\left[\prod_{j=1}^{n} x_j^{\alpha_j}\right] = \sum_{x_1=0}^{\infty} \sum_{x_2=0}^{\infty} \cdots \sum_{x_n=0}^{\infty} \prod_{j=1}^{n} x_j^{\alpha_j} P_{\boldsymbol{x}}(t) \tag{6}$$

with $\boldsymbol{\alpha} := [\alpha_1, \alpha_2, \cdots, \alpha_n]^T \in \mathbb{N}_0^n$ and refer to the sum $\sum_{i=1}^n \alpha_i$ as the order of the moment. A general form of the moment equation is then obtained as

$$\frac{d}{dt}\boldsymbol{m} = A\boldsymbol{m} + B\boldsymbol{u}, \tag{7}$$

where A and B are constant matrices, $\boldsymbol{m}$ is a vector of raw moments up to the μ-th order, and $\boldsymbol{u}$ is a vector of the $\mu + 1$-th or higher order moments $\boldsymbol{u}$ (see Supporting Information S.3 for details). Note that (2) is a special case of (7) with $B = \mathbf{0}$. Equation (7) implies that the μ-th order moments $\boldsymbol{m}$, which are the moments of our interest, depend on the higher order moments $\boldsymbol{u}$. Thus, it is not possible to uniquely determine the solution of the steady state moment equation $A\boldsymbol{m} + B\boldsymbol{u} = \mathbf{0}$ since there are more variables than equations. In fact, analytic steady state moments are available only in the special case of $B = \mathbf{0}$, in which case $\boldsymbol{m}$ is obtained by solving $A\boldsymbol{m} = \mathbf{0}$ as shown in (3) and (4). In general, $B = \mathbf{0}$ holds if and only if all reactions are the zero-th or the first order, that is, the reaction rates $w_i(\boldsymbol{x})$ are affine in $\boldsymbol{x}$ (see Supplementary Information S.3).

To see an example, we consider a negative feedback biocircuit in Fig. 3A, where the expression of the repressor protein is self-regulated by the negative feedback. This biocircuit, despite a slight extension of Fig. 2A, contains a bimolecular reaction, namely the binding of the repressor to the promoter whose propensity function is given by $w(\boldsymbol{x}) = k_5 x_2 x_3$. As a result, the matrix B in (7) is no longer zero, and there are infinitely many solutions for the steady state moment equation unless we know additional information that links $\boldsymbol{m}$ and higher order moments $\boldsymbol{u}$.

To constrain the solution, a key observation is that the variables $\boldsymbol{m}$ and $\boldsymbol{u}$ must constitute moments of some probability distribution defined on the positive orthant $\{[x_1, x_2, \cdots, x_n] \mid x_i > 0,\ i = 1, 2, \cdots, n\}$. An obvious necessary condition is that all entries of $\boldsymbol{m}$ and $\boldsymbol{u}$ must be positive according to the definition (6). In fact, there are tighter conditions that the variables $\boldsymbol{m}$ and $\boldsymbol{u}$ must satisfy to be moments of some probability distribution (Proposition A.1 in Supporting Information)[32, 33]. Incorporating these conditions, we can narrow possible combinations of raw moments $\boldsymbol{m}$ and $\boldsymbol{u}$ as specified by the following proposition.

Proposition. Consider stochastic chemical reactions modeled by the CME (1). The steady state moments of the probability distribution satisfy

$$\begin{aligned} &\mathbf{0} = A\boldsymbol{m} + B\boldsymbol{u}, \\ &H^{(\gamma_1)}(\boldsymbol{m}, \boldsymbol{u}) \succeq O, \\ &H_k^{(\gamma_2)}(\boldsymbol{m}, \boldsymbol{u}) \succeq O \quad (k = 1, 2, \cdots, n), \end{aligned} \tag{8}$$

where the matrices $H^{(\gamma_1)}(\boldsymbol{m}, \boldsymbol{u})$ and $H_k^{(\gamma_2)}(\boldsymbol{m}, \boldsymbol{u})$ represent moment matrices defined in (A.21) and (A.22) of Supporting Information. The symbol $X \succeq O$ represents that a matrix X is positive semidefinite.

This proposition implies that the raw moments $\boldsymbol{m}$ and $\boldsymbol{u}$ lie in the semi-algebraic set specified by (8). Although it is hard to uniquely determine $\boldsymbol{m}$ and $\boldsymbol{u}$ from these conditions, equations (8) imply that we can computationally search for possible combinations of $\boldsymbol{m}$ and $\boldsymbol{u}$ based on (8). In particular, we can find the upper and/or lower bounds of statistical values such as the covariance, CV and Fano factor of molecular copy numbers. In what follows, we show that the problem of finding the upper and the lower bounds of these statistics can be recast as a mathematical optimization problem, which we can solve efficiently using existing algorithms of mathematical programming.

We consider the negative feedback biocircuit in Fig. 3A and define x_1, x_2 and x_3 as the copy numbers of mRNA, repressor protein and free DNA, respectively. Our goal here is to

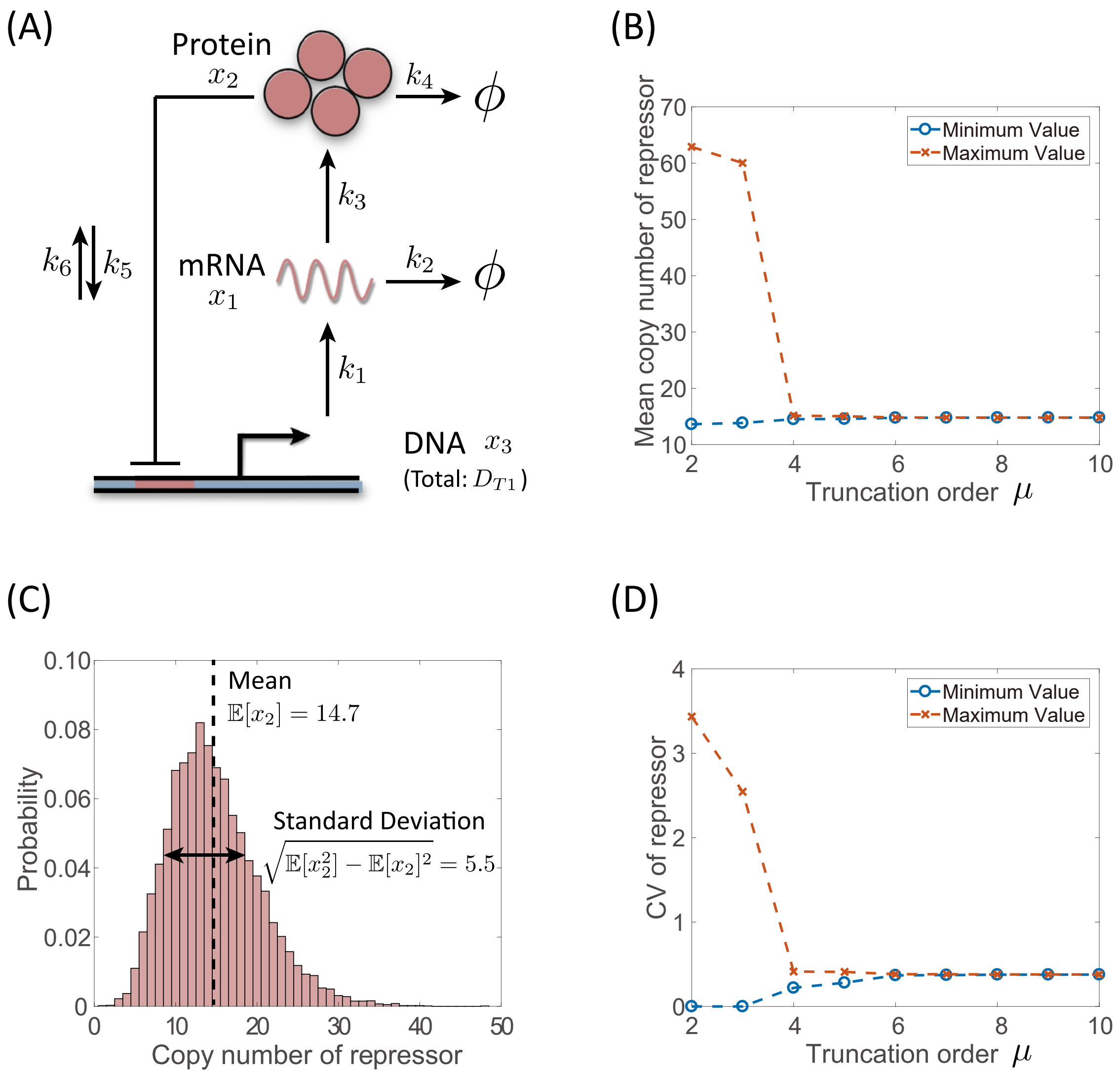


Figure 3: (A) Schematic diagram of a negative feedback biocircuit. (B) Distribution of the repressor copy number at $t = 1440$ min. (C) Computed upper and lower bounds of the mean copy number for different truncation orders μ. (D) Computed upper and lower bounds of the variance of the copy number for different truncation orders μ.

compute the mean and the CV of the repressor copy number x_2 without running stochastic simulations of single-cell trajectories. To this end, we use the semi-algebraic constraint (8) and formulate a maximization problem of the mean and the CV as

$$\max_{\boldsymbol{m}} h(\boldsymbol{m}) \ \text{ subject to } \ (8), \tag{9}$$

where $h(\boldsymbol{m})$ is defined by $h(\boldsymbol{m}) = m_{[0,1,0]^T} := \mathbb{E}[x_2]$ for the mean, and $h(\boldsymbol{m}) = \sqrt{m_{0,2,0} - m_{0,1,0}^2}/m_{0,1,0} := \sqrt{\mathbb{E}[x_2^2] - \mathbb{E}[x_2]^2}/\mathbb{E}[x_2]$ for the CV, respectively. Then, the solution of this problem gives upper bounds of the mean and the CV, respectively. An advantage of using the optimization approach is that we can leverage efficient algorithms for mathematical optimization, whose techniques were extensively studied in engineering science. In particular, the computation of these summary statistics can be recast as semi-definite programming (SDP) [34, 29], which is a subclass of convex optimization program with many practically useful properties such that it allows for finding global minimum (or maximum) with much less computational efforts than other mathematical optimization (Supplementary Information S.4 for details).

Using the SDP approach, we computed the lower and the upper bounds of the mean repressor copy number, $\mathbb{E}[x_2]$ for different values of μ, which is a user-specified parameter that determines the largest order of moments in the vector $\boldsymbol{m}$ in (8) (Fig. 3B). Note that the lower bounds are obtained by solving a similar form of optimization that maximizes $-h(\boldsymbol{m})$. As we increase μ, the gap between the upper and lower bounds decreases in general, allowing for better estimation of statistical values at the expense of computational time (see Supporting Information). For the biocircuit in Fig. 3A, the estimated mean copy number of the repressor $\mathbb{E}[x_2]$ converged to 14.7 (Fig. 3B), which agrees with the mean value of the approximate distribution computed by Monte Carlo simulations [20] (Fig. 3C).

The upper bound of the CV is an important performance norm to quantify the dispersion of the population distribution. Since the optimization of the CV in (9) is not directly solvable by SDP, we developed a procedure to recast the optimization problem (9) into a SDP form by introducing additional variables (see Supplementary Information S.4). Using this approach, we computed the upper bounds of the CV for different values of μ as illustrated in Fig. 3D, where the lower bounds were also computed for a reference. We observe that, similar to the mean value, the lower and the upper bounds of the CV approach as we increase the order of the moments μ, which implies that the estimation becomes more accurate.

Table 1: Summary statistics of stochastic biocircuits computable by semidefinite programming

Statistics	Mathematical representation
Upper and lower bounds of mean	$\mathbb{E}[x_i]$
Upper bound of variance and covariance	$\mathbb{E}[(x_i - \mathbb{E}[x_i])(x_j - \mathbb{E}[x_j])]$
Upper bound of coefficient of variation	$\sqrt{\mathbb{E}[x_i^2] - (\mathbb{E}[x_i])^2}/\mathbb{E}[x_i]$
Upper bound of Fano factor	$(\mathbb{E}[x_i^2] - (\mathbb{E}[x_i])^2)/\mathbb{E}[x_i]$
Upper bound of the largest confidence ellipsoid	

Similar mathematical techniques apply to other descriptive statistics and allow us to rigorously compute covariance, Fano factor and confidence ellipsoids of the molecular copy numbers using semidefinite programming (Table 1). Specific forms of these optimizations and their mathematical proofs are summarized in Supplementary Information S.4. As an illustrative example, we computed the largest confidence ellipsoid of mRNA and protein copy numbers of the transcription-translation circuit in Fig. 2A. The two-dimensional confidence ellipsoid allows for visualizing the correlation between the two molecules (Fig. S.1). In the design process, the confidence ellipsoids would be useful to investigate how tightly a target molecule is regulated by an upstream molecule. Other optimizations will be demonstrated in the following sections along with the design examples of a negative feedback biocircuit.

Synthesizing biocircuits with statistical design specifications

The process of biocircuit engineering requires many iterations of design-build-test cycles to achieve prescribed performance requirements. Since the specifications are possibly incompatible or conflicting, computational design tools are important to efficiently explore and find the feasible design space of biocircuits. Our optimization approach allows for rigorous characterization of biocircuit parameter space satisfying multiple design requirements described by statistical constraints (Table 1). Specifically, we use a set of inequalities to mathematically specify the design requirements of biocircuits. For example, $\gamma - \mathbb{E}[x] \leq 0$ implies that the copy number of a molecule, say x, must be more than γ, and $\sqrt{\mathbb{E}[x^2] - (\mathbb{E}[x])^2}/\mathbb{E}[x] \leq \delta$ implies that the CV must be less than δ at steady state. More formally, we denote biocircuit specifications by $f_i(\boldsymbol{m}) \leq 0$ $(i = 1, 2, \cdots, s)$ with raw moments $\boldsymbol{m}$. Using the convex optimization presented in the previous section, we can rigorously determine whether a given circuit design satisfies these performance specifications.

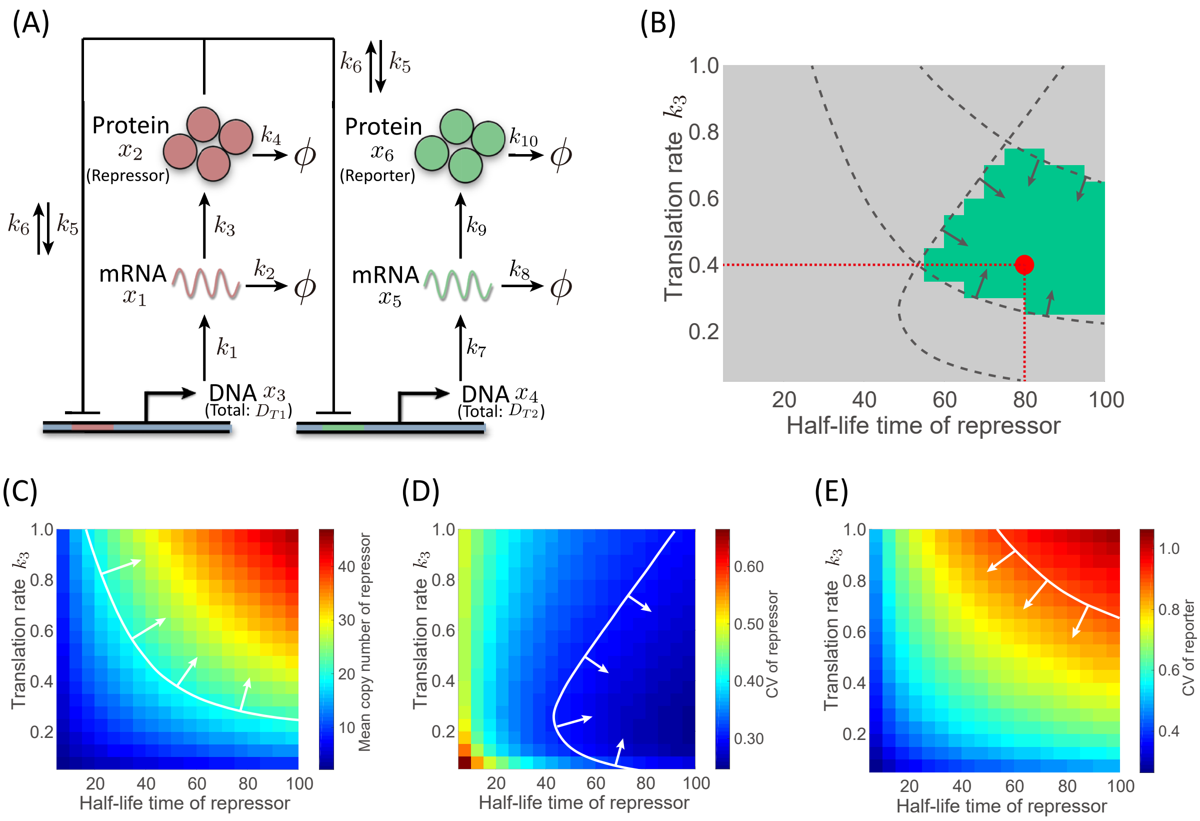


Figure 4: (A) Schematic diagram of a negative feedback biocircuit with a downstream reporter protein. (B) Parameter region satisfying the three design specifications. (C) The lower bound of the mean copy number of the repressor protein. (D) The upper bound of the CV of the repressor protein. (E) The upper bound of the CV of the reporter protein.

To demonstrate the optimization based synthesis method, we consider to design a biocircuit in Fig. 4A, where a reporter protein is added to the downstream of the negative feedback circuit in Fig. 3A. As design specifications, we require the biocircuit to satisfy the following three performance criteria at steady state: (i) the mean copy number of the repressor molecule is at least 20, (ii) the CV of the repressor protein is less than 0.30, and

(iii) the CV of the reporter protein is less than 0.90. These specifications are translated as

$$\begin{aligned} f_1(\boldsymbol{m}) &= -\mathbb{E}[x_2] + 20 \leq 0, \\ f_2(\boldsymbol{m}) &= \frac{\sqrt{\mathbb{E}[x_2^2] - \mathbb{E}[x_2]^2}}{\mathbb{E}[x_2]} - 0.30 \leq 0, \\ f_3(\boldsymbol{m}) &= \frac{\sqrt{\mathbb{E}[x_6^2] - \mathbb{E}[x_6]^2}}{\mathbb{E}[x_6]} - 0.90 \leq 0, \end{aligned} \tag{10}$$

where x_2 and x_6 denote the copy number of the repressor and the reporter proteins, respectively.

For illustration purpose, we consider two tuning parameters, the translation rate k_3 and the degradation rate k_4 of the repressor protein. Note that these parameters can be tuned, for examlpe, by engineering the ribosome binding site [35] and degradation tags of the protein, respectively. Using the semidefinite programs presented in the previous section, we produced a parameter map showing the feasible design space with which the biocircuit satisfies all of the three statistical design requirements in (10) (Fig. 4B). Figure 4B illustrates that the three design features specified by (10) are in a trade-off relationship in that moving a parameter to one direction satisfies one constraint but violates another. Thus, we need to carefully choose parameters in the middle of the parameter space. The parameter map provides valuable information to narrow the potential combinations of genetic parts to be tested and reduces the iterations of design-build-test cycles. To verify the result of the parameter space exploration, we simulated the stochastic biomolecular reactions of the negative feedback biocircuit using the stochastic simulation algorithm [20], where the parameters were taken from the feasible design space as illustrated by the red dot in Fig. 4B. The mean copy number of the repressor protein was 26.2 copy, the CV of the repressor and the reporter proteins were 0.273 and 0.725, respectively, which all meets the design specifications.

Noise attenuation requires balanced expression and repression

We further investigated the statistical values of the negative feedback biocircuit in detail to better understand the underlying mechanisms that limit the feasible design space in Fig. 4B and clarify design strategies (Fig. 4C–E). Figure 4C shows that increasing the translation rate of the repressor protein results in the increase of the repressor copy number at steady state despite the negative feedback. This implies that the translation of mRNA has a more

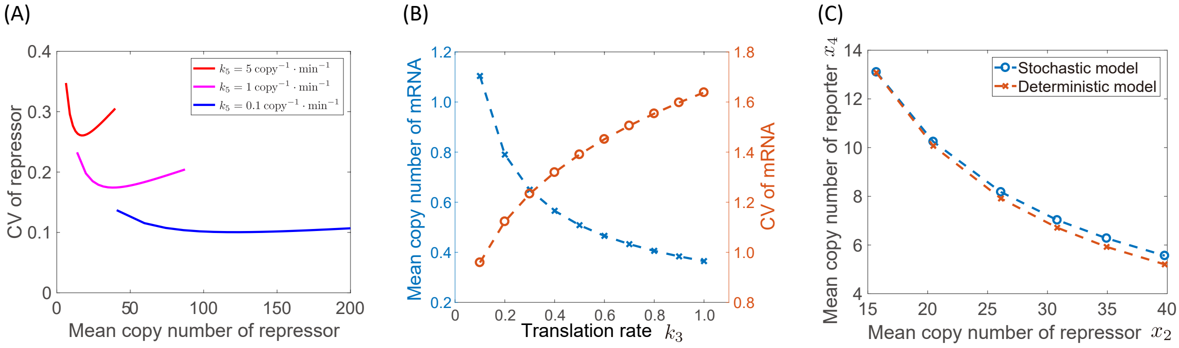


Figure 5: (A) Non-monotonic relation between the CV and the copy number of the repressor protein. (B) The mean and the CV of the mRNA copy number. (C) The mean and the CV of the repressor protein. The mean copy number does not follow Michaelis-Menten kinetics.

influence on the total copy number of the repressor protein than the negative feedback. Thus, a design strategy for meeting the specification $f_1(\boldsymbol{m}) \leq 0$, which is to maintain the copy number of the repressor at least 20 molecules is to increase the translation rate k_3. On the other hand, Fig. 4D illustrates that the CV of the repressor copy number does not decrease monotonically with k_3, suggesting that the two design features $f_1(\boldsymbol{m}) \leq 0$ and $f_2(\boldsymbol{m}) \leq 0$ are in a trade-off relationship.

It is interesting to observe that there is an optimal strength of translation k_3 that minimizes the CV of the repressor copy number (Fig. 4D). Since the intrinsic noise of biocircuits comes from the low copy nature of molecules, it is counterintuitive that both of the mean and the CV of the repressor increase at the same time in Fig. 4C, D. We observed that this trend is generic for a wide range of the repressor-promoter binding rate k_5, which directly controls the strength of the negative feedback (Fig. 5A). We suspect that this is due to a trade-off relation between the strength of repression and transcription. More specifically, increasing the translation rate results in the attenuation of noise due to the high copy numbers of the repressor protein, but at the same time, it also increases the variance of the mRNA copy number due to the strong repression as illustrated in Fig. 5B. Then, the highly stochastic mRNA transcription indirectly contributes to increasing the dispersion of the copy number of the repressor protein. Fig. 5A suggests that the former is dominant when the translation rate k_3 is small, but the latter becomes dominant as the increase of k_3. These analysis results suggest that balancing the repression and the expression is a key to attenuate the noise in negative feedback biocircuits.

Expression level of reporter protein is dependent on the structure of upstream biocircuits

A typical approach to probing protein concentrations, or internal states of biocircuits, without disrupting cells is to express a fluorescent reporter protein at the downstream of a target molecule. The internal states are then indirectly quantified based on the fluorescence measurements. Characterizing the expression levels of the reporter versus the target molecule is thus essential for rigorously quantifying the internal states of biocircuits. For the biocircuit in Fig. 4A, the mean expression level of the reporter protein is given by $(k_7k_9/k_8k_{10})\mathbb{E}[x_4]$ as suggested by (3), where $\mathbb{E}[x_4]$ represents the mean copy number of the free DNA that is not bound by the repressor protein. To calculate $\mathbb{E}[x_4]$, the moment equation is

$$\frac{d}{dt}\mathbb{E}[x_4] = k_6 D_{T_2} - k_6\mathbb{E}[x_4] - k_5\mathbb{E}[x_2 x_4]. \tag{11}$$

Equation (11) implies that $\mathbb{E}[x_4]$ depends on the second order moment $\mathbb{E}[x_2x_4]$, and thus, higher order moments are necessary to fully characterize $\mathbb{E}[x_4]$. In other words, the mean copy number of the reporter $\mathbb{E}[x_6]$ cannot be determined simply from the mean copy number of the repressor $\mathbb{E}[x_2]$ but it requires higher order statistics, which indirectly depends on moments of the upstream negative feedback pathways via moment matrices $H^{(\gamma_1)}(\boldsymbol{m}, \boldsymbol{u}) \geq O$ and $H_k^{(\gamma_2)(\boldsymbol{m},\boldsymbol{u})} \geq O$ in (8). This is in contrast with the deterministic modeling, where the steady state concentration of the free DNA x_4 is expressed by the Mechaelis-Menten equation

$$x_4 = \frac{K}{K + x_2} D_{T_2} \tag{12}$$

with $K = k_6/k_5$.

Using the convex optimization program, we characterized the reporter expression level versus the repressor copy number in Fig. 5C, where the Michaelis-Menten equation (12) is superimposed. The figure clearly illustrates that the reporter copy number deviates from the Michaelis-Menten kinetics with the maximum relative error of 6%, suggesting that the simple Michaelis-Menten kinetics is erroneous especially when the biocircuit is highly stochastic.

Sensitivity analysis of descriptive statistics

The ability of model-based biocircuit design is currently limited by the uncertainty of parameter values in mathematical models. The source of the uncertainty partly lies in misidentified parameters due to insufficient and noisy measurements, but more inherently, it lies in extrinsic perturbations to host cell environments such as growth conditions. As a result, the process of biocircuit engineering often requires ad-hoc tuning of circuit parameters to deal with the deviation of circuit performance from model predictions. Sensitivity analysis allows for quantifying the impact of model uncertainties on the behavior of biocircuits and finding sensitive design parameters that need special attention in the build process.

We developed semidefinite optimization programs to evaluate the sensitivity of the descriptive statistics in Table 1. Specifically, let $\boldsymbol{k}^* = [k_1^*, k_2^*, \cdots, k_r^*]^T$ denote a vector of nominal parameters with which a biocircuit satisfies performance specifications $f_i(\boldsymbol{m}) \leq 0$ $(i = 1, 2, \cdots, s)$. We consider a parametric perturbation $\boldsymbol{k}^* \pm \boldsymbol{\Delta} k_i$, where $\boldsymbol{\Delta} k_i := [0, 0, \cdots, 0, \Delta k_i, 0, \cdots, 0]^T$ is a (small) perturbation to the nominal parameter k_i^*. The goal of the sensitivity analysis is to find the range of the statistics under the perturbation $|\Delta k_i| \leq \delta$, where δ is a given constant. This can be formulated in an optimization form as

$$\begin{aligned} \max_{\Delta k_i, \boldsymbol{m}} \quad & h(\boldsymbol{m}) \\ \text{subject to} \quad & \boldsymbol{0} = A(\Delta k_i)\boldsymbol{m} + B(\Delta k_i)\boldsymbol{u}, \\ & -\delta \leq \Delta k_i \leq \delta, \\ & H(\boldsymbol{m}, \boldsymbol{u}) \succeq O, \\ & H_k(\boldsymbol{m}, \boldsymbol{u}) \succeq O \quad (k = 1, 2, \cdots, n), \end{aligned}$$

where we denote the perturbed coefficient matrices in (7) by $A(\Delta k_i)$ and $B(\Delta k_i)$. We convert this optimization program into a convex form to enable efficient computation of the worst-case statistics for all possible parameter combinations satisfying $|\Delta k_i| \leq \delta$ with mathematical rigor (see Supporting Information S.6 for details). In other words, the optimization program can strictly guarantee the robustness of biocircuits for all of the parameters satisfying $|\Delta k_i| \leq \delta$.

Using this approach, we performed sensitivity analysis of the negative feedback biocircuit

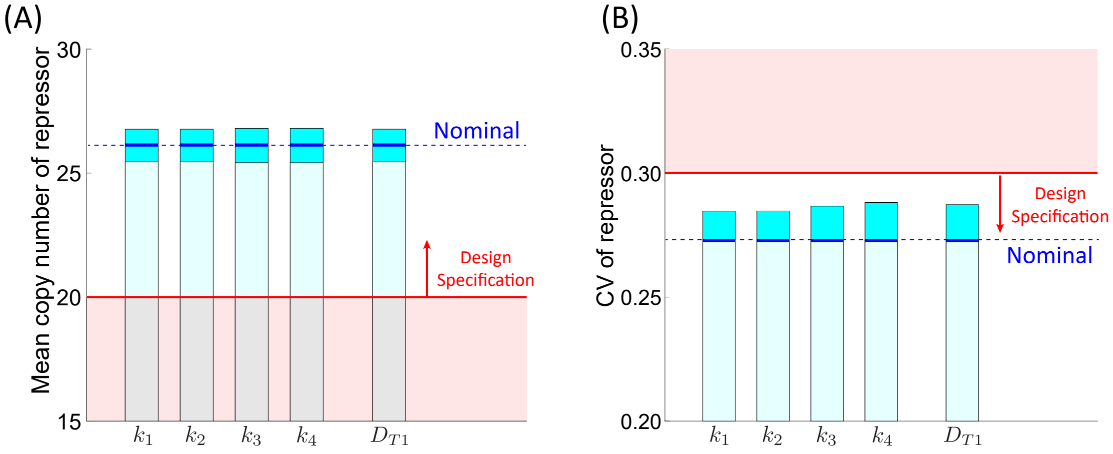


Figure 6: Sensitivity analysis for parameter perturbation. The red region shows the design specifications (10). (A) Sensitivity analysis of the mean copy number of the repressor protein. The blue region shows the worst-case upper and lower bounds for parameter perturbation. (B) Sensitivity analysis of the CV of the repressor protein. The blue region shows the worst-case upper bound for parameter perturbation.

in Fig. 4A around a nominal parameter value shown as the red dot in Fig. 4B. Initially, we computed the worst-case mean and the CV of the repressor protein X when k_1, k_2, k_3 and k_4 are perturbed within 5% of the nominal value (Fig. 6). The result shows that the deviation of the statistics is almost equal between the perturbations, implying that there is no highly sensitive parameters that significantly affect these statistics. Figure 6 also illustrates that the mean and the CV do not violate the design constraints $f_1(\boldsymbol{m}) = -\mathbb{E}[x] + 20 \leq 0$ and $f_2(\boldsymbol{m}) = (\sqrt{\mathbb{E}[x^2] - \mathbb{E}[x]^2})/\mathbb{E}[x] - 0.30 \leq 0$, which guarantees that the negative feedback biocircuit designed in Fig. 4 is robust against these parameter perturbations.

Another important but often overlooked design parameter of biocircuits is the plasmid copy number, which is controlled by the replication origin of a circuit plasmid. Although the plasmid copy number is assumed constant in Fig. 4B, variance of the plasmid copy number in real biological cells affects the behavior of biocircuits. To analyze the effect of plasmid copy numbers, we applied the same approach to computing the worst-case performance of the negative feedback biocircuit against 5% deviation of the copy number of the repressor plasmid from the nominal value D_{T1}. From Fig. 6, we can guarantee that the designed negative feedback biocircuit can operate within the pre-specified range of performance norms even under the extrinsic perturbation to the plasmid copy number.

Discussion

A promising approach toward robust engineering of complex biocircuits is to guarantee the performance of individual circuit components at high precision. The highly stochastic nature of biomolecular reactions, however, hinders reliable assessment of biocircuit behaviors in current synthetic biology. To advance a model-guided design approach, it is critical to develop design-oriented theoretical tools that can rigorously certify robustness of stochastic chemical reactions.

In this paper, we have presented an optimization based approach to designing stochastic biocircuits. The presented approach allows for specifying the design features of biocircuits using intuitive and well-defined metrics of descriptive statistics. The mathematical optimization algorithms enable systematic exploration of the design space to find parameter configurations satisfying the specifications. In contrast with approximation based approaches [25, 36], the presented method provides mathematically rigorous certification of circuit performance based on user-specified statistical norms. Thus, it is amenable for robust synthesis of stochastic biocircuits that require high reliability. Moreover, the convex nature of the optimization programs allows for efficient search of the optimal solutions by benefitting from existing algorithms of mathematical optimization.

To demonstrate these features, we have explored the design space of the negative feedback biocircuit in Fig. 4A and obtained the parameter map of feasible design space with which the biocircuit satisfies design requirements. In particular, the optimization based analysis elucidated that there is an optimal translation rate of the repressor protein that best attenuates intrinsic noise and that it is caused by the tradeoff relation between repression and expression. It is worth noting that a similar tradeoff relation was previously predicted for a metabolic pathway [37] and the repressilator [38] based on approximated model-based analyses. A similar U-shaped trend to Fig. 5A was also observed by experiments [39]. These examples suggest that even simple biocircuits can exhibit complex noise characteristics, which emphasizes the importance of advanced mathematical and computational frameworks for analyzing stochastic biocircuits.

Although not discussed, multi-modality of population distributions is one of the important design features that would likely to be included in the specifications of stochastic biocircuits. In synthetic biology, multimodal population distribution is often associated with multi-stability of the governing dynamics of biocircuits and is used to build switch-like systems as represented by the celebrated genetic toggle switch [15]. An optimization based approach was recently developed to design multimodal biocircuits by directly minimizing the deviation of the distribution from a desired shape [40]. As of yet, however, there has not been a well-defined statistical metric that can quantitatively certify the existence of multimodal distributions, though a recent study indicated that most information of bimodal distributions is encoded in a small number of low order moments [41]. Future work will aim to establish statistical criteria for more advanced design features to enhance rational engineering process of complex stochastic biocircuits.

Method

Stochastic simulations

The stochastic simulation algorithm [20] was used to simulate time trajectories of molecular copy counts. $10,000$ cells were simulated to draw a snapshot of the population distribution at 1440 minute in Fig. 2B, Fig. 3C. The following parameter values were used for the simulations. $k_1 = 0.2$ min^{-1}, $k_2 = \ln(2)/5$ min^{-1}, $k_3 = 0.5$ min^{-1}, $k_4 = \ln(2)/20$ min^{-1} $k_5 = 5$ copy$^{-1}\cdot$min^{-1}, $k_6 = 1$ min^{-1}, $k_7 = 0.2$ min^{-1}, $k_8 = \ln(2)/5$ min^{-1}, $k_9 = 0.5$ min^{-1}, $k_{10} = \ln(2)/20$ min^{-1}. $D_{T_1} = 20$ copy was used for the simulation in Fig. 2B, and $D_{T_1} = D_{T_2} = 50$ copy was used for Fig. 3C. The red dot in Fig. 4B corresponds to $k_3 = 0.4$ and $k_4 = \ln(2)/80$. The initial copy numbers were assumed $x_1 = 0$ and $x_2 = 0$ for Fig. 2B, and $x_1 = 0$ copy, $x_2 = 0$ copy and $x_3 = 0$ copy for Fig. 3B. All simulations were run by MATLAB 2016b.

Optimization based computation of statistics

The semidefinite programs were solved with MATLAB 2016b and Sedumi 1.32 solver [42], where the following options of the solver were used. $\text{pars.eps} = 0, \text{pars.alg} = 2, \text{pars.theta} = 0.01, \text{pars.beta} = 0.9, \text{pars.stepdif} = 1, \text{pars.free} = 1, \text{pars.cg.maxiter} = 500, \text{pars.cg.refine} = 10, \text{pars.cg.stagtol} = 5 \times 10^{-20}$, $\text{pars.cg.restol} = 5 \times 10^{-10}$, $\text{pars.chol.canceltol} = 10^{-20}$, $\text{pars.chol.maxuden} = 4000$. The variables $\boldsymbol{m}$ and $\boldsymbol{u}$ were normalized as appropriate by constants to avoid numerical instability.

The truncation order μ was set $\mu = 8$ to compute the mean and the CV of the repressor in Fig. 4C, D and, and $\mu = 6$ to compute the CV of the reporter protein in Fig. 4E.

For the analysis of negative feedback biocircuits in Fig. 5, the optimization problems were solved for different values of k_3. The other parameter values were set equal to those shown in the "stochastic simulations" section, and the truncation order was set $\mu = 8$. To vary the mean copy number of the repressor at steady state in Fig. 5A, the translation rate k_3 was scanned between 0.025 and 0.9. For Fig. 5C, $k_3 = 0.15, 0.25, 0.40, 0.55, 0.70, 0.90$ were used. The truncation order $\mu = 6$ was used for the sensitivity analysis in Fig. 6.

Acknowledgments

This work was supported in part by JSPS KAKENHI Grant Number JP16H07175 and Keio Gijuku Academic Development Funds.

References

[1] K. J. Astrom and P. R. Kumar, "Control: a perspective," *Automatica*, vol. 50, no. 1, pp. 3–43, 2014.

[2] D. Del Vecchio, A. J. Dy, and Y. Qian, "Control theory meets synthetic biology," *Journal of the Royal Society Interface*, vol. 13, no. 120, 2016.

[3] J. C. Anderson, C. A. Voigy, and A. P. Arkin, "Environmental signal integration by a modular AND gate," *Molecular Systems Biology*, vol. 3, p. 133, 2007.

[4] A. Tamsir, J. J. Tabor, and C. A. Voigt, "Robust multicellular computing using genetically encoded nor gates and chemical 'wires'," *Nature*, vol. 469, no. 7329, pp. 212–215, 2011.

[5] T.-S. Moon, C. Lou, A. Tamsir, B. C. Stanton, and C. A. Voigt, "Genetic programs constructed from layered logic gates in single cells," *Nature*, vol. 491, no. 7423, pp. 249–253, 2012.

[6] M. B. Elowitz and S. Leibler, "A synthetic oscillatory network of transcriptional regulators," *Nature*, vol. 403, no. 6767, pp. 335–338, 2000.

[7] T. Danino, O. Mondragón-Palomino, L. Tsimring, and J. Hasty, "A synchronized quorum of genetic clocks," *Nature*, vol. 463, no. 7279, pp. 326–330, 2010.

[8] H. Niederholtmeyer, Z. Sun, Y. Hori, E. Yeung, A. Verpoorte, R. M. Murray, and S. J. Maerkl, "Rapid cell-free forward engineering of novel genetic ring oscillators," *eLife*, vol. 4, p. e09771, 2015.

[9] L. Potvin-Trottier, N. D. Lord, G. Vinnicombe, and J. Paulsson, "Synchronous long-term oscillations in a synthetic gene circuit," *Nature*, vol. 538, pp. 514–517, 2016.

[10] J. W. Kotula, S. J. Kerns, L. A. Shaket, L. Siraj, J. J. Collins, J. C. Way, and P. A. Silver, "Programmable bacteria detect and record an environmental signal in the mammalian gut," *Proceedings of National Academy of Sciences of the United States of America*, vol. 111, pp. 4838–4843, 2014.

[11] L. Yang, A. A. K. Nielsen, J. Fernandez-Rodriguez, C. J. McClune, M. T. Laub, T. K. Lu, and C. A. Voigt, "Permanent genetic memory with $>$ 1 byte capacity," *Nature Method*, vol. 11, no. 12, pp. 1261–1266, 2014.

[12] H. H. McAdams and A. Arkin, "Stochastic mechanisms in gene expression," *Proceedings of National Academy of Sciences of the United States of America*, vol. 94, no. 3, pp. 814–819, 1997.

[13] M. B. Elowitz, A. J. Levine, E. D. Siggia, and P. S. Swain, "Stochastic gene expression in a single cell," *Science*, vol. 297, no. 5584, pp. 1183–1186, 2002.

[14] J. M. Raser and E. K. O'Shea, "Noise in gene expression: origins, consequences and control," *Science*, vol. 309, no. 5743, pp. 2010–2013, 2005.

[15] T. S. Gardner, C. R. Cantor, and J. J. Collins, "Construction of a genetic toggle switch in escherichia coli," *Nature*, vol. 403, no. 6767, pp. 339–342, 2000.

[16] D. Nevozhay, R. M. Adams, E. V. Itallie, M. R. Bennett, and G. Balazsi, "Mapping the environmental fitness landscape of a synthetic gene circuit," *PLOS Computational Biology*, vol. 8, no. 4, p. e1002480, 2012.

[17] M. Wu, R.-Q. Sub, X. Lia, T. Ellis, Y.-C. Lai, and X. Wang, "Engineering of regulated stochastic cell fate determination," *Proceedings of National Academy of Sciences of the United States of America*, vol. 110, no. 26, pp. 10610–10615, 2013.

[18] V. Hsiao, Y. Hori, P. W. K. Rothemund, and R. M. Murray, "A population-based temporal logic gate for timing and recording chemical events," *Molecular Systems Biology*, vol. 12, no. 869, 2016.

[19] S. R. Biggar and G. R. Crabtree, "Cell signaling can direct either binary or graded transcriptional responses," *The EMBO Journal*, vol. 20, no. 12, pp. 3167–3176, 2001.

[20] D. T. Gillespie, "A general method for numerically simulating the stochastic time evolution of coupled chemical reactions," *Journal of Computational Physics*, vol. 22, no. 4, pp. 403–434, 1976.

[21] B. Munsky and M. Khammash, "The finite state projection algorithm for the solution of the chemical master equation," *Journal of Chemical Physics*, vol. 124, no. 4, p. 044104, 2006.

[22] A. Gupta, J. Mikelson, and M. Khammash, "A finite state projection algorithm for the stationary solution of the chemical master equation," 2017. arXiv:1704.07259.

[23] N. G. van Kampen, *Stochastic processes in physics and chemistry.* North Holland, 3rd eddition ed., 2007.

[24] Y.-B. Zhao, J. Kim, and J. P. Hespanha, "Hybrid moment computation algorithm for biochemical reaction networks," in *Proceedings of IEEE Conference on Decision and Control*, pp. 1693–1698, 2010.

[25] A. Singh and J. P. Hespanha, "Approximate moment dynamics for chemically reacting systems," *IEEE Transactions on Automatic Control*, vol. 56, no. 2, pp. 414–418, 2011.

[26] J. Kuntz, P. Thomas, and M. B. G.-B. Stan, "Rigorous bounds on the stationary distributionsof the chemical master equation via mathematical programming," 2017. arXiv:1702.05468.

[27] Y. Sakurai and Y. Hori, "A convex approach to steady state moment analysis for stochastic chemical reactions," 2017. arXiv:1704.07722.

[28] K. R. Ghusinga, C. A. Vargas-Garcia, A. Lamperski, and A. Singh, "Bounds on stationary moments in stochastic chemical kinetics," 2016. arXiv:1612.09518.

[29] S. P. Boyd and L. Vandenberghe, *Convex Optimization.* Cambridge University Press, 2004.

[30] D. T. Gillespie, "A rigorous derivation of the chemical master equation," *Physica A*, vol. 188, no. 1–3, pp. 404–425, 1992.

[31] E. T. Denisov, O. M. Sarkisov, and G. I. Likhteshtein, *Chemical kinetics: fundamentals and new developments.* Elsevier, 2003.

[32] J. A. Shohat and J. D. Tamarkin, *The problem of moments.* American Mathematical Society, 1943.

[33] H. J. Landau, *Moments in Mathematics.* American Mathematical Society, 1987.

[34] S. Boyd, L. E. Ghaoui, E. Feron, and V. Balakrishnan, *Linear matrix inequalities in system and control theory.* Society for Industrial and Applied Mathematics, 1994.

[35] B. Reeve, T. Hargest, C. Gilbert, and T. Ellis, "Predicting translation initiation rates for designing synthetic biology," *Frontiers in Bioengineering and Biotechnology*, vol. 2, no. 1, pp. 1–6, 2014.

[36] T. T. Marquez-Lago and K. Burrage, "Binomial tau-leap spatial stochastic simulation algorithm for applications in chemical kinetics," *Journal of Chemical Physics*, vol. 127, no. 10, p. 104101, 2007.

[37] D. A. Oyarzun, J.-B. Lugagne, and G.-B. V. Stan, "Noise propagation in synthetic gene circuits for metabolic control," *ACS Synthetic Biology*, vol. 4, pp. 116–125, 2015.

[38] Y. Hori and R. M. Murray, "Engineering principles of synthetic biochemical oscillators with negative cyclic feedback," in *Proceedings of the 54th IEEE Conference on Decision and Control*, pp. 584–589, 2015.

[39] Y. Dublanche, K. Michalodimitrakis, N. Kummerer, M. Foglierini, and L. Serrano, "Noise in transcription negative feedback loops: simulation and experimental analysis," *Molecular Systems Biology*, vol. 2, p. 41, 2006.

[40] A. A. Baetica, Y. Yuan, J. Goncalves, and R. M. Murray, "A stochastic framework for the design of transient and steady state behavior of biochemical reaction networks," in *Proceedings of the 54th IEEE Conference on Decision and Control*, pp. 3199–3205, 2015.

[41] C. Zechner, J. Ruess, P. Krenn, S. Pelet, M. Peter, J. Lygeros, and H. Koeppl, "Moment-based inference predicts bimodality in transient gene expression," *Proceedings of National Academy of Sciences of the United States of America*, vol. 109, no. 21, pp. 8340–8345, 2012.

[42] J. F. Sturm, "Using SeDuMi 1.02, a MATLAB toolbox for optimization over symmetric cones," *Optimization Methods and Software*, vol. 11–12, pp. 625–653, 1999.