

Document downloaded from:

<http://hdl.handle.net/10251/151291>

This paper must be cited as:

Boada-Acosta, YF.; Vignoni, A.; Picó, J. (2020). Multiobjective Identification of a Feedback Synthetic Gene Circuit. *IEEE Transactions on Control Systems Technology*. 28(1):208-223. <https://doi.org/10.1109/TCST.2018.2885694>



The final publication is available at

<https://doi.org/10.1109/TCST.2018.2885694>

Copyright Institute of Electrical and Electronics Engineers

Additional Information

© 2020 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.

Multiobjective Identification of a Feedback Synthetic Gene Circuit

Yadira Boada, *Member, IEEE*, Alejandro Vignoni^{id}, *Member, IEEE*, and Jesús Picó^{id}, *Senior Member, IEEE*

Abstract—Kinetic (i.e., dynamic) semimechanistic models based on the first principles are particularly important in systems and synthetic biology since they can explain and predict the functional behavior that emerges from the time-varying concentrations in cellular components. However, gene circuit models are nonlinear higher order ones and have a large number of parameters. In addition, experimental measurements are often scarce, and enough signal excitability for identification cannot always be achieved. These characteristics render the identification problem ill-posed, so most gene circuit models present incomplete parameter identifiability. Thus, parameter identification of typical biological models still appears as an open problem, where ensemble modeling approaches and multiobjective optimization arise as natural options. We address the problem of identifying the stochastic model of a closed-loop synthetic genetic circuit designed to minimize the gene expression noise. The model results from the feedback interaction between two subsystems. Besides incomplete parameter identifiability, the closed-loop dynamics cannot be directly identified due to the lack of enough input signal excitability. We apply a two-stage approach. First, the open-loop averaged time-course experimental data are used to identify a reduced-order stochastic model of the system direct chain. Then, closed-loop steady-state stochastic distributions are used to identify the remaining parameters in the feedback configuration. In both cases, multiobjective optimization is used to address the parameter identifiability, providing sets of parameters valid for different state-space regions. The methodology gives good identification results, provides clear guidelines on the effect of the parameters under different scenarios, and it is particularly useful for easily combining time-course population averaged and steady-state single-cell distribution experimental data.

Index Terms—Closed-loop identification, feedback synthetic gene circuit, model reduction, multiobjective optimization, parameter identification, synthetic biology.

I. INTRODUCTION

BIOLÓGICAL circuits in synthetic biology are usually modeled by means of systems of ordinary differential equations (ODEs). These ODEs describe the time evolution

Manuscript received March 15, 2018; revised October 2, 2018; accepted December 1, 2018. Date of publication January 3, 2019; date of current version December 27, 2019. Manuscript received in final form December 4, 2018. This work was supported by the European Union and Spanish Government, MINECO/AEI/FEDER under Grant DPI2017-82896-C2-1-R. The work of Y. Boada was supported by the Universitat Politècnica de València under Grant FPI/2013-3242. Recommended by Associate Editor B. Jayawardhana. (Y. Boada and A. Vignoni contributed equally to this work.) (Corresponding author: Alejandro Vignoni.)

The authors are with the Instituto de Automática e Informática Industrial, Universitat Politècnica de València, 46022 Valencia, Spain (e-mail: yaboada@upv.es; alvig2@upv.es; jpico@upv.es).

Color versions of one or more of the figures in this article are available online at <http://ieeexplore.ieee.org>.

of the amounts of the involved species, such as messenger RNA (mRNA) or proteins. Starting from a set of biochemical reactions for the circuit, dynamic balances for the biochemical species can be obtained using well-established methods, such as the mass-action kinetics formalism [1], [2]. Even for small genetic circuits, the number of metabolic intermediate biochemical species involved in the whole system dynamics is typically much larger than the few variables of interest. As a consequence, the resulting nonlinear dynamic models are high dimensional, even for small circuits, and have a very large number of unknown parameters. Most of these parameters are essentially physical magnitudes reflecting stochastic binding/unbinding reactions at the molecular level, affected by many factors within the cell. Therefore, they are uncertain by nature. In addition, the available measurements are usually very scarce, being limited to a very small number of species. Parameter identification of these models is thus infeasible in practice. To cope with this situation, a first standard approach consists of obtaining reduced-order dynamic models. Model reduction of biochemical reaction systems is carried out by, exploiting the presence of different time scales [3], [4] and the existence of system invariant species [2], [5]. Most Often, it is desired that the reduced model still allows for some degree of mechanistic description of the system. Therefore, the reduction process should yield a more amenable model for computational analysis but avoiding excessive reduction that would lead to a lack of biological relevance. In particular, the species in the reduced model must not be a combination of other species, and the resulting lumped parameters must be easy to associate with the experimental tuning knobs available in the lab. This is specially important in synthetic biology, where circuit design and experimental tuning are central. Thus, both mechanistic interpretation and easiness to translate results to the biological lab are the desired properties that the reduced-order models should commonly have in synthetic biology. However, these models still inherit many of the problems present in their parent ones.

Identification of model parameters from measurements is an established problem in control systems technology, with a long tradition of methods applicable to both linear and nonlinear systems [6], [7]. Yet, parameter estimation in nonlinear dynamic models remains a very challenging inverse problem due to its nonconvexity, ill-conditioning caused by over-parametrization, experimental measurement errors, data scarcity, and uncertainty [8], [9]. Moreover, for nonlinear models, the amount of information collected from an experiment may strongly depend on the true value of the parameters [10].

Thus, the inference of model parameters from experimental data is ill-posed under these conditions. Lack of model identifiability due to ill-posedness of the identification problem is a pervasive issue in biological systems causing parameter degeneracy [11]–[14]. Standard traditional methods based on *best-fit* models try to find a unique solution (i.e., only one value for each parameter), the best fit, relying on single objective optimization that minimizes a measure of the model performance. Yet, for most biological models, as described earlier, it is not possible to uniquely estimate parameter values from experimental data that perform well under all experimental scenarios. The best-fit solution may be good for one set of experiments and bad for others, or it may be acceptable for all the experiments but not really good for any one.

Several methods have been proposed to tackle parameter degeneracy in biological contexts. For instance, extended Kalman filters have been used in [15] for parameter estimation and model selection in biological systems with scarce noisy measurements. Relaxation procedures providing feasible sets of parameters, i.e., consistent with measurements, have been used in [16]. On the other hand, ensembles of models, i.e., sets of models with different structures and/or parameter values, have received much attention in the last years. Ensemble models combine the output of a set of models, yielding improved prediction as compared to a single model. Several approaches have been proposed. In [17], a parametric representation of the Jacobian matrix using local linear models at each point in the parameter space was used to circumvent the lack of knowledge about the structure of the system kinetics. Then, the authors used the ensemble of models to elucidate the parameter regions associated with the specific experimentally observed dynamical behaviors. A similar approach was used in [18]. Data-driven generation of ensembles of models with different structures and parameters values has been done in [19]. The final ensemble prediction is obtained from a combination using the median of the predictions of the different models. In [11], an incremental method to generate ensemble models was used to model the metabolic pathways. The models in the ensemble share the same topology and only differ in their parameter values. These models represent regions in the parameters space with equivalent model prediction error (in the statistical sense) and are obtained by exploring the parameter space. Most often, current approaches use global optimization methods to cope with nonconvexity of the objective function, e.g., genetic algorithms [20], and evolutionary computation [21]–[23]. Evolutionary computation is becoming one of the preferred optimization techniques for the large parameter estimation problems arising in systems and synthetic biology [21], [23].

Although the identification problem can most often be naturally expressed as a multiobjective problem (MOP), this approach has seldom been used [24], [25]. The use of an expanded space of model performance objectives in a multiobjective setting leads to a set of optimal solutions, the so-called Pareto set. Sets of model parameters naturally arise, akin to ensemble modeling. This alleviates the problem of parameter identifiability. Formulation of the identification problem as an MOP also allows easy integration of experimental data of different nature requiring diverse performance indexes

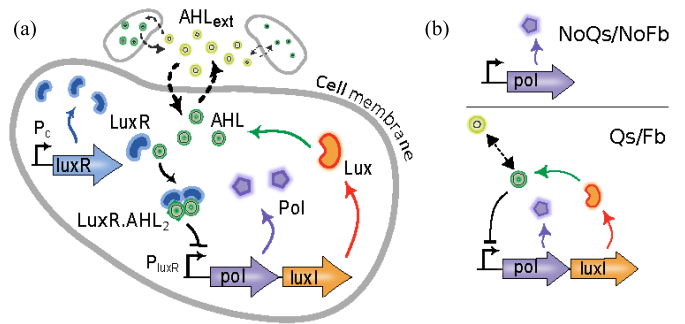


Fig. 1. (a) Representation of a cell incorporating the engineered synthetic circuit comprising cell-to-cell communication based on quorum sensing together with intracellular feedback. (b) Schematics representing the open-loop (NoQs/NoFb) and the closed-loop (Qs/Fb) circuits.

and fusion of information coming from different experiments (e.g., open-loop and closed-loop ones).

We propose a methodology based on multiobjective optimization design (MOOD) to perform parameter identification leading to ensembles of nonlinear models. The methodology uses a *global* multiobjective evolutionary algorithm (MOEA) to find the best approximation to the Pareto set of model parameters. The Pareto set together with the Pareto front regions are correlated with the experimental scenarios using hierarchical clustering and a multicriteria decision making (MCDM) strategy [26]. This allows to obtain the sets of model parameters that explain each scenario.

To show the applicability of the methodology, we performed the multiobjective optimization-based identification on a synthetic gene circuit designed to regulate noise in gene expression [27], [28]. We also show that the MOOD methodology allows easy integration and evaluation of experimental data of different nature. In our case, the closed-loop dynamics of the circuit cannot be directly identified due to lack of enough input signal excitability. Thus, we use both open-loop averaged time-course experimental data to identify a reduced-order model of the system direct chain and closed-loop steady-state population distributions to identify the remaining parameters of the stochastic model in the feedback configuration using a two-stage identification approach.

The rest of this paper is organized as follows. In Section II, the biological circuit is described and its reduced-order model derived. In Section III, we describe the proposed two stages identification methodology in detail. The results achieved are shown in Section IV. Finally, conclusions are drawn in Section V.

II. QS/FB GENE SYNTHETIC CIRCUIT

A. Description of the Genetic Circuit

Productivity in biotechnological processes decreases with the heterogeneity of the cell populations [29]. This one can be attributed to variability at the protein level caused by gene expression noise. The gene synthetic circuit depicted in Fig. 1(a) was designed in [27] and [28] to regulate the mean value of the expression of a protein of interest (PoI) while minimizing its variance. The circuit, hereafter denoted as QS/Fb, was implemented in the bacterial host *E. coli* and comprises two feedback loops.

On the one hand, the circuit uses the synthetic repressible promoter P_{luxR} designed in [30] to control transcription of the gene $luxI$ coexpressing the PoI and the protein LuxI. The promoter P_{luxR} is repressed by the transcription factor $(\text{LuxR} \cdot \text{AHL})_2$. The protein LuxR is expressed by the gene $luxR$ under control of the constitutive (i.e., constant) promoter P_c . The small signaling autoinducer molecule N-acyl-L-homoserine lactone (AHL) [31], [32] is synthesized by the LuxI protein. Proteins LuxR and AHL bind, forming the heterodimer $(\text{LuxR} \cdot \text{AHL})$, which subsequently dimerizes as the heterotetramer $(\text{LuxR} \cdot \text{AHL})_2$. This way, an intracellular negative feedback control is effectively implemented on the expression of LuxI.

On the other hand, the signaling molecule AHL induces a cell-to-cell communication mechanism via *quorum sensing* [33]. This molecule passively diffuses across the cellular membrane. Therefore, a continuous flow of AHL across the cell membrane appears as a function of the intracellular an extracellular AHL concentrations. Since production of PoI and LuxI are linked, cells producing less PoI than the average production of the cell population will have a positive inflow of extracellular AHL and the other way round. In this way, an external feedback loop inducing consensus at the cell population level is implemented.

If the gene $luxI$ is removed, AHL it is not produced. This results in the open-loop circuit NoQS/NoFb that only implements the constitutive expression of LuxR and a repressible expression of the PoI [see Fig. 1(b)]. In this case, the level of PoI expression can be repressed at different levels by adding varying concentrations of external AHL into the culture medium.

In Section II-D, we derive a stochastic reduced model for the closed-loop QS/Fb circuit. This model can be used to represent also the open loop (NoQs/NoFb) by setting the AHL synthesis rate to 0, which represents the absence of the $luxI$ gene. First, we derive a deterministic model from the set of relevant biochemical reactions. We will obtain a reduced-order model applying the *quasisteady-state* (QSS) approximation and considering *invariant species*. From this reduced-order model, we will obtain an equivalent set of biochemical pseudoreactions that will eventually be used to derive a stochastic model using the chemical Langevin equation (CLE) formalism.

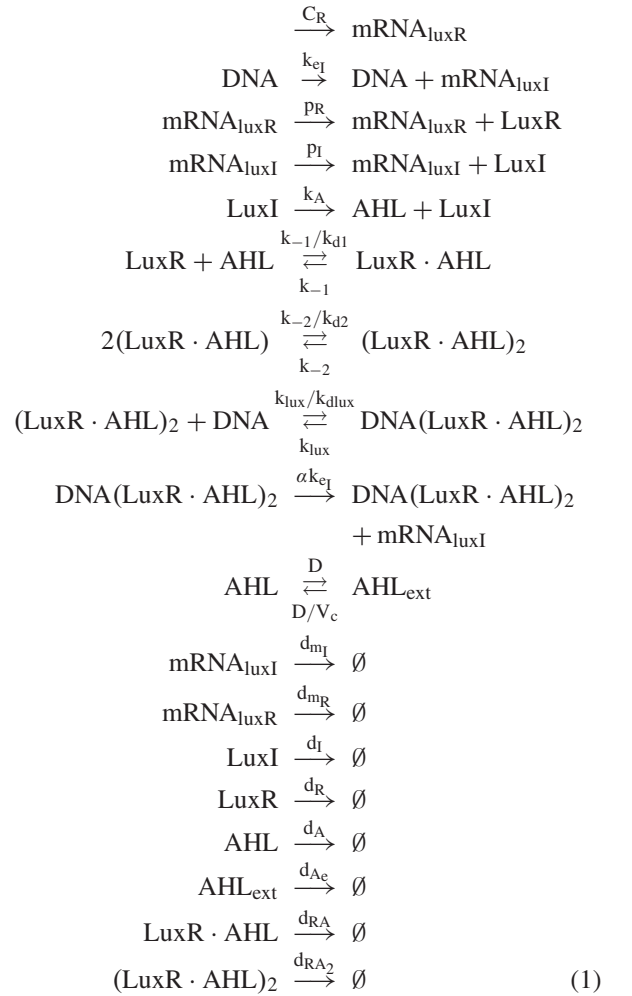
B. Deterministic Model

For the above-mentioned circuit, we consider the main biochemical reactions: transcription and translation of the genes involved, the hetero- and homodimerization reactions, and the diffusion of the inducer through the cell membrane. A set of reactions is obtained under the following assumptions:

- 1) We consider the cell contains enough resources (free RNA polymerase and ribosomes). This way, there is no need to consider the binding/unbinding reactions of RNA polymerase and ribosomes to DNA and mRNA, respectively.
- 2) Transcription of genes $luxI$ and $luxR$ is irreversible.
- 3) For the repressible promoter P_{luxR} , we consider promoter leakiness (i.e., transcription even in saturating presence of the repressor). The effective maximum transcription

rate of $luxI$ (without repression) is k_{e1} , and its basal transcription (leakage) is αk_{e1} with $\alpha < 1$.

- 4) There may be several copies of both genes $luxI$ and $luxR$ in the cell, e.g., by introducing them in separate *plasmids* with different plasmid copy numbers [34].
- 5) Translation is not a simple process [34]. We model it as an irreversible reaction with an average transcription rate accounting for the fact that binding of ribosomes to the ribosome binding site (RBS) is indeed reversible, and several ribosomes may translate a single mRNA molecule simultaneously.
- 6) Binding of LuxR and AHL, and their dimerization to form $(\text{LuxR} \cdot \text{AHL})_2$ are both reversible reactions.
- 7) The diffusion process of the intracellular (AHL) and extracellular (AHL_{ext}) inducer across the cell membrane is modeled as a pseudoreaction.



where the empty set \emptyset represents degradation of biochemical species. We denoted DNA as the free promoter of gene $luxI$, $\text{mRNA}_{\text{luxI}}$ and $\text{mRNA}_{\text{luxR}}$ are the mRNA transcribed from the genes $luxI$ and $luxR$, respectively, and LuxI and LuxR are their corresponding translated proteins. AHL is the intracellular inducer and AHL_{ext} the extracellular one, and $V_c = V_{\text{ext}}/V_{\text{cell}}$ is the ratio between the cell and the culture volumes.

The dynamic balance equations for each species in the above-mentioned reactions can be derived using the mass action kinetics formalism [1]. The resulting models assume

TABLE I
STATES IN THE COMPLETE DETERMINISTIC QS/FB MODEL

Variable	Biochemical species	Unit
n_1	mRNA of gene <i>luxI</i>	molecules
n_2	mRNA of gene <i>luxR</i>	molecules
n_3	Protein LuxI	molecules
n_4	Protein LuxR	molecules
n_5	Monomer (LuxR · AHL)	molecules
n_6	Dimer (LuxR · AHL) ₂	molecules
n_7	Free promoter DNA	molecules
n_8	Bound promoter DNA(LuxR · AHL) ₂	molecules
n_9	Intracellular inducer AHL	molecules
n_{10}	Extracellular inducer AHL _{ext}	molecules

that the amount of species transformed by the reactions depend solely on the current amount of species, the rates at which these reactions proceed, and the stoichiometry of the reactions [2]. We take into account the dilution effect caused by cells growth, assuming a constant specific growth rate μ included in the respective degradation rate parameter of each species. We assume that the number of copies of the genes *luxR* and *luxI* keeps constant over time. This is indeed the case if the genes are chromosomal ones. In case the genes are located in plasmids, we can assume that at each cell division, plasmids are first duplicated and, then, half of them will be inherited by each of the offspring cells. This is a valid approximation if we assume that we model the average cell.

The resulting deterministic dynamic model is given by the set of equations (2)–(11) representing the dynamics of each species inside the i th cell in a population of N cells. The model has $9N + 1$ states and 22 parameters, assuming identical cells. Table I describes each state in the model

$$\dot{n}_1^i = k_{e1}n_7^i + \alpha k_{e1}n_8^i - d_{m1}n_1^i \quad (2)$$

$$\dot{n}_2^i = C_R - d_{mR}n_2^i \quad (3)$$

$$\dot{n}_3^i = p_I n_1^i - d_I n_3^i \quad (4)$$

$$\dot{n}_4^i = p_R n_2^i + k_{-1}n_5^i - d_R n_4^i - \frac{k_{-1}}{k_{d1}}n_9^i n_4^i \quad (5)$$

$$\dot{n}_5^i = 2k_{-2}n_6^i + \frac{k_{-1}}{k_{d1}}n_9^i n_4^i + \left(-k_{-1} - d_{RA} - 2\frac{k_{-2}}{k_{d2}}n_5^i\right)n_5^i \quad (6)$$

$$\dot{n}_6^i = k_{lux}n_8^i + \frac{k_{-2}}{k_{d2}}n_5^i + \left(-k_{-2} - d_{RA2} - \frac{k_{lux}}{k_{dlux}}n_7^i\right)n_6^i \quad (7)$$

$$\dot{n}_7^i = k_{lux}n_8^i - \frac{k_{lux}}{k_{dlux}}n_6^i n_7^i \quad (8)$$

$$\dot{n}_8^i = -k_{lux}n_8^i + \frac{k_{lux}}{k_{dlux}}n_6^i n_7^i \quad (9)$$

$$\dot{n}_9^i = D \left(\frac{n_{10}}{V_c} - n_9^i \right) + k_A n_3^i + k_{-1}n_5^i - \left(\frac{k_{-1}}{k_{d1}}n_4^i + d_A \right) n_9^i \quad (10)$$

$$\dot{n}_{10} = D \left(-\frac{N n_{10}}{V_c} + \sum_{i=1}^N n_9^i \right) - d_{Ae} n_{10}. \quad (11)$$

Note that we have expressed diffusion as a biochemical reaction in (1) when it, actually, is a physical process that we modeled using a lumped approximation of the Fick's law [34], [35]. In addition, the extracellular inducer AHL_{ext} is not subject to dilution by growth rate. We assumed that the external inducer is introduced as a *bolus* injection. That

is, at time $t = 0$, an amount of AHL_{ext} is injected in the culture. This value is taken as the initial condition $n_{10}(0)$ for its corresponding dynamic balance.

C. Model Reduction

Using the model (2)–(11), we get a reduced-order model by applying the QSS approximation on the fast chemical reactions and taking into account invariant species derived from conservation laws. We aim at obtaining a reduced model more amenable for computational analysis but avoiding excessive reduction that would lead to a lack of biological relevance. On the one hand, the species we obtain in the reduced model are not the lumped ones. Reduced models accounting for total mRNA and total transcription factor have been proposed to match modeled species with measurable ones [36]. In our case, we explicitly modeled bound and unbound forms of the transcription factor, but the model accounts for the total LuxI protein. For our circuit, this is a good proxy for the amount of PoI if both are coexpressed, and transcriptional noise dominates. In the best case, when the PoI is in self-cleavable tandem fusion with LuxI, both will express in 1:1 stoichiometric ratio [37]. On the other hand, the resulting parameters in the reduced model, even if being lumped ones, are easy to associate with the tuning knobs available in the wet-lab implementation (relevant cases [38]), and their values are amenable to be obtained experimentally.

Conservation laws can be inferred from simple inspection in the model (2)–(11). Notice the sum of equation (8) representing the variation of free promoter plus equation (9) representing the variation of dimer bound to the promoter DNA(LuxR · AHL)₂ is 0. This implies that the sum of free and bound promoter is constant and equal to the (*plasmid*) copy number P_N of the gene *luxI*

$$\dot{n}_7^i + \dot{n}_8^i = 0 \rightsquigarrow n_7^i + n_8^i = P_N. \quad (12)$$

Next, we considered that the binding/unbinding reactions of the dimer (LuxR · AHL)₂ to the gene promoter DNA proceed much faster than translation and mRNA degradation, so they can be assumed to be at QSS. This is also equivalent to consider that k_{lux} is large enough so that using (12), we can approximate

$$\begin{aligned} \frac{1}{k_{lux}}\dot{n}_7^i = 0 &\rightsquigarrow n_7^i = P_N \left(\frac{k_{dlux}}{k_{dlux} + n_6^i} \right) \\ \frac{1}{k_{lux}}\dot{n}_8^i = 0 &\rightsquigarrow n_8^i = P_N \left(\frac{n_6^i}{k_{dlux} + n_6^i} \right). \end{aligned} \quad (13)$$

We also assume that mRNA is produced and degraded much faster than proteins. Hence, considering mRNA at QSS, we get algebraic expressions for mRNA_{luxI} (n_1^i) and mRNA_{luxR} (n_2^i)

$$\dot{n}_1^i = 0 \rightsquigarrow n_1^i = \frac{k_{e1}}{d_{m1}}(n_7^i + \alpha n_8^i) \quad (14)$$

and

$$\dot{n}_2^i = 0 \rightsquigarrow n_2^i = \frac{C_R}{d_{mR}}. \quad (15)$$

TABLE II
STATES IN THE DETERMINISTIC REDUCED MODEL

Variable	Biochemical species	Unit
n_1	Protein LuxI	molecules
n_2	Protein LuxR	molecules
n_3	Dimer (LuxR · AHL) ₂	molecules
n_4	Internal autoinducer AHL	molecules
n_5	External autoinducer AHL _{ext}	molecules
n_6	Monomer (LuxR · AHL)	molecules

From (13) and (14), replacing in (4), we have

$$\dot{n}_3^i = \frac{P_N k_{e1} \text{PI}}{d_{m1}} \left(\frac{k_{dlux} + \alpha n_6^i}{k_{dlux} + n_6^i} \right) - d_1 n_3^i \quad (16)$$

and from (13) and (15), replacing in (5), we have

$$\dot{n}_4^i = \frac{C_{RPR}}{d_{mR}} + k_{-1} n_5^i - \left(\frac{k_{-1}}{k_{d1}} n_9^i + d_R \right) n_4^i. \quad (17)$$

In addition, using (13) in (7), we obtain

$$\dot{n}_6^i = \frac{k_{-2}}{k_{d2}} (n_5^i)^2 - (k_{-2} + d_{RA2}) n_6^i. \quad (18)$$

Finally, we assumed that the monomer production is large as compared to that of the dimer, so that it can be considered at QSS, i.e., $\dot{n}_5^i = 0$ in (6). The resulting expression for n_5^i can be replaced in (5), (7), and (10).

All these approximations, after convenient renaming of the state variables as shown in Table II, lead to the reduced-order model (19)–(23) for the i th cell in a population of N cells

$$\dot{n}_1^i = \frac{C_1 \text{PI}}{d_{m1}} \left(\frac{k_{dlux} + \alpha n_3^i}{k_{dlux} + n_3^i} \right) - d_1 n_1^i \quad (19)$$

$$\dot{n}_2^i = \frac{C_{RPR}}{d_{mR}} + k_{-1} n_6^i - \left(\frac{k_{-1}}{k_{d1}} n_4^i + d_R \right) n_2^i \quad (20)$$

$$\dot{n}_3^i = \frac{k_{-2}}{k_{d2}} (n_6^i)^2 - (k_{-2} + d_{RA2}) n_3^i \quad (21)$$

$$\dot{n}_4^i = k_{-1} n_6^i + k_A n_1^i + D \left(\frac{n_5}{V_c} - n_4^i \right) - \left(\frac{k_{-1}}{k_{d1}} n_2^i + d_A \right) n_4^i \quad (22)$$

$$\dot{n}_5 = D \left(-\frac{N n_5}{V_c} + \sum_{i=1}^N n_4^i \right) - d_{A_e} n_5 \quad (23)$$

with

$$n_6^i = K \left[\sqrt{\frac{8 k_{-2} (2 k_{-2} k_{d1} n_3^i + k_{-1} n_2^i n_4^i)}{k_{d1} k_{d2} (d_{RA} + k_{-1})^2}} + 1 - 1 \right] \quad (24)$$

where C_1 is the plasmid copy number of gene *luxI* times its effective transcription rate. The model requires 18 parameters, as $(C_1 \text{PI}/d_{m1})$ and (C_{RPR}/d_{mR}) represent lumped effective expression rates, and the constant $K = (k_{d2}(d_{RA} + k_{-1}))/4k_2$. It has $4N + 1$ states and N algebraic equations. The reduction by $5N$ states makes this model more computationally efficient than the original one.

Notice the first term on the right-hand side of (19) is a Hill-like function with the Hill exponent equal to $n = 1$ [39], which together with the monomer algebraic equation (24) represents the transcription factor regulatory effect (repression

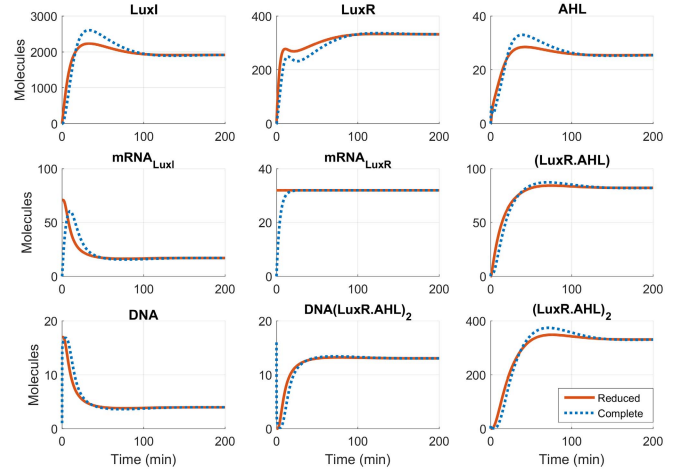


Fig. 2. Comparison of models. A 200-min-long simulation carried out over a single cell ($N = 1$) for both the reduced (solid line) and the complete model (dashed line). Both simulations used the same initial conditions, same parameter values, and same discretization step size ($\delta t = 10^{-3}$ s). The species eliminated by the reduction process [$\text{mRNA}_{\text{LuxI}}$, $\text{mRNA}_{\text{LuxR}}$, DNA, $\text{DNA}(\text{LuxR} \cdot \text{AHL})_2$, and $(\text{LuxR} \cdot \text{AHL})$] were calculated algebraically from the remaining species. As $N = 1$, the amount of molecules of AHL and AHL_{ext} is similar, so the AHL_{ext} plot was omitted in this figure.

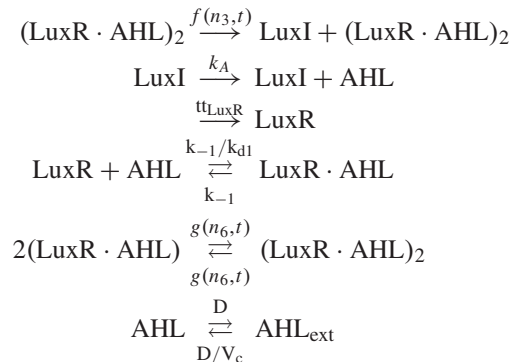
in our case) over the expression of protein LuxI. These expressions are equivalent to the ones that can be obtained using [36] for the multimer-dominant case.

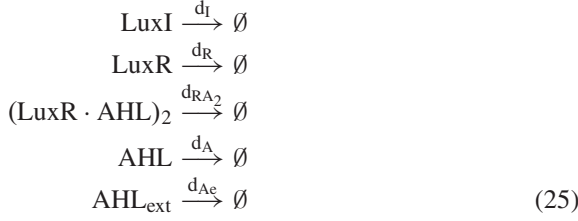
To compare the complete model with the reduced order one, we performed a series of simulation experiments. Fig. 2 shows some of the results for both the complete and the reduced models. The biochemical species are plotted in Fig. 2 for the reduced model (solid line) and the full one (dashed line).

The agreement between the results of both models was good enough for our purposes, without requiring any *ad hoc* adjustment. From a qualitative point of view, the transient regimes of both models are similar for all species. The length of the transients and the steady-state values coincide in both models.

D. Stochastic Model

To model gene expression intrinsic noise, we have derived a stochastic CLE model. Here, the mean of each species' trajectory corresponds to the trajectory of the deterministic reduced model (19)–(23). To this end, we first considered the following equivalent set of pseudoreactions for the i th cell drawn from the deterministic model:





where

$$f(n_3, t) \triangleq \frac{C_{\text{IPI}}}{d_{\text{mI}}} \left(\frac{k_{\text{dlux}} + \alpha_I n_3^i}{k_{\text{dlux}} + n_3^i} \right)$$

the Hill-like function associated with LuxI expression, $g(n_6, t)$ corresponds to the dimerization reflected in (24), $\text{tt}_{\text{LuxR}} = (C_{\text{RPR}}/d_{\text{mR}})$ represents the transcription–translation activity of *luxR*, and \emptyset denotes species degradation. From these reactions, it is possible to obtain the associated CLE [52]

$$d\mathbf{n}(t) = \mathbf{S} \cdot \mathbf{a}(\mathbf{n}(t))dt + \mathbf{S} \cdot \sqrt{\mathbf{a}(\mathbf{n}(t))} \cdot d\mathcal{W}(t) \tag{26}$$

where $\mathbf{n}(t) = [\mathbf{n}(t)^i, \dots, \mathbf{n}(t)^N, n_5]^T$ are the number of molecules of each species in the population, $\mathbf{a}(\mathbf{n}(t))$ the reaction propensities, and $\mathcal{W}_{(\mathbf{JN}+1) \times (\mathbf{JN}+1)}$ is a diagonal matrix of independent scalar Brownian motions [52] with $J = 13$ the number of reactions for the i th cell.

The stoichiometry matrix \mathbf{S} , whose elements are the stoichiometry submatrices for every cell \mathbf{S}_{cell} and the external stoichiometry \mathbf{S}_{ext} , has the structure

$$\mathbf{S} = \left[\begin{array}{c|c} \mathbf{S}_{\text{cell}} \otimes \mathbf{I}_N & \mathbf{0}_{4N \times 1} \\ \hline \mathbf{S}_{\text{ext}} \otimes \mathbf{1}_{1 \times N} & -1 \end{array} \right] \tag{27}$$

where \otimes is the Kronecker product, \mathbf{I}_N is the identity matrix of dimension $N \times N$, $\mathbf{0}_{4N \times 1}$ and $\mathbf{1}_{1 \times N}$ are the vectors of zeroes and ones, respectively. The coefficients in the stoichiometry matrices \mathbf{S}_{cell} and \mathbf{S}_{ext} , obtained from the set of pseudoreactions (25), are

$$\begin{aligned}
\mathbf{S}_{\text{cell}} &= \begin{bmatrix} 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & -1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 & 1 & -1 & -1 & 1 \end{bmatrix} \\
\mathbf{S}_{\text{ext}} &= [0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 1 \ -1].
\end{aligned}$$

The term $\mathbf{a}(\mathbf{n})$ in (26) is the associated vector of reaction propensities for the whole population of cells with

$$\mathbf{a}(\mathbf{n}) = \begin{bmatrix} \mathbf{a}(\mathbf{n})^1 \\ \mathbf{a}(\mathbf{n})^2 \\ \vdots \\ \mathbf{a}(\mathbf{n})^N \\ d_{A_e} n_5 \end{bmatrix}, \quad \mathbf{a}(\mathbf{n})^i = \begin{bmatrix} f(n_3^i, t) \\ d_1 n_1^i \\ \frac{C_{\text{RPR}}}{d_{\text{mR}}} \\ k_{-1} n_6^i \\ \frac{k_{-1}}{k_{d1}} n_2^i n_4^i \\ d_{\text{R}} n_2^i \\ \frac{k_{-2}}{k_{d2}} (n_6^i)^2 \\ k_{-2} n_3^i \\ d_{\text{RA}_2} n_3^i \\ k_{\text{A}} n_1^i \\ d_{\text{A}} n_4^i \\ D n_4^i \\ D/V_c n_5 \end{bmatrix}.$$

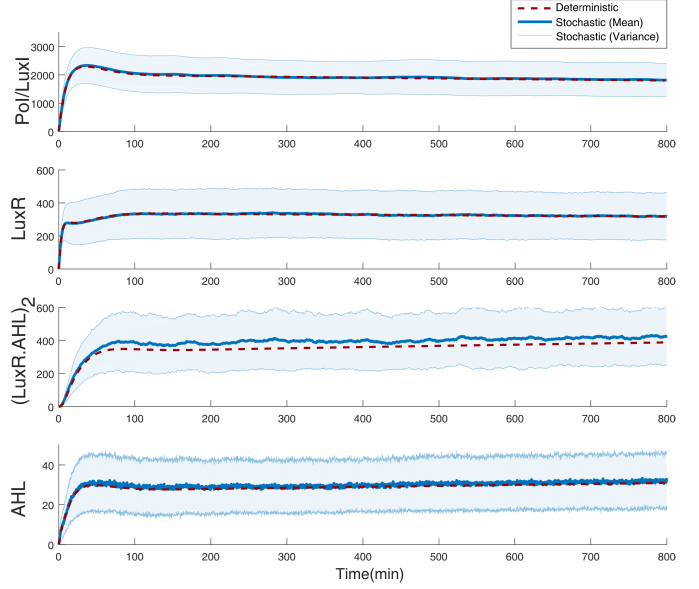


Fig. 3. Comparison of the deterministic and stochastic models. A 800-min-long simulation of the deterministic reduced model is compared with a simulation of the stochastic model carried out over a population of cells ($N = 240$). The deterministic trajectory (dashed red line) is plotted along the stochastic mean (solid blue line) and standard deviation (cyan shadow).

For the computational analysis, we used the Euler–Maruyama discretization (28) of the CLE stochastic model (26)

$$\mathbf{n}(t + \delta t) = \mathbf{n}(t) + \mathbf{S} \cdot \mathbf{a}(\mathbf{n})\delta t + \mathbf{S} \cdot \mathcal{N} \cdot \sqrt{\mathbf{a}(\mathbf{n})}\sqrt{\delta t} \tag{28}$$

where $\mathcal{N}_{(\mathbf{JN}+1) \times (\mathbf{JN}+1)}$, with $J = 13$, is a diagonal matrix of continuous normal random variables with zero mean and unit variance.

Note that, we used the lumped propensity functions derived from the reduced model, like the $f(n_3^i, t)$ Hill-like function associated with LuxI repression. This approach has already been used in [53]. We validated it for our model (see Appendix I) by simulating the pseudoreaction associated with $f(n_3^i, t)$ using the CLE approach and comparing the result with the one obtained by simulating the expanded set of corresponding original reactions using the Gillespie’s direct method Stochastic Simulation Algorithm (SSA). For the rest of the analysis, we performed stochastic simulations (28) using $N = 240$ cells [28] with a total time of 400 min using $\delta t = 25 \times 10^{-4}$ s. Extrinsic noise was modeled by randomizing the values of the model parameters [54], [55], an approach that can easily be integrated within the CLE framework. We assumed a normal distribution to generate the model parameters for each i th cell in the population. Under these conditions, we compared the output of the deterministic and the stochastic models. The results are shown in Fig. 3. The mean expression of the species in the stochastic model is very close to their corresponding trajectory in the deterministic one.

E. Model Parameters

Table III describes the model parameters for both the Qs/Fb and the NoQs/NoFb circuits, with 24 and 22 parameters,

TABLE III
PARAMETERS OF THE GENE CIRCUIT MODEL

Fixed Parameters	Description	Value	Reference
C_I	Plasmid copy number times <i>luxI</i> transcription rate	$17.5 \text{ molecules} \cdot \text{min}^{-1}$	estimated
C_R	Plasmid copy number times <i>luxR</i> transcription rate	$7.9 \text{ molecules} \cdot \text{min}^{-1}$	estimated
α	Basal expression of <i>luxI</i>	0.01	estimated
k_{-1}	Dissociation rate of $(\text{LuxR} \cdot \text{AHL})$	10 min^{-1}	[40]
k_{-2}	Dissociation rate of dimer $(\text{LuxR} \cdot \text{AHL})_2$	1 min^{-1}	[28]
k_{d1}	Dissociation constant of $(\text{LuxR} \cdot \text{AHL})$	100 molecules	[41]
k_{d2}	Dissociation constant of $(\text{LuxR} \cdot \text{AHL})_2$	20 molecules	[42]
d_R	Degradation rate of LuxR	0.2 min^{-1}	[43], and refs. therein
d_A	Degradation rate of AHL	0.057 min^{-1}	[33], [44]
d_{A_e}	Degradation rate AHL_{ext}	0.04 min^{-1}	[31], [33], [44]
d_{RA}	Degradation rate of $(\text{LuxR} \cdot \text{AHL})$	0.156 min^{-1}	[45] and refs. therein
d_{RA_2}	Degradation rate of $(\text{LuxR} \cdot \text{AHL})_2$	0.017 min^{-1}	[28]
d_{m_I}, d_{m_R}	Degradation rate of $\text{mRNA}_{\text{LuxI}}$ and $\text{mRNA}_{\text{LuxR}}$	0.247 min^{-1}	[46], [47]
D	Diffusion rate of AHL through the cell membrane	2 min^{-1}	[35], [48]
V_{cell}	Typical volume of <i>E. coli</i> .	$1.1 \cdot 10^{-9} \mu\text{L}$	[49]
V_{ext}	Typical volume of a plate reader well	$200 \mu\text{L}$	
Unknown Parameters	Description	Range of values	Reference
p_I	Translation rate of $\text{mRNA}_{\text{LuxI}}$	$[0.1 \ 10] \text{ min}^{-1}$	[39], [49]
d_I	Degradation rate of LuxI	$[0.005 \ 0.1] \text{ min}^{-1}$	[50], [51]
$k_{d\text{lux}}$	Dissociation constant of $(\text{LuxR} \cdot \text{AHL})_2$ to the P_{luxR} promoter	$[1 \ 1000] \text{ molecules}$	[45] and refs. therein
p_R	Translation rate of $\text{mRNA}_{\text{LuxR}}$	$[0.1 \ 10] \text{ min}^{-1}$	[39], [49]
K_{pr}	Fluorescence to number of molecules ratio (plate reader gain)	$[0.04 \ 5]$	
k_A	Synthesis rate of AHL by LuxI (Qs/Fb circuit only)	$[0.005 \ 0.1] \text{ min}^{-1}$	[27]
K_{fc}	Fluorescence to number of molecules ratio (flow cytometer gain)	$[0.04 \ 5]$	

respectively. Recall the model (28) also represents the NoQs/NoFb CLE-based model if the synthesis rate is set to $k_A = 0$. Out of these parameters, 17 were obtained from the literature and were kept fixed (top of Table III). Some of these parameters were calculated as follows.

- 1) C_I is the LuxI transcription rate times the LuxI plasmid copy number. The typical transcription rate in *E. coli* is $\approx 600\text{--}6000 \text{ bp/min}$ [34]. The LuxI length is 582 bp [56]. The LuxI transcription rate is $(600 \text{ bp/min})/582 \text{ bp} = 1.03 \text{ min}^{-1}$; using the plasmid vector pBR322 with [15, 20] copies/cell, we estimated $C_I = 17 * 61.03 = 17.5 \text{ molecules} \cdot \text{min}^{-1}$.
- 2) C_R is the transcription rate obtained as before times the LuxI plasmid copy number. We used the plasmid vector pACYC184 with 10 copies/cell, the minimum transcription rate 600 bp/min, and the LuxR length 756 bp [56]. Hence, the plasmid copy number times LuxR transcription rate is $C_R = (10 * 600 \text{ bp/min})/756 \text{ bp} = 7.9 \text{ molecules} \cdot \text{min}^{-1}$.
- 3) The degradation rates $d_{m_I}, d_{m_R}, d_I, d_R, d_A,$ and d_{RA} include the dilution effect due to cell growth. We considered a specific growth rate $\mu_{\text{spe}} = 0.011 \text{ min}^{-1}$ corresponding to a cell doubling time of 63 min obtained experimentally.
- 4) The degradation rate $d_{RA_2} = 0.017 \text{ min}^{-1}$ of the transcription factor $(\text{LuxR} \cdot \text{AHL})_2$ only depends on the specific growth rate μ_{spe} , assuming $(\text{LuxR} \cdot \text{AHL})_2$ is much more stable than the other species in the system [57], [45].

The remaining seven parameters are easy to associate with the experimental tuning knobs available in the lab and will be estimated in this paper. For them, initial range intervals

were set, taking into account the plausible biological values from the literature.

III. MULTIOBJECTIVE OPTIMIZATION AND PARAMETER IDENTIFICATION

A. Two-Stage Identification Approach

In our circuit, the only available experimental data are limited to two classes. These are typically available in a standard laboratory: 1) time-course averaged measurements of cells growth and total production of the PoI in a culture and 2) steady-state distributions of protein production across the cells population. In this last case, time snapshots can be obtained with low frequency using a lab-intensive and costly approximate procedure.

Recall that the goal of the circuit is to minimize output variance for a given mean. To this end, it integrates two feedback loops to achieve robustness with respect to the external perturbations. This makes the identification problem harder due to the practical lack of excitability of the input. Indeed, time-course data of the closed-loop circuit output, both the mean and the distribution across the cell population, show very little variation, even for the large perturbations we can introduce experimentally. As seen in Fig. 4(a), even after a large perturbation of 50 nM AHL introduced in the culture medium, the closed-loop circuit output (population distribution for different times after the induction) remains almost unaltered. On the contrary, the open-loop circuit can be easily excited [see Fig. 4(b)]. Thus, we resort to a two-stage identification procedure, where we first identify the open-loop circuit using easy-to-get experimental time-course bulk measurements, i.e., the mean of the cells population behavior.

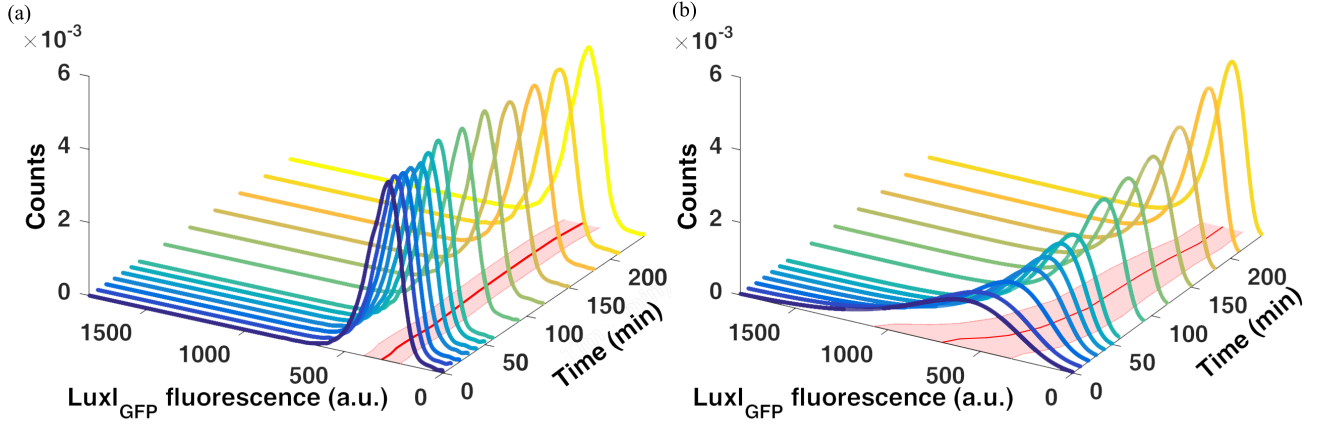


Fig. 4. Population distribution snapshots taken with a flow cytometer for some time points after AHL induction with 50 nM. (a) Experimental data for the closed-loop circuit. (b) Experimental data for the open-loop circuit.

Therefore, the relevant model for this first stage is the deterministic model given by (19)–(23).

In the second stage, once we have estimated the open-loop circuit parameters, we extend the identified open-loop dynamics to get the full closed-loop circuit. Here, the goal is to estimate only the remaining parameters k_A , K_{fc} . The available data are the steady-state population distribution of the PoI. Therefore, we use the stochastic model (28) derived from the deterministic one to predict such distribution.

Notice that the single-stage optimization could be implemented. Yet, even if parameters search space dimension increases only by two (the open-loop parameters plus the two closed-loop ones), the computationally expensive stochastic simulations will have to be evaluated in each iteration for the whole set of open-loop parameters plus the two closed-loop ones. The two-stage approach appreciably reduces the number of stochastic evaluations required.

B. Multiobjective Optimization Approach

As mentioned before, parameter degeneracy is often the price to pay for having models of biological gene networks and circuits that allow for a relevant degree of mechanistic biological interpretation and have parameters that can be associated with available experimental lab tuning knobs. To alleviate degeneracy, the use of an expanded space of model performance objectives in a multiobjective setting leads to a set of optimal solutions that can be clustered. In this way, sets of model parameters naturally arise, each valid to describe an experimental scenario.

Consider a dynamic model described by ODEs of the following general form:

$$\Sigma: \begin{cases} \dot{\mathbf{x}}(t) = \mathbf{f}(\mathbf{x}(t), \mathbf{u}(t), \theta) \\ \mathbf{y} = \mathbf{g}(\mathbf{x}(t), \theta) \\ x_0 = \mathbf{x}(t_0), \theta \end{cases} \quad (29)$$

where $\mathbf{x} \in \mathbf{X} \subset \mathbb{R}^n$ is the state vector, $\mathbf{u} \in \mathbf{U} \subset \mathbb{R}^r$ is a control input, $\mathbf{f} : \mathbb{R}^{n \times r \times q} \rightarrow \mathbb{R}^n$ a vector field defined in \mathbf{X} , $\mathbf{y} \in \mathbf{Y} \subset \mathbb{R}^m$ the measurable output or observable vector, and $\theta \in \mathbb{R}^q$ is a real-valued vector of parameters. Each input/output experiment performed in the system Σ will provide a particular value $\hat{\mathbf{y}}$ of the observables vector \mathbf{y} . The parameter identification problem consists on finding the

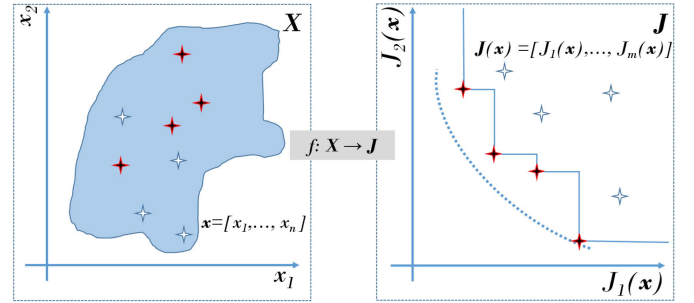


Fig. 5. Pareto optimality and dominance concepts. A Pareto front (dotted line in objective space J) is approximated by a set of solutions (depicted with stars) selected from the feasible decision space X . The red solutions are nondominated solutions in the set, and therefore, they are used to build a Pareto front approximation (solid line). The remainder solutions are dominated solutions. Adapted from [24].

best value of θ , such that some cost function $J(\theta)$ related to the model prediction error $\mathbf{y}(\theta) - \hat{\mathbf{y}}$ is minimized.

As referred in [58], an MOP can be stated as follows:

$$\min_{\theta} \mathbf{J}(\theta) = [J_1(\theta), \dots, J_m(\theta)] \quad (30)$$

$$\text{s.t. } \mathbf{K}(\theta) \leq 0$$

$$\mathbf{L}(\theta) = 0$$

$$\underline{\theta}_i \leq \theta_i \leq \bar{\theta}_i \quad \forall i = [1, \dots, n] \quad (31)$$

where $\theta = [\theta_1, \theta_2, \dots, \theta_n]$ is defined as the decision vector, $\mathbf{J}(\theta)$ is the objective vector, $\mathbf{K}(\theta)$ and $\mathbf{L}(\theta)$ are the inequality and equality constraint vectors, respectively, and $\underline{\theta}_i$ and $\bar{\theta}_i$ are the lower and upper bounds in the decision space Θ . The solution of this problem, instead of one unique solution as in the case of single objective optimization, is a set of solutions defined as the *Pareto Set* Θ_P . Each solution in this set corresponds to an optimal objective vector in the *Pareto Front* \mathbf{J}_P . All solutions in the *Pareto Set* are Pareto-optimal nondominated solutions and differ from each other in the tradeoff of objectives each one represents.

Definition 1 (Pareto Optimality [24]): An objective vector $\mathbf{J}(\theta^1)$ is Pareto optimal if there is not another objective vector $\mathbf{J}(\theta^2)$, such that $\mathbf{J}_i(\theta^2) \leq \mathbf{J}_i(\theta^1)$ for all $i \in [1, 2, \dots, m]$ and $\mathbf{J}_j(\theta^2) < \mathbf{J}_j(\theta^1)$ for at least on j , $j \in [1, 2, \dots, m]$. See Fig. 5.

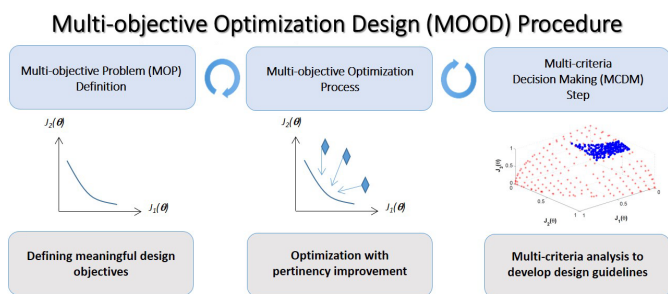


Fig. 6. Three steps of the MOOD procedure with the specific problems to solve.

Definition 2 (Dominance [59]): An objective vector $\mathbf{J}(\theta^1)$ is dominated by another objective vector $\mathbf{J}(\theta^2)$ if and only if $\mathbf{J}_i(\theta^2) \leq \mathbf{J}_i(\theta^1)$ for all $i \in [1, 2, \dots, m]$ and $\mathbf{J}_j(\theta^2) < \mathbf{J}_j(\theta^1)$ for at least on j , $j \in [1, 2, \dots, m]$. This is denoted as $\mathbf{J}(\theta^2) \preceq \mathbf{J}(\theta^1)$. See Fig. 5.

Notice that, in general, the *Pareto Front* \mathbf{J}_P is unknown, and it is only possible to rely on an approximation of the *Pareto Front* \mathbf{J}_P^* and of its corresponding approximation of the *Pareto Set* θ_P^* .

In order to successfully implement this approach, the following three fundamental steps are required [58]:

- 1) the MOP definition;
- 2) the optimization process;
- 3) and the MCDM stage.

This overall MOOD procedure, depicted in Fig. 6, enables to analyze the existing tradeoffs between the objectives and, accordingly, choose the preferable solution [24].

1) Multiobjective Problem Definition: At this point, the objective vector function has to be defined in order to solve our parameter identification problem. For the case of our two-stage identification problem, we must define the objective vector functions related to the minimization of the error between, respectively, the following:

- 1) time-course plate reader experimental observations and open-loop model predictions;
- 2) steady-state flow cytometry experimental observations and closed-loop model predictions.

These will be detailed later on.

2) Multiobjective Optimization Process: The multiobjective optimization process finds the best parameters θ_p^* that produce the best Pareto front approximation \mathbf{J}_P^* . For problems with a large number of decision variables, as it is our case, it is more efficient to use an appropriate multiobjective optimization algorithm to approximate this solution.

In this paper, we used the sp-MODE¹ algorithm, an MOEA based on differential evolution, which uses a spherical pruning to approximate the Pareto front. sp-MODE improves: 1) the convergence by using an external file to store solutions and include them in the evolutionary process; 2) spreading by using the spherical pruning mechanism [60]; and 3) pertinency of solutions by means of a basic bound mechanism in the objective space, as described in [61].

3) Multicriteria Decision Making Stage: The selection of the preferable solution according to designer's criteria takes place in an *a posteriori* multicriteria analysis of the Pareto front approximation. Tools that simplify the visualization and the analysis of the tradeoff among competing objectives are very useful. Such visualization and analysis is not a trivial task when the number of objectives is larger than three, and/or the number of decision variables in the Pareto set is large. Desirable characteristics to analyze and visualize the results include the following.

- 1) Enabling comparison of design alternatives (analyze different solutions).
- 2) Enabling comparison of design concepts (analyze different Pareto front approximations).
- 3) *Completeness*: All relevant information should be contained in the visualization.
- 4) *Persistence*: All the relevant information should be retained in the designer's mind.
- 5) *Simplicity*: The visualization should be easily understandable.

We used the visualization tool level diagrams [62], and it is freely available implementation for designers LD-Tool.² LD-Tool allows to correlate design objectives with decision variables by providing two graphs. The first graph contains each objective, where its Y-axis is the p-norm $\|J(\theta)\|_p$ of the objectives vector and the X-axis corresponds to each objective value $J_a(\theta)$ [see Fig. 7(a)]. The second graph shows $\|J(\theta)\|_p$ with respect to each decision variable [see Fig. 7(c)]. Thus, a given solution will have the same y-value in all graphs. In addition, the solutions were clustered using a hierarchical clustering algorithm, and all the graphs were colored according to the resulting clusters.

a) Open-loop identification using bulk time-series data: For the time-series data sets, we used our implementation of the synthetic circuit (see Appendix II) and followed the protocol in Appendix III. The samples were measured using a 96-well Plate Reader (Biotek Cytation3 Imaging) at every $T = 10$ min during approximately 400 min. Corresponding absorbance (OD) and bulk green fluorescent protein (GFP) fluorescence (F) of the cell population at different induction levels of AHL were obtained. The subsequent analysis was performed using custom scripts in MATLAB. The time-series data represent the open-loop NoQS/NoFb circuit dynamics across the population for values of the AHL inducer circuit input signal corresponding to 0, 1, 10, and 50 nM.

To compare these measurements with the model results, we first eliminated the effect of the culture medium absorbance and the autofluorescence of the cells, subtracting these measurements from the corresponding data sets. The Plate Reader gain was also removed. Then, we obtained the fluorescence per cell ($FOD = F/OD$) for each induction. As expected, the circuit shows higher fluorescence (higher FOD) at AHL = 0 nM than for the other data sets. Therefore, it was defined as the output signal baseline. The data sets for AHL values

¹Tool available in <http://www.mathworks.com/matlabcentral/fileexchange/39215>

²Tool available at <http://www.mathworks.com/matlabcentral/fileexchange/24042>

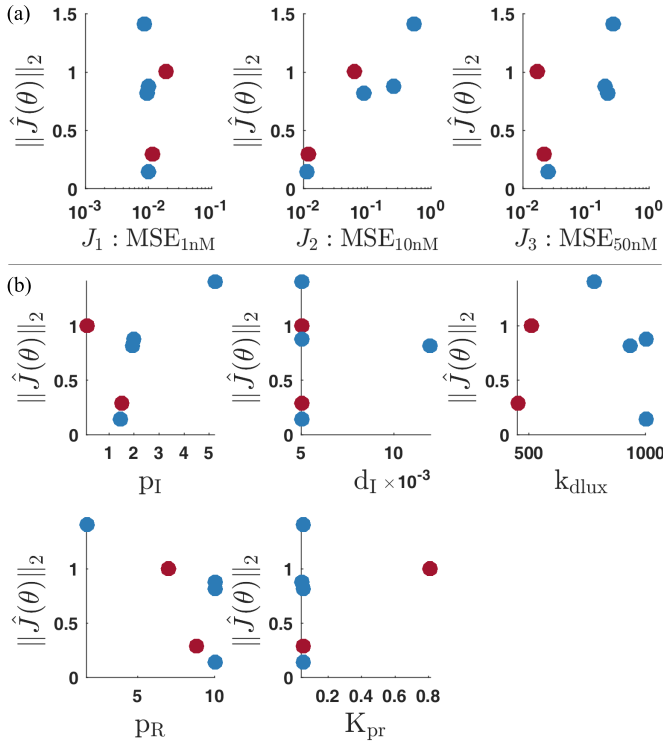


Fig. 7. Identification results. (a) LD-modified representation of the Pareto Front for each objective. (b) LD-modified representation of the Pareto Set for the five parameters. Colors reflect each resulting cluster.

of 1, 10, and 50 nM were normalized with respect to the baseline before they were used in the identification process.

The definition of the MOP consists of formulating the objectives to be optimized. We minimize the mean square error of the temporal profile of FOD (GFP fluorescence per cell) for the different induction levels of the input signal AHL taken as different objectives

$$J_{[a=1,\dots,3]}(\theta_{ol}) = \frac{1}{n} \sum_{q=1}^n \frac{1}{m} \sum_{k=1}^m (\hat{n}_{1_{aq}}(k) - n_{1_{aq}}(kT))^2 \quad (32)$$

where \hat{n}_1 is the experimental observation of FOD at the instant k , the index $a = 1$ corresponds to AHL = 1 nM, $a = 2$ to AHL = 10 nM, and $a = 3$ to AHL = 50 nM, $q = [1, \dots, n]$ is the observation sample (replica) measured at the point k for each objective, and m is the total number of temporal samples. The input stimulus is applied at $t_0 = 0$. The predicted observation n_1 is the result of the model simulation.

Finally, we looked for the values of the decision variables θ described in Table III which minimize all objectives $J(\theta)$. These three objectives are in conflict, so a tradeoff must be reached. Our problem can be formulated as an MOP

$$\begin{aligned} \min_{\theta \in \mathfrak{R}^5} J(\theta_{ol}) &= [J_1(\theta_{ol}), \dots, J_3(\theta_{ol})] \in \mathfrak{R}^3 \\ \text{s.t. equations (19)–(23)} \end{aligned} \quad (33)$$

where recall that the deterministic model (19)–(23) is used, as the time-series data corresponds to bulk averaged data. By using the deterministic model, we introduced some randomness by allowing the values of the parameters in each of

the N cells of the simulated population to be drawn from a normal distribution with mean equal to the estimated value of the parameter or to the one obtained from the literature for the fixed parameters. This extrinsic randomness partly accounts for the variance of the mean observed across the sample replicas.

b) Closed-loop identification using population distribution data: After choosing a preferable solution obtained from the first identification stage, we used the stochastic model to perform a second stage of the identification. In this one, we obtained the values for the closed-loop parameters, that is, the synthesis rate of AHL by LuxI (k_A) and the ratio between fluorescence and number of molecules (flow cytometer gain K_{fc}). In this context, flow cytometry data provide the distributions of fluorescent output species coming from many individual cells. These distributions are steady-state measurements of the population fluorescence at a given time t .

For this particular case, the measures were taken with the BD FACSCalibur flow cytometer under the protocol in Appendix IV. Thus, the absolute value of the relative errors for the mean e_μ and the noise strength e_{η^2} were optimized. These objectives are expressed by the indexes

$$\begin{aligned} J_4(\theta_{cl}) &= e_{\mu(QS/Fb)} = \left| \frac{\mu(QS/Fb) - K_{fc} \hat{\mu}(QS/Fb)}{\mu(QS/Fb)} \right| \\ J_5(\theta_{cl}) &= e_{\eta^2(QS/Fb)} = \left| \frac{\eta_{(QS/Fb)}^2 - \hat{\eta}^2(QS/Fb)}{\eta_{(QS/Fb)}^2} \right| \end{aligned} \quad (34)$$

where $\mu(QS/Fb)$ is the mean of the experimental flow cytometry data of the protein GFP/LuxI and $\eta_{(QS/Fb)}^2$ is its corresponding noise strength as defined in (35). The simulated mean and noise strength ($\hat{\mu}$ and $\hat{\eta}^2$, respectively) of GFP/LuxI (n_1) were computed from the stochastic simulation using the model (28) and the previously selected preferable solution obtained from the open-loop identification stage (with the corresponding identified parameter values).

Both GFP/LuxI $\hat{\mu}$ and $\hat{\eta}^2$ were obtained at steady state over the population of cells using both the laws of total expectation and total variance. The result is the set of equations (35), where $n_1^i(kT)$ is the value of protein GFP/LuxI (in number of molecules) at time instant kT for the i th cell, $k \in \mathcal{N}$, k_0T is the time instant at which steady state is reached, k_fT is the end of the simulation, and N is the number of cells of the population. The GFP/LuxI mean over the population at time kT is $m(kT)$, and its variance is $s^2(kT)$. The long-term mean of GFP/LuxI is $\hat{\mu}$ and the variance $\hat{\eta}^2$

$$\begin{aligned} m(kT) &= \frac{1}{N} \sum_{i=1}^N n_1^i(kT) \\ s^2(kT) &= \frac{1}{N} \sum_{i=1}^N (n_1^i(kT) - m(kT))^2 \\ \hat{\mu} &= \frac{1}{(k_f - k_0)T} \sum_{k=k_0}^{k_f} m(kT) \end{aligned}$$

$$\begin{aligned}\widehat{\sigma}^2 &= \frac{1}{(k_f - k_0)T} \sum_{k=k_0}^{k_f} s^2(kT) + \frac{1}{(k_f - k_0)T} \\ &\quad \times \sum_{k=k_0}^{k_f} (m(kT) - \widehat{\mu})^2 \\ \widehat{\eta}^2 &= \frac{\widehat{\sigma}^2}{\widehat{\mu}^2}.\end{aligned}\quad (35)$$

The MO-based identification process looked for the values of the closed-loop parameters $\theta_{cl} = [k_A, K_{fc}]$ that minimize the objectives $J_{4,5}(\theta_{cl})$ achieving a tradeoff between. This, again, can be formulated as an MOP

$$\begin{aligned}\min_{\theta_{cl} \in \mathfrak{N}^2} J(\theta_{cl}) &= [J_4(\theta_{cl}), J_5(\theta_{cl})] \in \mathfrak{R}^2 \\ \text{s.t. equation (28)}.\end{aligned}\quad (36)$$

IV. RESULTS

We carried out a first optimization of (30) to get an initial estimate of the Pareto front and the unknown parameters in the open-loop model. From this preliminary optimization, we found appropriate minimum and maximum limits for each objective, the so-called *pertinency* of each $J_i(\theta)$, which were used to enhance the search of the Pareto front in a narrower region of the parameters space. In both cases, the optimization was done using sp-MODE starting with an initial population of candidate solutions chosen randomly from a uniform distribution in the parameters space.

In the next step, an approximation of the Pareto front with six solutions was obtained [see Fig. 7(a)], together with the Pareto set containing their corresponding parameters [see Fig. 7(b)]. These solutions were classified using the *k-means* algorithm [63], yielding two clusters showing a tradeoff between the different objectives. This clustering helps to choosing the best parameters for different scenarios.

The analysis of the Pareto front [see Fig. 7(a)] shows the classical tradeoffs. For instance, red points, which are the best solutions for J_3 , correspond to one of the worst solutions for J_1 . The blue solutions are the best solutions for J_1 but not so good for J_2 and J_3 . However, in the Pareto set [see Fig. 7(b)], some trends can be seen. The LuxI translation rate, p_I , has an opposite trend to the one of the translation rate of LuxR, p_R . The degradation of the measured protein d_I has consistent values in both clusters (for all the solutions). This value is in the limit of the initial interval for this parameter and is compatible with the slowest measured growth rate of the microorganisms 140 min and equivalent to a degradation rate of $d_I = 0.005 \text{ min}^{-1}$ (including the growth rate as mentioned in Section II). Note, in particular, that a value of this parameter lying in the boundary of the initial interval means that the protein is very stable. It is not actively degraded by the cell and only dilutes due to growth rate.

Fig. 8 shows the values obtained for all five parameters in each cluster. The Y-axis reflects the initial search interval for each parameter. Out of five estimated parameters, two had approximately the same value in both clusters: $d_I \approx 0.005 \text{ min}^{-1}$ and $K_{pr} \approx 0.056$. Notice that even the parameters with the most different values in both clusters lie

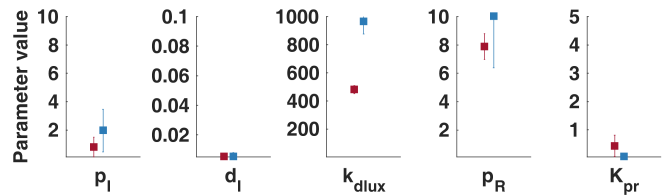


Fig. 8. Values of the estimated parameters for the two resulting clusters. The range of the Y-axis corresponds to the initial search interval for each parameter. Filled boxes represent the mean value of the parameter in each cluster, and vertical bars represent their standard deviation.

TABLE IV
PARAMETER VALUES OF THE SELECTED SOLUTION

Parameter	Description	Solution from red cluster	Solution from blue cluster
p_I	Translation rate of $\text{mRNA}_{\text{LuxI}}$	0.42 min^{-1}	1.98 min^{-1}
d_I	Degradation rate of LuxI	0.005 min^{-1}	0.015 min^{-1}
k_{dlux}	Dissociation constant of $(\text{LuxR} \cdot \text{AHL})_2$ to the lux promoter	550 molecules	932 molecules
p_R	Translation rate of $\text{mRNA}_{\text{LuxR}}$	6.15 min^{-1}	10 min^{-1}
K_{pr}	Fluorescence to number of molecules ratio (plate reader gain)	0.8	0.05

within close orders of magnitude: p_I is in the order of 2 min^{-1} and p_R is in the order of 8 min^{-1} . In addition, k_{dlux} has a range in the hundreds of molecules (500–1000 molecules), smaller than the initial search range (1–1000 molecules).

As a first validation, we selected two solutions from the resulting clusters, described in Table IV. Fig. 9(a) and (b) shows the good agreement between the model predictions and the experimental data for different induction levels (see Appendix III). The validation was performed with data sets not previously used for identification. The experimental data sets and the simulated data sets were treated as explained in Section III.

Note that all the solutions obtained are optimal in the sense of Pareto. It is the job of the engineer to select the most appropriate solution. To continue with the identification process, we choose the solution corresponding to the red cluster in Table IV as preferred one, as it better predicts high induction input values. This selection is useful for a model that will not be used with zero or low induction levels. The reason behind this choice is that the closed-loop circuit has a k_A different than zero, which means that there will be an induction level different than zero. With these parameters, for the open-loop circuit, the GFP/LuxI protein estimations coming from long-term distributions generated by the closed-loop stochastic model were used to optimize the best values for θ_{cl} . The Pareto front and set obtained are depicted in Fig. 10. As seen, one of these optimal solutions is very close to the ideal one [lowest norm to the ideal (0, 0) point]. Following the MCDM process, the chosen values for both the feedback and the flow cytometer gain were:

$$k_A = 0.048 \text{ min}^{-1}, \quad K_{fc} = 0.0777. \quad (37)$$

Interestingly, the value (values) for the closed-loop gain k_A is very close to the value previously reported in the literature $k_A = 0.04 \text{ min}^{-1}$ ([27] and references therein), thus

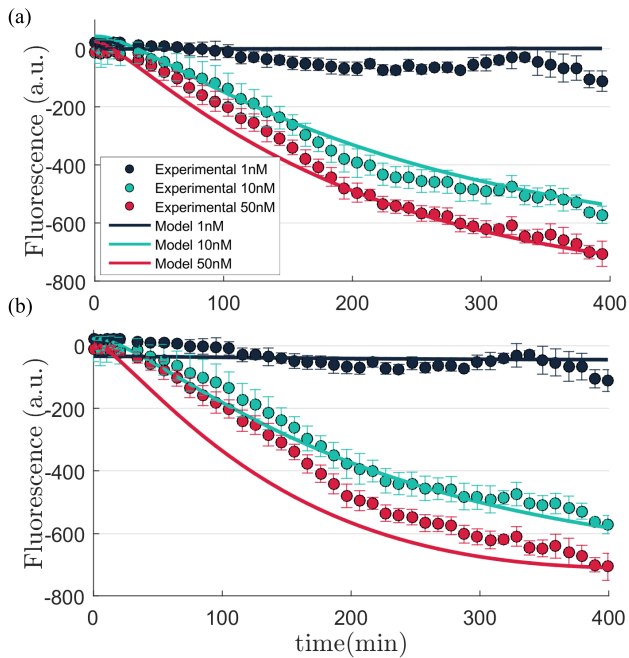


Fig. 9. Comparison between predicted and experimental temporal data. (a) Using parameter values from a solution in the red cluster for the three levels of input induction. (b) Using parameter values from a solution in the blue cluster for the three levels of input induction.

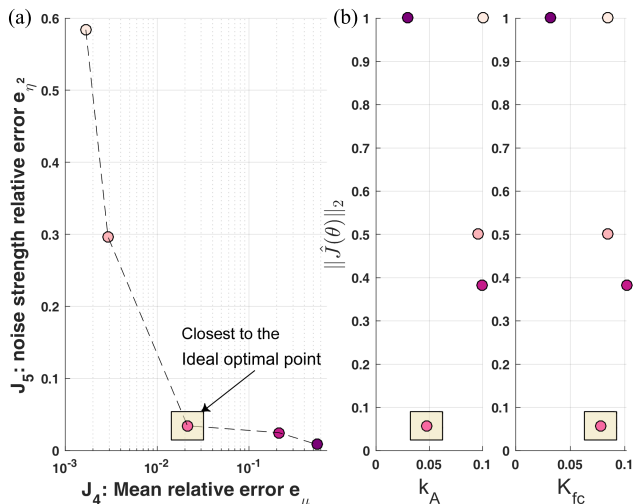


Fig. 10. (a) Pareto front showing the typical tradeoff between both relative errors J_4 and J_5 . (b) In the Pareto set θ_{cl} , the best optimal value for the feedback gain $k_A = 0.048 \text{ min}^{-1}$ was chosen.

validating both the model and the experimental setup to obtain the data.

Finally, we used the chosen values for validation. The resulting histograms are shown in Fig. 11. Histograms corresponding to the simulations of the stochastic model are shown in solid colors. The purple histogram corresponds to the open-loop circuit and the orange histogram to the closed-loop one. The experimental data obtained using flow cytometry are plotted in black dashed lines and black lines for the open- and closed-loop circuits, respectively (when the simulation data match the experimental one, the histograms colors become light orange or light purple accordingly). Notice the good

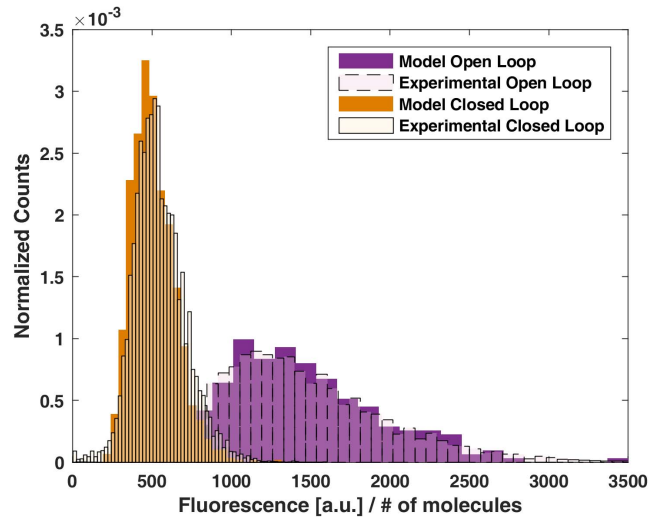


Fig. 11. Validation of predicted and experimental population distribution data for the open-loop and the closed-loop circuits. Open-loop Circuit: experimental histogram in black dashed lines and simulation in purple lines. Closed-Loop Circuit: experimental histogram in black lines and simulation in orange lines.

matching between the histograms from their superposition, validating both the open-loop identification process and, most importantly, the reduced model obtained for the feedback synthetic gene circuit.

V. CONCLUSION

Parameter degeneracy is often the price to pay for having models of biological gene networks and circuits that allow for a relevant degree of mechanistic biological interpretation and have parameters that can be associated with available experimental lab tuning knobs. These semimechanistic nonlinear biological models, usually arising from model order reduction of high-dimensional models based on first principles, inherit some of the problems of their parent models: lack of identifiability, low number of measurable variables and scarcity of available measurements, and inherent stochasticity of the parameters due to unmodelled dynamic interactions between the circuit and the host microorganism. In addition, many gene circuits, both natural and engineered, integrate feedback loops to achieve robustness against external perturbations. This makes the identification problem harder due to the practical lack of excitability of the input signals that can be applied to perturb the system.

Not less important is the limitation of experimental data to two broad classes typically available in a standard laboratory: time-course averaged measurements of cells growth and total production of a PoI in a culture, and steady-state distributions of protein production across the population of cells.

To cope with these problems, ensembles of models, i.e., sets of models with different structures and/or parameter values, have received much attention in the last years. Here, we have proposed a methodology, akin to ensemble modeling. Different from obtaining the model prediction as a weighted average of the individual models in the set, we express the identification problem as a multiobjective optimization one.

The use of an expanded space of model performance objectives in a multiobjective setting leads to a set of optimal solutions that can be clustered. In this way, sets of model parameters naturally arise, each one valid to describe an experimental scenario. By the formulation of the identification problem as a multiobjective optimization, one also allows easy integration of experimental data of different nature, requiring diverse performance indexes and fusion of information coming from different experiments.

We addressed the practical case of identifying the model parameters of a feedback circuit design to control the variance of the production of a PoI. For this circuit, only the steady-state distribution of protein production across the population of cells was available for the original feedback configuration. Besides, a circuit implementing the open-loop direct chain was built. For this last, time-course averaged measurements and a deterministic model were available. We used these to estimate sets of parameters describing the open-loop dynamics for different magnitudes of a perturbing input signal. In a second stage, we extended the identified open-loop dynamics to get the full closed-loop circuit. Then, we used the steady-state population distributions of production of the PoI and a stochastic model derived from the deterministic one to estimate the remaining parameters.

The methodology gives good identification results in the open-loop identification stage that can be translated into a complete identification of feedback circuit in the second stage of the identification, one of the closed loop. In the practical case addressed here, we obtain values for the feedback gain close to the ones reported in the literature.

APPENDIX I NONLINEAR PROPENSITIES

Following an approach used in [53], we validated the use of a high-order propensity function by simulating the pseudoreaction associated with $f(n_3, t)$ using the CLE approach and then comparing this result with the one obtained by simulating the set of corresponding original reactions using the Gillespie's direct method SSA.

In particular, denoting as n_3^i , $(\text{LuxR} \cdot \text{AHL})_2$ for the i th cell, the Hill-like function

$$f(n_3, t) \triangleq \frac{C_{\text{IPI}}}{d_{\text{mI}}} \left(\frac{k_{\text{dlux}} + \alpha_{\text{I}} n_3}{k_{\text{dlux}} + n_3} \right)$$

represents how the dimer $(\text{LuxR} \cdot \text{AHL})_2$ inhibits transcription of DNA from the corresponding gene $luxI$ into its mRNA, which, in turn, is translated into protein LuxI. Recalling the fast reactions (1) involving messenger $\text{mRNA}_{\text{luxI}}$

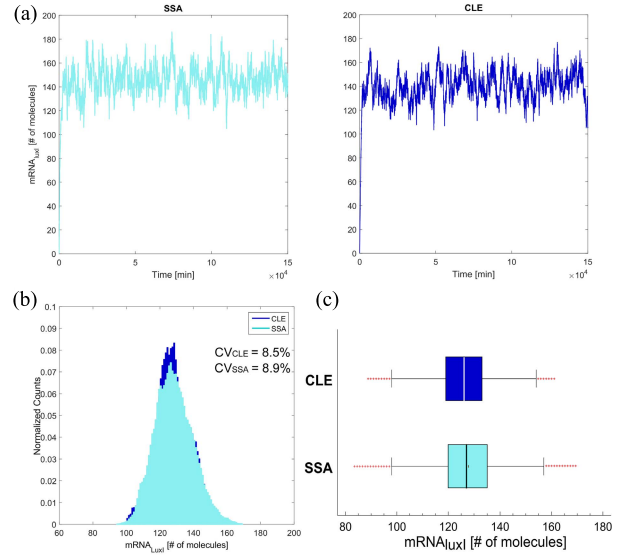
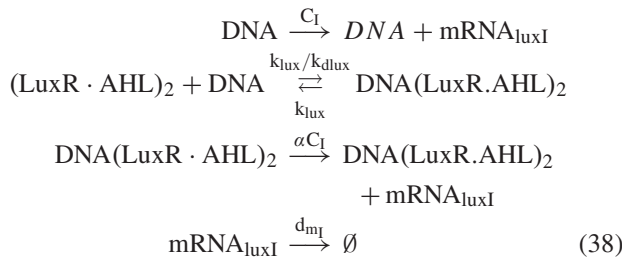
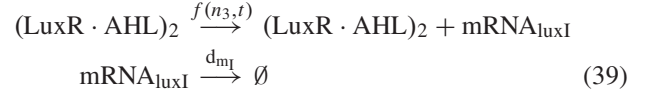


Fig. 12. SSA and CLE comparison for a complex propensity. (a) One realization of $\text{mRNA}_{\text{luxI}}$ made using the SSA (cyan color) and the CLE (blue color), respectively. Both trajectories match during a large temporal window (15×10^4 min). (b) Histograms with similar means and covariances. (c) Two medians' $\mu_{\text{SSA}} = 127.7$ and $\mu_{\text{CLE}} = 126.1$ molecules are statistically the same in the Box-and-Whisker plots.

they were approximated into two equivalent reactions as



where $f(n_3, t)$ describes the time evolution of $\text{mRNA}_{\text{luxI}}$ in the same way than in (38). To validate the propensity function $f(n_3, t)$, both sets of reactions were simulated. For one single cell ($i = 1$) and with the same conditions, reactions (39) were ran using the CLE approach, and reactions (38) were simulated using the Gillespie's SSA direct method.

Fig. 12(a) depicts, for one realization, how the SSA trajectory (left-top) matches very well with the CLE trajectory (right-top) during the whole simulation time. Both SSA and CLE trajectories have similar distributions with small differences between their first statistical moments, so $\mu_{\text{SSA}} \approx \mu_{\text{CLE}}$ and $\sigma_{\text{SSA}} \approx \sigma_{\text{CLE}}$ [see Fig. 12(b)]. It can be seen that the noise strength of $\text{mRNA}_{\text{luxI}}$ for the SSA distribution ($\eta_{\text{SSA}}^2 = 0.008$) matches closely with the same for the CLE ($\eta_{\text{CLE}}^2 = 0.0072$).

Finally, Fig. 12(c) shows the Box-and-Whisker plots of mRNA of the $luxI$ SSA and CLE realizations. Their medians (red line) are practically the same, and the Kruskal–Wallis test [64] reveals that there is no statistically significant difference between their medians at the 95.0% confidence level [test statistic, p-value] = $[-2.09067 \times 10^6, 1.0]$.

APPENDIX II EXPERIMENTAL IMPLEMENTATION

Following the implementation in [28], we engineered and implemented the closed-loop QS/Fb and open-loop NoQS/NoFb circuits *in vivo*. We used components from the iGEM Registry of Standard Biological Parts. All parts were

cloned using the Biobrick's foundation 3 Antibiotic Assembly method and confirmed by sequencing.

The circuit QS/Fb couples both QS-based cell-to-cell communication and the negative feedback subsystems. It was split in two subunits integrated in different plasmids. On the one hand, plasmid pCB2tc contains the gene *luxR* (part BBa_C0062) coding for the protein LuxR constitutively expressed under the control of a medium strength promoter (part BBa_J23106) and a strong RBS (part BBa_B0034). This insert was cloned into the pACYC184 plasmid cloning vector (p15A origin, chloramphenicol/tetracycline). On the other hand, plasmid pYB06ta contains gene *luxI* (part BBa_C0161) under control of the P_{luxR} repressible promoter (part BBa_R0062) and a strong RBS (part BBa_B0034). The strong RBS BBa_B0034 and the GFP (part BBa_E0040) were inserted using GIBSON assembly (NEB Catalog Number E2611S) upstream of *luxI*, right after the P_{luxR} promoter. This way, GFP, used as PoI [in Fig. 1(a)] is coexpressed with LuxI. None of them are tagged for degradation, so the main decay source for both proteins is the growth-related dilution. Therefore, their dynamics can be considered as equivalent. These constructions were inserted into the pBR322 plasmid cloning vector (pMB1 origin, ampicillin/tetracycline). Finally, both plasmids pCB2tc and pYB06ta were cotransformed in competent cells (DH-5 α , Invitrogen).

As a control network, we implemented the circuit NoQS/NoFb that removes both QS and the feedback loop. To this end, the above-mentioned plasmid pCB2tc was cotransformed with the plasmid pAV02ta (pMB1 origin, ampicillin/tetracycline) containing only GFP downstream of the P_{luxR} repressible promoter (part BBa_R0062) and the the strong RBS (part BBa_B0034). Both were cloned in the pBR322 plasmid cloning vector.

To obtain the data sets used in this paper, we followed the same protocol for bacterial growth as in [28], and the resulting samples were measured using the BD FACSCalibur flow cytometer (original default configuration parameters) and analyzed using custom scripts in MATLAB for the distribution of protein levels across the population of cells. For the time-series data sets, we followed the same protocol as in [65], and the resulting samples were measured and incubated with the appropriate induction using the Biotek Cytation3 Imaging Plate Reader and analyzed using custom scripts in MATLAB.

APPENDIX III

EXPERIMENTS FOR TIME-SERIES DATA SETS

The following experiments were performed. *E. coli* cells (Top 10, NEB) carrying the pCB2tc and pAV02ta plasmids for the open-loop circuit and the pCB2tc and pYB06ta plasmids for the closed-loop circuit were grown overnight in LB medium with the appropriate antibiotics. Then, 96 well plates were inoculated at $OD_{600} \approx 0.05$ and incubated to reach an optical density $OD_{600} \approx 0.1$. At this point, selected wells were induced with appropriate concentrations of AHL (N-3-Oxohexanoyl-L-homoserine lactone, Santa Cruz Biotechnology Catalog Number SC205396) and incubated for 400 min. Measurements were taken with a Biotek Cytation3 Imaging Plate Reader with the following protocol: 7'

shaking, OD measurement, and Fluorescent measurement. Each condition was performed in four replicates (data sets) for each condition. For each experiment, we have four conditions (ranging from 0- to 50-nM AHL), four data sets, and absorbance and fluorescence measurements every 10 min during 400 min of incubation after induction.

APPENDIX IV

EXPERIMENTS FOR STEADY-STATE POPULATION DISTRIBUTION DATA SETS

The following experiments were performed. Two sets of *E. coli* cells (cloning strain DH-5 α) carrying the closed-loop QS/Fb and the open-loop NoQS/NoFb circuits, respectively, were inoculated from -80 °C stocks into 3 mL of LB with appropriate antibiotics, followed by an overnight incubation at 37 °C and 250 rpm in 14-mL culture tubes. When the cultures reached an optical density $OD_{600} = 4$ (Eppendorf BioPhotometer D30), the overnight cultures were diluted 500-fold ($OD_{600} = 0.02$) into M9 medium with appropriate antibiotics. These were used to inoculate new cultures that were incubated for 7 hr (37 °C, 250 rpm, 14-mL culture tubes) until they reached an OD_{600} between 0.2 and 0.3. At this point, cell growth and protein expression were interrupted by transferring the culture into an ice-water bath for 10 min. Next, 50 μ L of each tube were transferred into 1 mL of phosphate-buffered saline with 500 μ g/mL of the transcription inhibitor rifampicin (PBS + Rif) in one 5-mL test tube and incubated during 1 hr in a water bath at 37 °C so that transcription kept was blocked and GFP had time to mature and fold properly. After that, samples were measured using the BD FACSCalibur flow cytometer (original default configuration parameters), and data were analyzed using custom scripts in MATLAB.

REFERENCES

- [1] V. Chellaboina, S. P. Bhat, W. H. Haddad, and D. S. Bernstein, "Modeling and analysis of mass-action kinetics," *IEEE Control Syst. Mag.*, vol. 29, no. 4, pp. 60–78, Aug. 2009.
- [2] J. Picó, A. Vignoni, E. Picó-Marco, and Y. Boada, "Modelling biochemical systems: From mass action kinetics to linear noise approximation," *Revista Iberoamer. Autom. Inf. Ind.*, vol. 12, no. 3, pp. 241–252, 7 2015.
- [3] L. A. Segel and M. Slemrod, "The quasi-steady-state assumption: A case study in perturbation," *SIAM Rev.*, vol. 31, no. 3, pp. 446–477, 1989.
- [4] T. P. Prescott and A. Papachristodoulou, "Layered decomposition for the model order reduction of timescale separated biochemical reaction networks," *J. Theor. Biol.*, vol. 356, pp. 113–122, Sep. 2014.
- [5] B. Mélykúti, J. P. Hespanha, and M. Khammash, "Equilibrium distributions of simple biochemical reaction systems for time-scale separation in stochastic reaction networks," *J. Roy. Soc. Interface*, vol. 11, no. 97, p. 20140054, Aug. 2014.
- [6] L. Ljung and T. Söderström, *Theory and Practice of Recursive Identification*. Cambridge, MA, USA: MIT Press, 1986.
- [7] E. Walter and L. Pronzato, *Identification of Parametric Models: From Experimental Data*. London, U.K.: Springer-Verlag, 1997.
- [8] A. Gábor and J. R. Banga, "Robust and efficient parameter estimation in dynamic models of biological systems," *BMC Syst. Biol.*, vol. 9, no. 1, p. 74, 2015.
- [9] H.-M. Kaltenbach, S. Dimopoulos, and J. Stelling, "Systems analysis of cellular networks under uncertainty," *FEBS Lett.*, vol. 583, no. 24, pp. 3923–3930, 2009.
- [10] L. Pronzato and A. Pázman, *Design of Experiments in Nonlinear Models (Lecture Notes in Statistics)*, vol. 212. New York, NY, USA: Springer-Verlag, 2013.
- [11] G. Jia, G. Stephanopoulos, and R. Gunawan, "Ensemble kinetic modeling of metabolic networks from dynamic metabolic profiles," *Metabolites*, vol. 2, no. 4, pp. 891–912, 2012.

- [12] A. F. Villaverde and J. R. Banga, "Reverse engineering and identification in systems biology: Strategies, perspectives and challenges," *J. Roy. Soc. Interface*, vol. 11, no. 91, p. 20130505, 2014.
- [13] D. Henriques, A. F. Villaverde, M. Rocha, J. Saez-Rodriguez, and J. R. Banga, "Data-driven reverse engineering of signaling pathways using ensembles of dynamic models," *PLoS Comput. Biol.*, vol. 13, no. 2, p. e1005379, 2017.
- [14] V. Hsiao, A. Swaminathan, and R. M. Murray, "Control theory for synthetic biology: Recent advances in system characterization, control design, and controller implementation for synthetic biology," *IEEE Control Syst. Mag.*, vol. 38, no. 3, pp. 32–62, Jun. 2018.
- [15] G. Lillacci and M. Khammash, "Parameter estimation and model selection in computational biology," *PLoS Comput. Biol.*, vol. 6, no. 3, p. e1000696, 2010.
- [16] J. Hasenauer, S. Waldherr, N. Radde, M. Doszczak, P. Scheurich, and F. Allgöwer, "A maximum likelihood estimator for parameter distributions in heterogeneous cell populations," *Procedia Comput. Sci.*, vol. 1, no. 1, pp. 1655–1663, 2010.
- [17] R. Steuer, T. Gross, J. Selbig, and B. Blasius, "Structural kinetic modeling of metabolic networks," *Proc. Nat. Acad. Sci.*, vol. 103, no. 32, pp. 11868–11873, 2006.
- [18] M. A. Samee *et al.*, "A systematic ensemble approach to thermodynamic modeling of gene expression from sequence data," *Cell Syst.*, vol. 1, no. 6, pp. 396–407, 2015.
- [19] A. F. Villaverde *et al.*, "BioPreDyn-bench: A suite of benchmark problems for dynamic modelling in systems biology," *BMC Syst. Biol.*, vol. 9, no. 1, p. 8, 2015.
- [20] M. Srinivas and L. M. Patnaik, "Genetic algorithms: A survey," *Computer*, vol. 27, pp. 17–26, Jun. 1994.
- [21] C. G. Moles, P. Mendes, and J. R. Banga, "Parameter estimation in biochemical pathways: A comparison of global optimization methods," *Genome Res.*, vol. 13, no. 11, pp. 2467–2474, 2003.
- [22] M. Ashyraliyev, J. Jaeger, and J. G. Blom, "Parameter estimation and determinability analysis applied to *Drosophila* gap gene circuits," *BMC Syst. Biol.*, vol. 2, no. 1, p. 83, 2008.
- [23] J. A. Egea, R. Martí, and J. R. Banga, "An evolutionary method for complex-process optimization," *Comput. Oper. Res.*, vol. 37, no. 2, pp. 315–324, 2010.
- [24] J. V. Carrau, G. Reynoso-Meza, S. García-Nieto, and X. Blasco, "Enhancing controller's tuning reliability with multi-objective optimisation: From *Model in the loop* to *Hardware in the loop*," *Eng. Appl. Artif. Intell.*, vol. 64, pp. 52–66, Sep. 2017.
- [25] V. Maheshwari, G. P. Rangaiah, and L. Samavedham, "Multiobjective framework for model-based design of experiments to improve parameter precision and minimize parameter correlation," *Ind. Eng. Chem. Res.*, vol. 52, no. 24, pp. 8289–8304, 2013.
- [26] G. Reynoso-Meza, X. Blasco, J. Sanchis, and M. Martínez, "Controller tuning using evolutionary multi-objective optimisation: Current trends and applications," *Control Eng. Pract.*, vol. 28, pp. 58–73, Jul. 2014.
- [27] A. Vignoni, D. A. Oyarzún, J. Picó, and G.-B. Stan, "Control of protein concentrations in heterogeneous cell populations," in *Proc. Eur. Control Conf. (ECC)*, 2013, pp. 3633–3639.
- [28] Y. Boada, A. Vignoni, and J. Picó, "Engineered control of genetic variability reveals interplay among quorum sensing, feedback regulation, and biochemical noise," *ACS Synth. Biol.*, vol. 4, no. 9, pp. 1796–1806, 2017.
- [29] R. L. Fernandes *et al.*, "Experimental methods and modeling techniques for description of cell population heterogeneity," *Biotechnol. Adv.*, vol. 29, no. 6, pp. 575–599, 2011.
- [30] K. A. Eglund and E. P. Greenberg, "Conversion of the *Vibrio fischeri* transcriptional activator, LuxR, to a repressor," *J. Bacteriol.*, vol. 182, no. 3, pp. 805–811, 2000.
- [31] H. B. Kaplan and E. P. Greenberg, "Diffusion of autoinducer is involved in regulation of the *Vibrio fischeri* luminescence system," *J. Bacteriol.*, vol. 163, no. 3, pp. 1210–1214, 1985.
- [32] C. Fuqua, M. R. Parsek, and E. Greenberg, "Regulation of gene expression by cell-to-cell communication: Acyl-homoserine lactone quorum sensing," *Annu. Rev. Genet.*, vol. 35, no. 1, pp. 439–468, 2001.
- [33] A. L. Schaefer, D. L. Val, B. L. Hanzelka, J. E. Cronan, Jr., and E. P. Greenberg, "Generation of cell-to-cell signals in quorum sensing: Acyl homoserine lactone synthase activity of a purified *Vibrio fischeri* LuxI protein," *Proc. Nat. Acad. Sci. USA*, vol. 93, no. 18, pp. 9505–9509, 1996.
- [34] B. Alberts *et al.*, *Essential Cell Biology*, 3rd ed. New York, NY, USA: Garland, 2009.
- [35] T. F. Weiss, *Cellular Biophysics*, vol. 1. Cambridge, MA, USA: MIT Press, 1996.
- [36] E. J. Hancock, G.-B. Stan, J. A. Arpino, and A. Papachristodoulou, "Simplified mechanistic models of gene regulation for analysis and design," *J. Roy. Soc. Interface*, vol. 12, no. 108, p. 20150312, 2015.
- [37] X. Chen, E. Pham, and K. Truong, "TEV protease-facilitated stoichiometric delivery of multiple genes using a single expression vector," *Protein Sci.*, vol. 19, no. 12, pp. 2379–2388, Dec. 2010.
- [38] J. A. J. Arpino *et al.*, "Tuning the dials of synthetic biology," *Microbiology*, vol. 159, pp. 1236–1253, Jul. 2013.
- [39] U. Alon, *An Introduction to Systems Biology: Design Principles of Biological Circuits*. London, U.K.: Chapman & Hall, 2007.
- [40] M. Weber and J. Buceta, "Dynamics of the quorum sensing switch: Stochastic and non-stationary effects," *BMC Syst. Biol.*, vol. 7, p. 6, 2013.
- [41] M. L. Urbanowski, C. P. Lostro, and E. P. Greenberg, "Reversible acyl-homoserine lactone binding to purified *Vibrio fischeri* LuxR protein," *J. Bacteriol.*, vol. 186, no. 3, pp. 631–637, Jan. 2004.
- [42] J. G. Harman, "Allosteric regulation of the cAMP receptor protein," *Biochim. Biophys. Acta-Protein Struct. Mol. Enzymol.*, vol. 1547, no. 1, pp. 1–17, 2001.
- [43] Y. Boada, G. Reynoso-Meza, J. Picó, and A. Vignoni, "Multi-objective optimization framework to obtain model-based guidelines for tuning biological synthetic devices: An adaptive network case," *BMC Syst. Biol.*, vol. 10, no. 1, p. 27, 2016.
- [44] G. F. Kaufmann *et al.*, "Revisiting quorum sensing: Discovery of additional chemical and biological functions for 3-oxo-N-acylhomoserine lactones," *Proc. Nat. Acad. Sci. USA*, vol. 102, no. 2, pp. 309–314, 2005.
- [45] N. E. Buchler, U. Gerland, and T. Hwa, "Nonlinear protein degradation and the function of genetic circuits," *Proc. Nat. Acad. Sci. USA*, vol. 102, no. 27, pp. 9559–9564, 2005.
- [46] C. Roberts *et al.*, "Characterizing the effect of the *staphylococcus aureus* virulence factor regulator, SarA, on log-phase mRNA half-lives," *J. Bacteriol.*, vol. 188, no. 7, pp. 2593–2603, 2006.
- [47] M. Santillán and M. C. Mackey, "Dynamic regulation of the tryptophan operon: A modeling study and comparison with experimental data," *Proc. Nat. Acad. Sci. USA*, vol. 98, no. 4, pp. 1364–1369, 2001.
- [48] P. Nilsson *et al.*, "Kinetics of the AHL regulatory system in a model biofilm system: How many bacteria constitute a 'quorum'?" *J. Mol. Biol.*, vol. 309, no. 3, pp. 631–640, 2001.
- [49] R. Milo *et al.* (2010). *BIONUMB3R5, the Database of Useful Biological Numbers*. [Online]. Available: <http://bionumbers.hms.harvard.edu/>
- [50] A. B. Goryachev, D. J. Toh, and T. Lee, "Systems analysis of a quorum sensing network: Design constraints imposed by the functional requirements, network topology and kinetic constants," *Biosystems*, vol. 83, nos. 2–3, pp. 178–187, 2006.
- [51] R. Milo, R. Phillips, and N. Orme, *Cell Biology by the Numbers*. New York, NY, USA: Garland, 2016.
- [52] D. J. Higham, "Modeling and simulating chemical reactions," *SIAM Rev.*, vol. 50, no. 2, pp. 347–368, 2008.
- [53] M. L. Woods, M. Leon, R. Perez-Carrasco, and C. P. Barnes, "A statistical approach reveals designs for the most robust stochastic gene oscillators," *ACS Synth. Biol.*, vol. 5, no. 6, pp. 459–470, Feb. 2016.
- [54] J. Joo, S. J. Plimpton, and J. L. Faulon, "Statistical ensemble analysis for simulating extrinsic noise-driven response in NF- κ B signaling networks," *BMC Syst. Biol.*, vol. 7, no. 1, p. 45, 2013.
- [55] T. Toni and B. Tidor, "Combined model of intrinsic and extrinsic variability for computational network design with application to synthetic biology," *PLoS Comput. Biol.*, vol. 9, no. 3, p. e1002960, 2013.
- [56] Biobrick Foundation. (2006). *Part Registry*. Accessed: Feb. 20, 2015. [Online]. Available: <http://partsregistry.org/>
- [57] S. Basu, Y. Gerchman, C. H. Collins, F. H. Arnold, and R. Weiss, "A synthetic multicellular system for programmed pattern formation," *Nature*, vol. 434, no. 7037, pp. 1130–1134, Apr. 2005.
- [58] K. Miettinen, F. Ruiz, and A. P. Wierzbicki, "Introduction to multiobjective optimization: Interactive approaches," in *Multiobjective Optimization*. Berlin, Germany: Springer, 2008, pp. 27–57.
- [59] R. A. Krohling, F. Hoffmann, and L. S. Coelho, "Co-evolutionary particle swarm optimization for min-max problems using Gaussian distribution," in *Proc. IEEE Congr. Evol. Comput. (CEC)*, vol. 1, Jul. 2004, pp. 959–964.
- [60] G. Reynoso-Meza, J. Sanchis, X. Blasco, and M. Martínez, "Design of continuous controllers using a multiobjective differential evolution algorithm with spherical pruning," *Applications of Evolutionary Computation*. Berlin, Germany: Springer, 2010, pp. 532–541.

- [61] G. Reynoso-Meza, J. Sanchis, X. Blasco, and J. M. Herrero, "Multiobjective evolutionary algorithms for multivariable PI controller design," *Expert Syst. Appl.*, vol. 39, no. 9, pp. 7895–7907, 2012.
- [62] X. Blasco, J. M. Herrero, J. Sanchis, and M. Martínez, "A new graphical visualization of n -dimensional Pareto front for decision-making in multiobjective optimization," *Inf. Sci.*, vol. 178, no. 20, pp. 3908–3924, 2008.
- [63] S. Lloyd, "Least squares quantization in PCM," *IEEE Trans. Inf. Theory*, vol. IT-28, no. 2, pp. 129–137, Mar. 1982.
- [64] W. H. Kruskal and W. A. Wallis, "Use of ranks in one-criterion variance analysis," *J. Amer. Stat. Assoc.*, vol. 47, no. 260, pp. 583–621, 1952.
- [65] Y. Boada, A. Vignoni, G. Reynoso-Meza, and J. Picó, "Parameter identification in synthetic biological circuits using multi-objective optimization," *IFAC-PapersOnLine*, vol. 49, no. 26, pp. 77–82, 2016.



Yadira Boada (M'13) was born in Quito, Ecuador, in 1985. She received the Engineering degree in electronics and control from the Escuela Politécnica Nacional, Quito, in 2010, and the M.Sc. degree in automatic control and industrial computing from the Universitat Politècnica de València, Valencia, Spain, in 2013, where she is currently pursuing the Ph.D. degree in control engineering, robotics and industrial informatics.

In 2015, she was with the Georgiev Lab, University of West Bohemia, Pilsen, Czech Republic, for a doctoral stay, where she was involved in the synthetic biology implementation of genetic circuits. In 2016, she was with the MOSAIC Group, Max Planck Institute of Molecular Cell Biology and Genetics, Dresden, Germany, for a doctoral stay, where she was involved in the efficient implementation of algorithms for stochastic simulation. She has been carrying out her research with the Synthetic Biology and Biosystems Control Lab, Institute ai2, Universitat Politècnica de València. She has co-authored more than 20 journal articles and contributions in conferences. Her current research interests include the methodologies for design and optimal tuning of feedback control gene synthetic networks with application in the biotechnology industry. She uses concepts and methodologies drawn from the system dynamics and feedback control, simulation of stochastic systems, and multiobjective optimization. Having a background in engineering, she validates her *in silico* designs by implementing synthetic gene circuits in bacteria.



Alejandro Vignoni (M'10) was born in La Plata, Buenos Aires, Argentina in 1984. He received the Engineering degree in electronics from the Universidad Nacional de La Plata, La Plata, Argentina, in 2008, and the Ph.D. degree in control engineering, robotics and industrial informatics from the Universitat Politècnica de València, Valencia, Spain, in 2014.

From 2015 to 2017, he was a Post-Doctoral Fellow with the Center for Systems Biology Dresden, Dresden, Germany, and the Max Planck Institute of Molecular Cell Biology and Genetics, Dresden.

Since 2017, he has been a Lecturer with the Department of Systems Engineering and Control and a Researcher with the Synthetic Biology and Biosystems Control Lab, Institute ai2, Universitat Politècnica de València. He has co-authored more than 30 journal articles and contributions in conferences, with more than 182 citations. The main question underpinning his research is how interactions among different levels of hierarchical organizations allow us to reveal emergent behaviors in complex systems, especially biological ones. He develops new algorithms and methodologies to design synthetic gene circuits for biotechnology by applying systems dynamics, feedback control, and sliding modes theory to the modeling, simulation, and control of biological systems.

Dr. Vignoni was a recipient of several awards and scholarships, including the ELBE Postdoctoral Fellowship from the Max Planck Society. He has tutored several Master Thesis and is Cotutoring a Ph.D. thesis.



Jesús Picó (M'89–SM'08) received the M.S.Eng. degree in engineering science and the Ph.D. degree in control engineering from the Universitat Politècnica de València (UPV), Valencia, Spain, in 1989 and 1996, respectively.

He is currently a Full Professor with the Department of Systems Engineering and Control, UPV, where he leads the Synthetic Biology and Biosystems Control Lab. Since 2017, he has been serving as a member of the Scientific Board of the Division for Coordination, Evaluation and Scientific-Technical Follow-up at the Spanish National Research Agency. He has published over 60 journal papers and book chapters on application of automatic control and multivariate statistic methodologies to bioprocesses control and estimation, and systems, and synthetic biology. His current research interests include sliding mode observers and nonlinear adaptive control of bioreactors, control of cell gene synthetic circuits considering performance constraints and optimal allocation of cell resources, and multiobjective optimization for analysis, synthesis, and tuning gene synthetic circuits.

Dr. Picó is a member of the IFAC Technical Committee on Biosystems and Bioprocesses. He has been a member of the Spanish Board of International Federation of Automatic Control (Comité Español de Automática).