Can optimal experimental design serve as a tool to characterize highly non-linear synthetic circuits?

Maxim Kryukov, Arthur Carcano, Grégory Batt, Jakob Ruess

Can optimal experimental design serve as a tool to characterize highly non-linear synthetic circuits?

Maxim Kryukov*,1,2, Arthur Carcano*,1, Gregory Batt[1] and Jakob Ruess[1]

***Abstract*— One of the most crippling problems in quantitative and synthetic biology is that models aiming to describe the real mechanisms of biochemical processes inside cells typically contain too many unknown parameters to be reliably inferable from available experimental data. Recent years, however, have seen immense progress in the development of experimental platforms that allow not only to measure biological systems more precisely but also to administer external control inputs to the cells. Optimal experimental design has been identified as a tool that can be used to decide how to best choose these control inputs so as to excite the systems in ways that are particularly useful for learning the biochemical rate constants from the corresponding data. Unfortunately, the experiment that is best to learn the parameters of a system depends on the precise values of these parameters, which are naturally unknown at the time at which experiments need to be designed. In this paper, we use a recently constructed genetic toggle switch as a case study to investigate how close to the best possible experiment we can hope to get with the most widely used optimal design approaches in the field. We find that, for strongly nonlinear systems such as the toggle switch, reliably predicting the information that can be gained from *a priori* fixed experiments can be difficult if the system parameters are not known very precisely. This suggests that a better strategy to guarantee informative experiments might be to use feedback control and to adjust the experimental plan in real time.**

I. INTRODUCTION

At the core of synthetic biology is the idea to rationally design biological circuits with new functionality by composing appropriate modules from existing repositories [1]. For this to be possible, mathematical models of the available modules must exist, and allow us to predict the behavior of their possible compositions. In recent years, however, it has become apparent that existing modules have been understood and characterized to an insufficient level of detail to satisfyingly predict how they will interact with each other or with the cellular environment [2], [3]. As a consequence, we routinely fail to predict the functionality of composed circuits reliably enough to serve as the basis of rational circuit design - and synthetic biology remains a field relying more on trial and error than on formal theoretical guidance.

One of the many reasons for this failure is that in early studies the characterization of modules often remained only qualitative [4], for instance the expression of a gene could be represented as a logic gate responding to the presence or absence of a regulating transcription factor, neglecting the dynamical aspects of the real biological process and quantizing all available data.

As qualitative representations proved to be insufficient to reliably predict the functionality of composed circuits, we started to increasingly turn to mechanistic dynamical systems models, derived from the actual biochemical reactions taking place in cells, which was made possible by improvements in experimental technologies that enabled the collection of ever more precise quantitative and time resolved data. However, it was quickly found that learning unique parameter values for such models from experimental data is often impossible due to structural or practical non-identifiabilites of the models, leading parts of the scientific community to the conclusion that mechanistic models in biology are generally "sloppy" [5], that is badly identifiable from data. While this observation is undoubtedly true, assigning sloppiness as a feature of models neglects that the identifiability of parameters depends primarily on the data that is available to constrain their values.

Recently, biology has been more and more penetrated by a fact long known in the field of systems identification: the successful identification of a system depends on the inputs that are used to excite it [6]. Hence, optimal experimental design approaches started to be developed for mechanistic models (either stochastic or deterministic) of biochemical reaction networks [7], [8], [9], as well as different types of possible experiments [10], [11] on comparably simple biological systems [12], [13]. Unfortunately, these approaches suffer from a fundamental, and inherently unresolvable, problem: the experiment that is optimal for characterizing a system depends on the precise dynamics of the system, which are unknown at the time when the experiment is designed, lest there would be no need to characterize the system in the first place.

In this paper, we study the possible consequences and implications of this problem on the case study of an externally controllable genetic toggle switch that has recently been implemented in *Echerichia Coli* [14]. This toggle switch is a bistable system consisting of two proteins that repress each other's production. Cells typically reside in one of two stable configurations where one protein is present while the production of the other is repressed. It is also remarkably challenging to excite the dynamics of the switch with external control inputs such that cells are forced out of the vicinity of the stable equilibria. The toggle switch therefore provides an interesting, but challenging, case study to test the limits of our current experimental design approaches.

* Both MK and AC contributed equally to this manuscript. [1] MK, AC, GB and JR are with Inria, Inria Saclay - Ile de France, 91120 Palaiseau, France, and Institut Pasteur, 75015 Paris, France. [2] MK is with the Department of Biological and Medical Physics, Moscow Institute of Physics and Technology, 141700, Russian Federation. kryukov.ma@phystech.edu, {arthur.carcano, gregory.batt, jakob.ruess}@inria.fr

The manuscript is organized as follows. In Section II, we introduce the genetic toggle switch together with the ordinary differential equations model that has been used to describe it in reference [14]. We introduce the Fisher information matrix (FIM) to quantify the amount of information provided by given experiments and discuss how the FIM is typically used for optimal experimental design. In Section III.A, we demonstrate how the informativeness of an experiment calculated according to the FIM depends on the model parameters that are used for the information calculation. We show *in silico* that imprecise knowledge of only some model parameters can be enough to suggest experimental designs that are almost useless. In Section III.B, in order to test if this can also happen in real experimental studies, we identify model parameters from data of preliminary characterization experiments performed in [14] and investigate how far from optimal FIM-based designs of additional experiments with these parameters are. In Sections III.C and III.D, we investigate the dependence on the model parameters of the calculated information for many different experiments and we study if feedback control can be used as a strategy to guarantee informative experiments. Section IV presents a discussion and our conclusions.

II. BACKGROUND

A. The genetic toggle switch

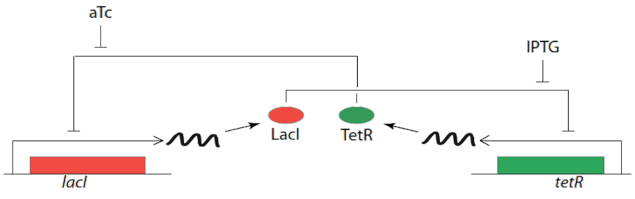

Fig. 1. A simplified schematic representation of the toggle switch. *LacI* and *TetR* are two proteins coded by *lacI* and *tetR*, respectively. Curly lines represent mRNAs. *TetR* inhibits the production of *LacI* and vice versa. aTc and IPTG are chemical inducers binding to the proteins. aTc prevents TetR from inhibiting *LacI* production and IPTG prevents *LacI* from inhibiting *TetR* production. "$\rightarrow$" represents production, and "$\dashv$" stands for inhibition.

In this manuscript, we investigate a recently published version of the toggle switch [14]. Shortly, the system consists of two genes *lacI* and *tetR* and their respective proteins - *LacI* and *TetR* (Fig. 1), which inhibit each other's production by binding as repressors to the respective genes. The system displays two stable equilibrium states. The toggle switch is an important circuit as it can be used to implement binary cell-decision making or cell fate differentiation. An important aspect of reference [14] is that the authors constructed an experimental platform based on microfluidics that gave them full control over the chemical environment in which cells are growing. This enabled them to externally modulate the repression activity of the two proteins by supplying small molecules that bind to the LacI and TetR repressor proteins, respectively, and impair their binding to the promoters of the genes (Fig. 1). The authors in [14] used these external control inputs to drive the toggle switch to its unstable equilibrium. Here, we will study how the control inputs can be used to excite the system in particularly informative ways.

In [14], the toggle switch has been quantitatively described by a deterministic mathematical model of ordinary non-linear differential equations that is capable of explaining the data of a large number of experiments in which the chemical environment of the cells has been dynamically varied in various ways. In this manuscript, we will use this model as ground truth and to simulate the behavior of the real system for different experiments that are designed when only partial information of the system is available. The model is based on the following chemical reactions:

- Transcription

$$\varnothing \xrightarrow{f_{\mathrm{L}}^{\mathrm{m}}(\mathrm{TetR},\,\mathrm{aTc})} \mathrm{mRNA_L}, \qquad \varnothing \xrightarrow{f_{\mathrm{T}}^{\mathrm{m}}(\mathrm{LacI},\,\mathrm{IPTG})} \mathrm{mRNA_T},$$

- Translation

$$\mathrm{mRNA_L} \xrightarrow{\kappa_{\mathrm{L}}^{\mathrm{p}}} \mathrm{mRNA_L} + \mathrm{LacI}, \qquad \mathrm{mRNA_T} \xrightarrow{\kappa_{\mathrm{T}}^{\mathrm{p}}} \mathrm{mRNA_T} + \mathrm{TetR},$$

- Degradation/Dilution

$$\mathrm{mRNA_L} \xrightarrow{g_{\mathrm{L}}^{\mathrm{m}}} \varnothing, \ \mathrm{mRNA_T} \xrightarrow{g_{\mathrm{T}}^{\mathrm{m}}} \varnothing, \ \mathrm{LacI} \xrightarrow{g_{\mathrm{L}}^{\mathrm{p}}} \varnothing, \ \mathrm{TetR} \xrightarrow{g_{\mathrm{T}}^{\mathrm{p}}} \varnothing$$

In these reactions, $f_{\mathrm{L}}^{\mathrm{m}}(\mathrm{TetR},\,\mathrm{aTc}) = \kappa_{\mathrm{L}}^{\mathrm{m0}} + \kappa_{\mathrm{L}}^{\mathrm{m}} \cdot h^{-}(\mathrm{TetR} \cdot h^{-}(\mathrm{aTc}, \theta_{\mathrm{aTc}}, \eta_{\mathrm{aTc}}), \theta_{\mathrm{TetR}}, \eta_{\mathrm{TetR}})$ and $f_{\mathrm{T}}^{\mathrm{m}}(\mathrm{LacI},\,\mathrm{IPTG}) = \kappa_{\mathrm{T}}^{\mathrm{m0}} + \kappa_{\mathrm{T}}^{\mathrm{m}} \cdot h^{-}(\mathrm{LacI} \cdot h^{-}(\mathrm{IPTG}, \theta_{\mathrm{IPTG}}, \eta_{\mathrm{IPTG}}), \theta_{\mathrm{LacI}}, \eta_{\mathrm{LacI}})$ are gene regulation functions, $\kappa_{\mathrm{L/T}}^{\mathrm{m0}}$, $\kappa_{\mathrm{L/T}}^{\mathrm{m}}$, $\kappa_{\mathrm{L/T}}^{\mathrm{p}}$, $g_{\mathrm{L/T}}^{\mathrm{m}}$ and $g_{\mathrm{L/T}}^{\mathrm{p}}$ are leakage transcription, transcription, translation, mRNA degradation, and protein degradation rate parameters for *lacI* (subscript L) and *tetR* (subscript R), and $h^{-}(x, \lambda, \eta) = 1/(1 + (x/\lambda)^{\eta})$ is a decreasing Hill function. Furthermore, the model distinguishes cell-external $u_{\mathrm{aTc/IPTG}}$ and internal aTc/IPTG concentrations of the chemical inducers. Combining everything together the dynamics of the reaction network can be captured by a system of six ordinary differential equations:

$$\begin{cases} \frac{d(\mathrm{mRNA_{LacI}})}{dt} = \kappa_{\mathrm{L}}^{\mathrm{m0}} + \frac{\kappa_{\mathrm{L}}^{\mathrm{m}}}{1+\left(\frac{\mathrm{TetR}}{\theta_{\mathrm{TetR}}} \times \frac{1}{1+(\mathrm{aTc}/\theta_{\mathrm{aTc}})^{\eta_{\mathrm{aTc}}}}\right)^{\eta_{\mathrm{TetR}}}} - g_{\mathrm{L}}^{\mathrm{m}} \times \mathrm{mRNA_{LacI}} \\ \frac{d(\mathrm{mRNA_{TetR}})}{dt} = \kappa_{\mathrm{T}}^{\mathrm{m0}} + \frac{\kappa_{\mathrm{T}}^{\mathrm{m}}}{1+\left(\frac{\mathrm{LacI}}{\theta_{\mathrm{LacI}}} \times \frac{1}{1+(\mathrm{IPTG}/\theta_{\mathrm{IPTG}})^{\eta_{\mathrm{IPTG}}}}\right)^{\eta_{\mathrm{LacI}}}} - g_{\mathrm{T}}^{\mathrm{m}} \times \mathrm{mRNA_{TetR}} \\ \frac{d(\mathrm{LacI})}{dt} = \kappa_{\mathrm{L}}^{\mathrm{p}} \times \mathrm{mRNA_{LacI}} - g_{\mathrm{L}}^{\mathrm{p}} \times \mathrm{LacI} \\ \frac{d(\mathrm{TetR})}{dt} = \kappa_{\mathrm{T}}^{\mathrm{p}} \times \mathrm{mRNA_{TetR}} - g_{\mathrm{T}}^{\mathrm{p}} \times \mathrm{TetR} \\ \frac{d(\mathrm{aTc})}{dt} = \begin{cases} \mathrm{k_{aTc}^{in}} \quad (u_{\mathrm{aTc}} - \mathrm{aTc}), & \text{if } u_{\mathrm{aTc}} > \mathrm{aTc} \\ \mathrm{k_{aTc}^{out}} \quad (u_{\mathrm{aTc}} - \mathrm{aTc}), & \text{if } u_{\mathrm{aTc}} \leqslant \mathrm{aTc} \end{cases} \\ \frac{d(\mathrm{IPTG})}{dt} = \begin{cases} k_{\mathrm{IPTG}}^{\mathrm{in}} \quad (u_{\mathrm{IPTG}} - \mathrm{IPTG}), & \text{if } u_{\mathrm{IPTG}} > \mathrm{IPTG} \\ k_{\mathrm{IPTG}}^{\mathrm{out}} \quad (u_{\mathrm{IPTG}} - \mathrm{IPTG}), & \text{if } u_{\mathrm{IPTG}} \leqslant \mathrm{IPTG} \end{cases} \end{cases}$$

The parameters identified in [14] are listed in Table I. Experimentally, concentrations of the two proteins can be measured every five minutes, at time points $t_0 = 0, \ldots, t_S$. To capture technical noise in data acquisition, we assume that measurements of both proteins are corrupted by independent additive Gaussian noise:

$$y_i(t_s) = x_i(t_s) + \varepsilon_i, \quad \varepsilon_i \sim_{iid} \mathcal{N}(0, \sigma^2), \quad i = 1, 2, \quad (1)$$

where $s = 0, \ldots, S$, x_1 and x_2 are the concentrations of LacI and TetR, respectively and $\sigma = 50$.

TABLE I

THE "REAL" VALUES OF THE SYSTEM PARAMETERS

κ_L^{m0}	0.0032	g_L^m	0.1386	η_{LacI}	2
κ_T^{m0}	0.119	g_T^m	0.1386	k_{aTc}^{out}	3000
κ_L^p	0.9726	η_{aTc}	2	θ_{IPTG}	0.0906
θ_{TetR}	29.9995	η_{IPTG}	2	κ_T^p	1.1697
θ_{LacI}	31.9132	g_T^p	0.0165	k_{IPTG}^{in}	$2.1832 \cdot 10^3$
k_{aTc}^{in}	369.6145	θ_{aTc}	11.6531	κ_L^m	8.3
κ_T^m	2.06	g_L^p	0.0165	η_{TetR}	2
k_{IPTG}^{out}	540				

B. The Fisher information matrix

To quantify the information about the model parameters that can be gained from a given experiment, we use – as is widely done in experimental design [13], [8], [15] – the determinant of the Fisher information matrix (FIM), $\det(\mathcal{I}(\lambda))$, for the vector of parameters λ. The FIM is defined as

$$[\mathcal{I}(\lambda)]_{k,l} = \mathbb{E}\left[\left(\frac{\partial}{\partial \lambda_k} \log f(Y;\lambda)\right)\left(\frac{\partial}{\partial \lambda_l} \log f(Y;\lambda)\right)\middle|\lambda\right],$$

where Y is the observation vector, $f(Y;\lambda)$ is the data likelihood for parameters λ and the expectation is taken over all possible realizations of the data. It can be readily seen that the FIM depends on the parameters λ that are used to evaluate it. For this reason, optimal experimental design is often implemented in iterations of design and experimentation, that is an experiment is designed using the current best estimate of the model parameters, the data is collected, and the parameter estimates are updated before the next experiment is designed [13].

For the Gaussian measurement noise model (Eq.1) that we use in this paper, the entries for the FIM can be calculated straightforwardly from the solution of the system of ODEs and its derivatives to model parameters according to

$$[\mathcal{I}(\lambda)]_{k,l} = \frac{1}{\sigma^2} \sum_{s=1}^{S} \sum_{i=1}^{2} \frac{\partial\, x_i(t_s,\lambda)}{\partial \lambda_k} \frac{\partial\, x_i(t_s,\lambda)}{\partial \lambda_l}. \quad (2)$$

Given that an analytical solution of the ODE system is not available, we evaluate these terms numerically using the `SENS_SYS` function for MATLAB [16].

III. RESULTS

A. Local design can lead to uninformative experiments

As a start, we aimed to demonstrate possible pitfalls in local experimental design in a setting as simple as possible. To this end, we considered an experiment that can intuitively be recognized as uninformative for characterizing the toggle switch where the applied inducers profiles fails to drive the system out of the vicinity of the stable equilibrium and so all cells remain close to their initial condition, displaying almost no dynamics. Calculating the FIM of this experiment with the parameter values λ^{real} identified in [14] (listed in Table I) confirms that this experiment is indeed largely uninformative (see Figure 2, top panel). If we aim to use experimental design in practice, the "real" parameter values of the toggle switch would, however, not be available and information would have to be calculated with estimated parameter values λ^{est}. To test what effect this can have on the calculated information values, we changed the values of 2 parameters, θ_{aTc} from 11.6531 to $\theta_{aTc}^{est} = 5$ and θ_{IPTG} from 0.0906 to $\theta_{IPTG}^{est} = 0.16$, and recalculated the FIM for the considered experiment. With this imprecise knowledge of only 2 parameters, cells are predicted to leave the vicinity of the stable equilibrium and circle around the unstable equilibrium (see Figure 2, bottom panel). Near the unstable equilibrium, the repression of both proteins is partially active at the same time, which implies that parameters characterizing the repression functions can be recovered from the observed dynamics. As a consequence, the experiment is evaluated as highly informative according to λ^{est} and would be a likely result of optimal experimental design procedures based on evaluation of the FIM at λ^{est}.

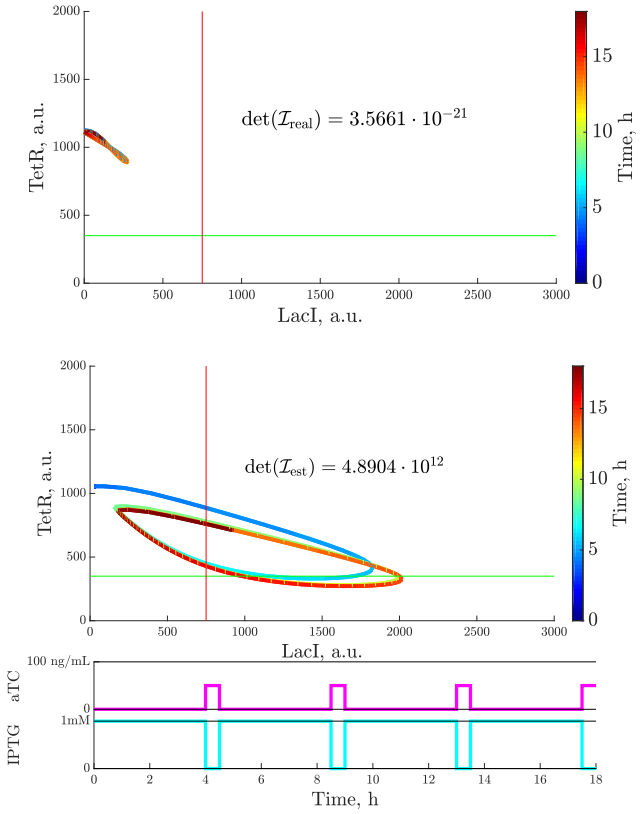

Fig. 2. The phase space trajectory for two sets of parameters. The top panel shows the (uninformative) real dynamics, while the bottom panel shows the (much more informative) system dynamics predicted with the incorrect parameter values. Inducer profiles of this experiment is shown at the bottom of the figure. The (real) unstable equilibrium point is located where the red and green vertical and horizontal lines cross.

TABLE II

COMPARISON OF "REAL" AND ESTIMATED PARAMETER VALUES

	λ^{real}	λ^{est}
$k_{\text{IPTG}}^{\text{in}}$	$2,1832 \cdot 10^3$	$5.5946 \cdot 10^3$
$k_{\text{IPTG}}^{\text{out}}$	540	576.7975
$k_{\text{aTc}}^{\text{in}}$	369.6145	307.6686
$k_{\text{aTc}}^{\text{out}}$	3000	$3.75921 \cdot 10^3$
θ_{IPTG}	0.0906	0.1275
θ_{aTc}	11.6531	18.9523

B. Local design with parameters learned from data

Are the pitfalls highlighted in the previous section only theoretical concerns or will they also be encountered when optimal design is used in practice to characterize a system?

To investigate this question, we considered a scenario of reference [14] that was encountered during the characterization of the toggle switch. In particular, the authors started their study by performing six experiments that were designated to characterize the toggle switch but then realized that these six experiments are not sufficiently informative to identify all the model parameters and that more experiments would be needed. To mimic this situation we took the data from the six characterization experiments (see Figure 3 in supplementary material of reference [14]) and used the CMA-ES [17] optimization routine to search for the maximum-likelihood estimators of six parameters of the model (see Table II) while the remaining parameters were assumed to be estimated correctly [1]. The results of the search, that is our best guess λ^{est} of the model parameters after having performed the six characterization experiments, are listed in Table II.

Further experiments are then needed to improve the estimates and could be designed using optimal experimental design. To test which experiment would be suggested by an optimal design based on local evaluation of the FIM at λ^{est}, we implemented an optimization scheme (based also on CMA-ES) that searches through experiments with periodically changing concentrations of the inducers IPTG and aTc (to use low-dimensional parametrization of the space of all possible experiments). More formally, we solved:

$$e^* = \arg\max_{e \in \mathcal{E}} \left\{ \det \left(\mathcal{I}(\lambda^{\text{est}}, e) \right) \right\}$$

where an experiment e is characterized by four parameters: the duration of IPTG induction (during which the aTc level is zero), the duration of aTc induction (during which the IPTG level is zero), the concentration of IPTG during induction and the concentration of aTc during induction.

The experiment e^* returned by this search and the corresponding system trajectories according to λ^{est} and λ^{real} are displayed in Figure 3. This experiment alternates between 0.44 mM of IPTG for 190 minutes and 47 ng/mL of aTc for 95 minutes. The top panel shows the system trajectory for the estimated parameters, that is the trajectory that has been

[1]In practice all parameter estimates would be imprecise. We chose to consider this simplified scenario here so that our conclusions are not a result of convergence problems of the parameter search.

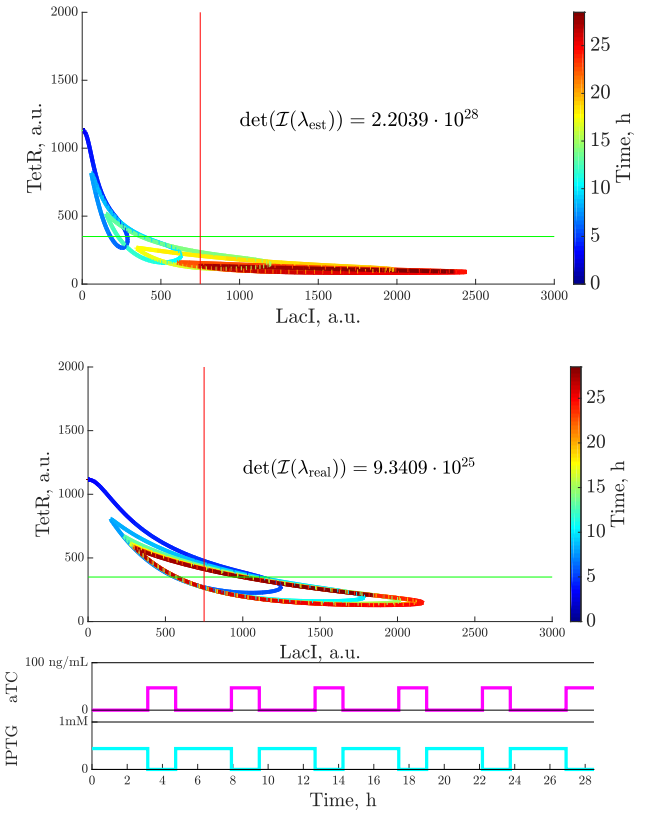

Fig. 3. The phase space trajectory of the toggle-switch during the experiment designed based on parameters estimates, with corresponding inducer profile. The upper graph shows the trajectory for the infered parameters, with which the experiment was designed, while the bottom pane shows the trajectory for the actual parameters.

deemed to allow optimal learning of the parameters: it covers a large area of the phase space and remains away of the two attracting equilibrium points for a substantial amount of time. The bottom panel shows the behavior of the system with the true parameters, that is the behavior that it will actually exhibit when the displayed inducer profiles are applied. Also here, the system visits different parts of the state space indicating that the designed experiment is at least reasonably informative. To test this quantitatively, and to investigate if there exist other experiments that are significantly more informative for the real system, we calculated the true FIM of this experiment $\mathcal{I}(\lambda^{\text{real}}, e^*)$ and performed another search for the best possible experiment, but this time using the true system parameters to evaluate the FIM:

$$e^*_{real} = \arg\max_{e \in \mathcal{E}} \left\{ \det \left(\mathcal{I}(\lambda^{\text{real}}, e) \right) \right\}$$

The results showed that e^*_{real} is more informative than e^*, but the difference turned out to be comparably small. We conclude that, in our scenario where the parameter estimates λ^{est} have been learned from several experiments and are already quite close to the true values, local experimental design can suggest useful experiments to finalize the characterization of the system.

C. Simple experiments provide predictable information

The success of the local design in the previous section raises the question of whether we can expect similar results in situations where only less precise estimates of the model parameters are available. A natural approach to investigate this question in a somewhat general way is to calculate how much the calculated information values for different experiments vary with the parameters. In particular, if we find that experiments that are very informative for the real system are only evaluated as informative for parameter sets close to the true parameters, then we should expect that these experiments will not be chosen by local design approaches when only imprecise estimates of the parameters are available. To test this, we sampled 150 parameter sets from uniform distributions covering the true parameters orders of magnitude[2] and calculated the information of all experiments performed in [14] with all sampled parameter values. In Figure 4 the calculated values of the determinant of the FIM are displayed in boxplots showing the distribution of this determinant for the sampled parameters.

There are several interesting conclusions that can be drawn from this analysis. First, simple non-periodic experiments ("*n*th calibration" in Fig. 4) that switch only once or twice between each inducer, such as the characterization experiments performed in [14], may or may not provide useful information, but their calculated informativeness depends only little on the exact values of the model parameters. Second, complex experiments – such as the periodic forcing experiments ("*n*th Kapitza") of [14] performed in order to drive the system to its unstable equilibrium point – have an informativeness that is very sensitive to the specific parameter values and therefore difficult to predict *a priori*. We conclude that complex experiments can in principle lead to very high information contents of the data but that they are difficult to tune precisely. If already relatively good estimates of the model parameters are available, this tuning can be successful as demonstrated in the previous section. However, if the parameters of the system are only poorly known, there is a significant risk that the attempt to design complex experiments will lead to experiments that are uninformative (as, for instance, demonstrated in Figure 2). As a final note, we point out that the experiment optimally designed for the true parameters in the previous section is evaluated as relatively uninformative for the majority of the sampled parameter values. It is therefore extremely unlikely that this experiment would ever be designed in any practical scenario where the true parameters are not known.

D. Feedback control leads to informative experiments

Finally, in light of the results obtained in this paper, we asked ourselves if, rather than optimizing information *a priori* over the space of possible inputs, a better strategy for experimental design might not be to design the trajectory that we would like the system to take and to realize this trajectory experimentally using feedback control. If the controller is well designed, this has the advantage that the control inputs can be adapted in real time to ensure that the system stays close to the chosen trajectory, even if the *a priori* knowledge of the model parameters turns out to have been imprecise when the experiment is performed. Completely uninformative experiments where the system gets stuck in a stable equilibrium, such as discussed in Figure 2, would therefore be prevented through adjustments to the inputs made by the controllers. To quantify whether feedback control experiments reliably provide high information content we calculated the FIM $\mathcal{I}(\lambda^{\text{real}}, e)$ for all the control experiments performed in [14] (bang-bang, weak PI, strong PI, see Figures 2, 3, and Sup.7 in the reference). We find that the information content of all these experiments, while being lower than the information content of the experiment designed in Section III.B with relatively precise parameters estimates λ^{est}, is still satisfactory and higher than the information content of all other experiments performed in reference [14]. Importantly, these experiments can be performed when only little knowledge about the system is available since they only require a useful and reachable trajectory or set point that serves as a control target.

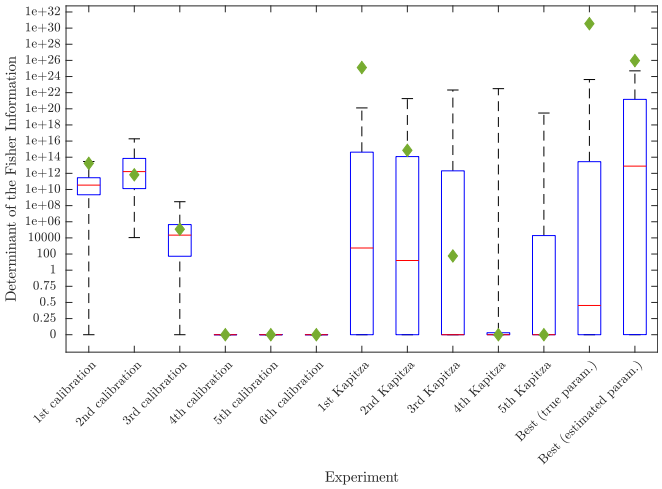

Fig. 4. Box plot of the distribution for sampled parameters, of the informativeness of different experiments. For each column, the blue box horizontal edges indicate the 25th and 75th percentile, the red line indicates the median, the whiskers indicate the most extreme data points, and the green diamond indicates the determinant of the experiment for the true parameters.

IV. CONCLUSIONS

In this work, using a genetic toggle switch as a case study, we aimed to investigate whether experimental design approaches based on the FIM can be employed to help with the characterization of biological circuits displaying dynamical responses that are strongly, and non-linearly, depending on the details of the stimulation profile that is used to drive the system.

We found that calculating information locally in parameter space using the current best parameter estimates can lead to informative designs when relatively precise estimates of

[2]More formally, for each true parameter λ^{real} the bounds of the distribution were $10^{\lfloor \log_{10}(\lambda^{real}) \rfloor}$ and $10^{\lfloor \log_{10}(\lambda^{real}) \rfloor + 1} - 1$

the model parameters are already available but might also lead to experiments that are largely uninformative when little knowledge about the system is available. Our results (Figure 4) also showed that robust experimental design approaches that aim to take uncertainty in model parameters into account, for instance averaging the FIM over "prior" distributions on model parameters, can be expected to favor simple experiments that have a decent, albeit relatively far from optimal, information content. Nonetheless, one should note that the requirement for computational experimental design to be useful in practice is not that the chosen designs need to be optimal or even just close to optimal, but rather that the iterative design of experiments leads to convergence of parameter estimates and that it does so faster and with less experimental effort compared to an expert designing the experiments [18]. Indeed, we found that the best experiment according to the parameter estimates λ^{est} that we learned from experimental data is close to optimal suggesting that iterative experimental design procedures could be successful. On the other hand, the results in Figure 4 show that the calculated information values can vary over orders of magnitude when the parameters are changed and that the success of local design approaches cannot really be guaranteed and suggests that we should also explore other possible directions for the design of experiments.

For instance, our results in Section III-D show that a strategy that essentially guarantees that informative data is obtained, is to use feedback control to drive the system to its unstable equilibrium, thereby deciding what control signals to send on the fly rather than trying to pre-choose informative signals before the experiment is started. It needs to be pointed out, however, that these control experiments are taken from reference [14] and have been carried out after the characterization of the system with a hard-coded set point for the controllers placed on the known location of the unstable equilibrium point. For the characterization of a system, this location will not be known and could not be used as a set point. Nevertheless, planning trajectories that visit different parts of the state space and using control algorithms to ensure that the system follows these trajectories seems perfectly possible for the toggle switch as well as more generally for other synthetic circuits for which it is difficult to predict whether or not a pre-specified stimulation profile will excite the system in any useful way [19]. Moreover, such approaches could be implemented in an online fashion, that is iterating experimental design and learning of the model parameters while the experiment is running [20], and/or making use of experimental platforms that allow sending input signals to individual cells and thereby to control different cells to follow their individual target trajectories [21], [22].

ACKNOWLEDGMENTS

This work was partially supported by EC funding H2020 FET OPEN 766840-COSY-BIO, and by the ANR under grants ANR-16-CE33-0018 (MEMIP) and ANR-16-CE12-444 0025 (COGEX). M.K. acknowledges support from the Amgen Scholars program.

REFERENCES

[1] P. Purnick and R. Weiss, "The second wave of synthetic biology: from modules to systems," *Nature Reviews Molecular Cell Biology*, vol. 10, pp. 410–422, 2009.

[2] R. Kwok, "Five hard truths for synthetic biology," *Nature*, vol. 463, pp. 288–290, 2010.

[3] S. Jayanthi, K. S. Nilgiriwala, and D. Del Vecchio, "Retroactivity controls the temporal dynamics of gene transcription," *ACS Synthetic Biology*, vol. 2, no. 8, pp. 431–441, 2013.

[4] C. Guet, M. Elowitz, W. Hsing, and S. Leibler, "Combinatorial synthesis of genetic networks," *Science*, vol. 296, no. 5572, pp. 1466–1470, 2002.

[5] R. Gutenkunst, J. Waterfall, F. Casey, K. Brown, C. Myers, and J. Sethna, "Universally sloppy parameter sensitivities in systems biology models," *PLoS Computational Biology*, vol. 3, no. 10, p. e189, 2007.

[6] N. Braniff and B. Munsky, "New opportunities for optimal design of dynamic experiments in systems and synthetic biology," *Current Opinion in Systems Biology*, vol. 9, pp. 42–48, 2018.

[7] E. Balsa-Canto, A. Alonso, and J. Banga, "Computational procedures for optimal experimental design in biological systems." *IET systems biology*, vol. 2, pp. 163–172, 2008.

[8] J. Ruess, A. Milias-Argeitis, and J. Lygeros, "Designing experiments to understand the variability in biochemical reaction networks," *Journal of the Royal Society Interface*, vol. 10, p. 20130588, 2013.

[9] Z. Fox and B. Munsky, "Designing Single-Cell Experiments with Discrete Stochastic Models," *Biophysical Journal*, vol. 114, p. 205A, 2018.

[10] C. Zechner, P. Nandy, M. Unger, and H. Koeppl, "Optimal variational perturbations for the inference of stochastic reaction dynamics," *IEEE 51st Annual Conference on Decision and Control (CDC). Maui, Hawaii, USA.*, 2012.

[11] J. Ruess and J. Lygeros, "Identifying stochastic biochemical networks from single-cell population experiments: a comparison of approaches based on the Fisher information." *IEEE 52nd Annual Conference on Decision and Control (CDC). Florence, Italy.*, 2013.

[12] S. Bandara, J. Schloder, R. Eils, H. Bock, and T. Meyer, "Optimal Experimental Design for Parameter Estimation of a Cell Signaling Model," *PLoS Computational Biology*, vol. 5, no. 11, p. e1000558, 2009.

[13] J. Ruess, F. Parise, A. Milias-Argeitis, M. Khammash, and J. Lygeros, "Iterative experiment design guides the characterization of a light-inducible gene expression circuit," *PNAS*, vol. 112, no. 26, pp. 8148–53, 2015.

[14] J.-B. Lugagne, S. Sosa Carrillo, M. Kirch, A. Köhler, G. Batt, and P. Hersen, "Balancing a genetic toggle switch by real-time feedback control and periodic forcing," *Nature Communications*, vol. 8, no. 1, p. 1671, 2017.

[15] M. Komorowski, M. J. Costa, D. A. Rand, and M. P. Stumpf, "Sensitivity, robustness, and identifiability in stochastic chemical kinetics models," *Proceedings of the National Academy of Sciences*, vol. 108, no. 21, pp. 8645–8650, 2011.

[16] V. Molla and R. Padilla, "Description of the matlab functions sens_sys and sens_ind," *Universidad Politécnica de Valencia, Spain*, 2002.

[17] N. Hansen, "The cma evolution strategy: A tutorial," 2016.

[18] D. Hagen, J. White, and B. Tidor, "Convergence in parameters and predictions using computational experimental design." *Interface Focus*, vol. 3, p. 20130008, 2013.

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

[20] L. Bandiera, Z. Hou, V. Kothamachu, E. Balsa-Canto, P. Swain, and F. Menolascina, "On-line optimal input design increases the efficiency and accuracy of the modelling of an inducible synthetic promoter." *Processes*, vol. 6, no. 9, p. 148, 2018.

[21] R. Chait, J. Ruess, T. Bergmiller, G. Tkačik, and C. Guet, "Shaping bacterial population behavior through computer-interfaced control of individual cells," *Nature Communications*, vol. 8, no. 1, p. 1535, 2017.

[22] M. Rullan, D. Benzinger, G. Schmidt, A. Milias-Argeitis, and M. Khammash, "An optogenetic platform for real-time, single-cell interrogation of stochastic transcriptional regulation," *Molecular Cell*, vol. 70, no. 4, pp. 745–756, 2018.