2022 European Control Conference (ECC)
July 12-15, 2022. London, United Kingdom

Performance and robustness analysis of control strategies for ameliorating cellular host-circuit interactions

Alexander P. S. Darlington and Declan G. Bates

Abstract— **Recent work on engineering synthetic cellular circuitry has shown that non-regulatory interactions brought about through competition for shared gene expression resources, such as ribosomes, can result in degraded performance (where circuit behaviour deviates from design specifications) or even failure (qualitatively different functionality). Numerous feedback control strategies have been proposed to decouple co-expressed genes in simple genetic circuits; ranging from feedback within the circuit, resource allocation schemes and growth-based feedback. In this work, we utilise a whole cell mathematical model, which captures key gene expression trade-offs, to compare these control strategies for their ability to ameliorate the impact of resource limitations, maintain growth and assess their robustness to host uncertainty and environmental variation.**

I. INTRODUCTION

A key goal of modern synthetic biology is to efficiently design and implement novel genetic regulatory networks within living cells. These small networks ('circuits') can be used to integrate external stimuli and guide microbial behaviour with numerous potential applications in a range of sectors [1]. However at present circuits show high levels of context dependency, with previously characterised modules showing different behaviours when implemented in different circuits, or circuit behaviour changing when implemented in different microbial strains [2].

A key cause of this context dependent failure of complex circuits composed of multiple genes is the host's innate biosynthetic constraints. Several studies have identified translational capacity as a key limitation; expression of genes from one module sequesters free ribosomes from other modules resulting in the emergence of a coupling effect not apparent from the circuit topology (e.g. [3]). Recent reports suggest metabolic resources also have significant effects [4], [5]. In Fig. 1, processes 1 and 2 are linked to each other and other host processes through common resource pools (either ribosomes or metabolites or both). As process 1 utilises the cell's resources through k_1, a corresponding disturbance d_1 is applied to the resource pool which impacts the value of k_2 and k_{host} driving process 2 and the host processes. Perturbation of the resource pool by synthetic circuits also reduces host gene expression and creates a growth defect commonly referred to as burden (e.g. [6]).

Funding from The Leverhulme Trust (grant RPG-2017-284) and Royal Academy of Engineering under the Research Fellowship scheme. High performance computing facilities provided by Warwick Integrative Synthetic Biology Centre (grant BB/M017982/1). The authors are with the Warwick Integrative Synthetic Biology Centre, School of Engineering, University of Warwick, Coventry, UK. (e-mail: a.darlington.1@warwick.ac.uk)

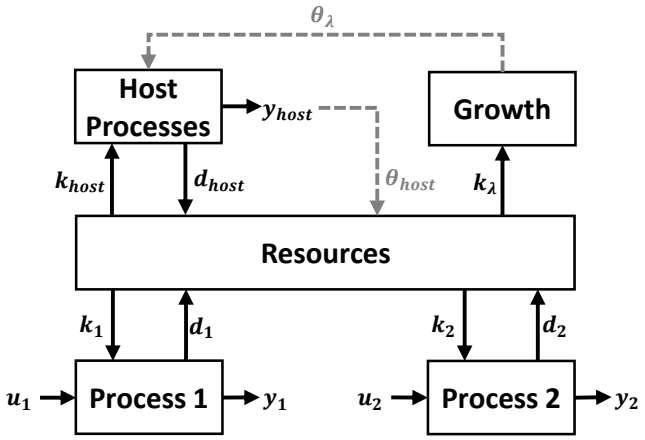

Fig. 1. Processes 1 and 2 compete for common pools of resources. This can be interpreted as a disturbance d_i which impacts the magnitude of k_j and results in coupling between the modules. Host processes also utilise the same resource pool; hence the output y_1 is determined by both its own input u_1 and indirectly by other system inputs (e.g. u_2.) The resource pool also sets cellular growth rate. The cell has endogenous feedback systems which are shown in grey (discussed further in [7]).

Feedback control provides a means to reduce these resource limitations by modifying the resource usage of each process in response to resources available or load (e.g. modifying the input u_2 in response to k_2 or modifying k_2 in response to the disturbance d_1). A number of recent works have reported designs of synthetic genetic feedback and feedforward systems to manage these resource limitations and/or host-circuit interactions. Here we review the biological mechanisms of these controllers before assessing their performance and robustness.

In [8], Shopera *et al.* created a proportional negative feedback controller based on an inhibitory transcription factor protein which is co-produced with the gene of interest and inhibits transcription of the genetic module containing itself and the output gene. Huang *et al.* created a post-transcriptional quasi-integral control system where an activator transcription factor is co-produced with the gene of interest. This activator enhances the expression of a small RNA which targets the process's mRNA for elimination [9].

In addition to these intra-module based systems, centralised control systems have been produced. In [10], we designed and implemented a control system which acts to divert ribosomes between host-specific and circuit-specific pools. A constitutively expressed regulator senses the demands of the synthetic circuit and increases or decreases production of circuit-specific quasi-orthogonal ribosomes. In [11], Ceroni *et al.* utilise growth burden as an input to a dCas9-based

Authorized licensed use limited to: University of Warwick. Downloaded on April 12,2023 at 09:31:25 UTC from IEEE Xplore. Restrictions apply.

control system. Decreases in growth rate are sensed and result in increased expression of a guide RNA which is taken up by a dCas9 protein. The guide RNA targets the dCas9 to the circuit promoters and inhibits their activation. Hence circuit activity is tuned down as cell growth decreases (and *vice versa*).

Barajas *et al.* recently reported a new burden alleviation feedforward controller based on manipulating global transcription rates in *E. coli* [12]. The synthetic system co-opts the endogenous ppGpp-based stringent response system; co-produced with the circuit genes is a mutant SpoT protein which only degrades ppGpp (the production domain is removed). This shifts the cell's physiological state into the ribosome production/enzyme inhibition regime in response to circuit expression.

In this paper we compare the performance of these new control systems in simple two gene circuits. In Section II, we review the control systems and develop mathematical models of the proposed control systems which take into account key host processes. In Section III we investigate the ability of the controllers to decouple genes in a simple two gene circuit and identify clear performance trade-offs. We then analyse the robustness of the controllers to variation in host strains.

II. Host-aware models of genetic controllers

In this Section, we develop models of the simple gene circuits and control systems in a host-aware framework. This framework accounts for key host constraints, including those imposed by metabolism, competition for ribosomes and growth.

A. Summary of the host-aware model

To account for host processes we embed the circuit and controller models within a previously described coarse grained model of *E. coli* [10]. This non-linear model is composed of 16 coupled ordinary differential equations which track the time evolution of a simple metabolism (consisting of an imported substrate s_i and universal energy carrier e), gene expression of a coarse grained transcriptome and proteome (consisting of transporters p_T, metabolic enzymes p_E, host proteins p_H and ribosomal proteins p_R), translational resource biogenesis and cell growth. The simple model of gene expression accounts for the production of mRNAs (m_x), the formation of mRNA-ribosome translation complexes (c_x) and final protein production (p_x) for all species in the proteome (i.e. $x \in \{T,\ E,\ H,\ R\}$). The rates of transcription and translation are both energy-dependent, hence incorporating natural metabolic feedback. Ribosomes are produced in a multistep resource biogenesis reaction with the production of ribosomal RNAs and r-proteins (p_r) which react to produce functional free ribosomes (R). All components are diluted at the cell's growth rate (λ) which is dynamically calculated within the model based on the level of the universal energy carrier and number of translating ribosomes. For a full description and parametrisation of the model see [10].

B. A summary of the core gene expression model

We assume that mRNAs are born spontaneoulsy at rate ω, scaled by the cell's internal energy e ($T_X(e)$ function), and a regulatory function $\mathcal{R}(\cdot)$.

$$\emptyset \xrightarrow{\omega \cdot T_X(e) \cdot \mathcal{R}(\cdot)} m \tag{1}$$

Free host ribosomes (R) bind these mRNAs to form translation complexes c. These translation complexes give rise to final proteins p.

$$m + R \underset{u}{\overset{b}{\rightleftharpoons}} c \xrightarrow{T_L(c,\ e)} p \tag{2}$$

All species dilute due to growth (λ) and the mRNAs and proteins also undergo first-order decay at rates d_m and d_p respectively:

$$m \xrightarrow{(\lambda + d_m)} \emptyset \quad c \xrightarrow{\lambda} \emptyset \quad p \xrightarrow{(\lambda + d_p)} \emptyset \tag{3}$$

Applying the Law of Mass Action to these reactions results in the following dynamic model of core gene expression:

$$\frac{dm}{dt} = \omega \cdot T_X(e) \cdot \mathcal{R}(\cdot) - b \cdot m \cdot R + u \cdot c - (\lambda + \delta_m) \cdot m \tag{4}$$

$$\frac{dc}{dt} = b \cdot m \cdot R - u \cdot c - T_L(c,\ e) - \lambda \cdot c \tag{5}$$

$$\frac{dp}{dt} = T_L(c,\ e) - (\lambda + \delta_p) \cdot p \tag{6}$$

$T_X(\cdot)$ and $T_L(\cdot)$ account for the energy dependence of transcription and translation:

$$T_X(e) = \frac{e}{\theta + e} \qquad T_L(c,\ e) = \frac{1}{n} \cdot \left(\frac{\gamma_{max} \cdot e}{\kappa_\gamma + e} \right) \cdot c \tag{7}$$

θ and κ_γ represent the energy thresholds for transcription and translation. γ_{max} is the maximal rate of translational elongation in amino acids per min. n is the length of the protein in amino acids. Note that γ_{max} and κ_γ are common to all species, see [10].

C. Open loop model

The circuit model is composed of two unregulated genes. The genes follow the dynamics of the core gene expression model (Section II-B). To distinguish between the genes, we introduce a subscript notation with each variable and constant of a gene being denoted with an i. That is, m_i is the mRNA of gene i and ω_i is the maximal transcription rate of gene i etc. The unregulated nature of their transcription means that $\mathcal{R}(\cdot)$ is set to 1.

D. Protein-based proportional feedback (fbprot)

To incorporate negative autoregulation developed by Shopera *et al.* we modify the mRNA birth rate regulation with a negative Hill function

$$\mathcal{R}_1(p_1) = \frac{1}{1 + (p_1/\kappa_1)^{h_1}} \tag{8}$$

To account for the co-translated autoregulator, we increase the circuit protein length from n_1 to $n_1 + n_f$, where n_f is the length of the regulator (here 300 amino acids).

Authorized licensed use limited to: University of Warwick. Downloaded on April 12,2023 at 09:31:25 UTC from IEEE Xplore. Restrictions apply.

E. sRNA-based integral feedback (fbrna)

To model integral feedback we first introduce a small RNA species s_1 which eliminates the circuit mRNA by an elimination reaction:

$$m_1 + s_1 \xrightarrow{\beta_1} \emptyset \tag{9}$$

s_1 is transcribed in a similar manner to m_1 but with its birth regulated by the circuit protein, modelled by an activatory Hill function. The dynamics of the small RNA are given by:

$$\frac{ds_1}{dt} = \omega_{s,1} \cdot T_X(e) \cdot \left(\frac{(p_1/\kappa_1)^{h_1}}{1 + (p_1/\kappa_i)^{h_1}} \right) \ ... \\ ... \ - \beta_1 \cdot s_1 \cdot m_1 - (\lambda + \delta_s) \cdot s_1 \tag{10}$$

The impact of the elimination of the mRNA means that circuit mRNA dynamics are updated to:

$$\frac{dm_1}{dt} = \omega_1 \cdot T_X(e) - b_1 \cdot m_1 \cdot R + u_1 \cdot c_1 \ ... \\ ... \ - \beta_1 \cdot s_1 \cdot m_1 - (\lambda + \delta_m) \cdot m_1 \tag{11}$$

The dynamics of the translation complex and protein are not changed from Eq. 5 and Eq. 6. To account for the co-translated autoregulator, the circuit protein length n_1 is also increased to $n_1 + n_f$.

F. Resource-based feedback (fboribo)

To model translational resource-based feedback, we first introduce a new ribosomal RNA species ρ and new orthogonal ribosome P. Orthogonal ribosomes are produced from host r-proteins:

$$\rho + p_R \underset{\varrho_r}{\overset{\varrho_f}{\rightleftharpoons}} P \tag{12}$$

The ribosomal RNAs are born spontaneously at rate ω_ρ, scaled by $T_X(e)$ and the impact of the inhibition by the controller protein p_f as described above. The dynamics of the orthogonal rRNA are given by:

$$\frac{d\rho}{dt} = \omega_\rho \cdot T_X(e) \cdot \left(\frac{1}{1 + (p_f/\kappa_f)^{h_f}} \right) \ ... \\ ... \ - \varrho_f \cdot \rho \cdot p_R + \varrho_r \cdot P - (\lambda + \delta_\rho) \cdot \rho \tag{13}$$

The dynamics of the production of the constitutively expressed controller protein follow those in Eq. 4 to 6, but with the host ribosomes R replaced with the orthogonal ribosomes P. The equations describing the dynamics of the circuit genes are not changed bar the replacement of R with P. The dynamics of free orthogonal ribosomes are given by:

$$\frac{dP}{dt} = \varrho_f \cdot p_R \cdot \rho - \varrho_r \cdot P - \lambda \cdot P \ ... \\ + \sum_Y \Big(T_L(c_Y,\ e) - b_Y \cdot m_Y \cdot P + u_Y \cdot c_Y \Big) \tag{14}$$

where the set $Y = \{1, 2, f\}$.

G. Growth-based feedback (fbgrow)

We model the growth based input of the new RNA g whose expression falls in response to growth (i.e. it rises in response to burden).

$$\emptyset \xrightarrow{\omega_g \cdot T_X \cdot \mathcal{R}(\lambda)} g \tag{15}$$

We model the actuator protein, dCas9, in two stages. First the core protein p_k is constitutively expressed with dynamics of m_k and c_k the same as depicted in Section II-B. In the second step the protein binds the new guide RNA species g to form the functional protein k:

$$g + p_k \underset{u_g}{\overset{b_g}{\rightleftharpoons}} K \tag{16}$$

The dynamics of these species are given by:

$$\frac{dg}{dt} = \omega_g \cdot T_X(e) \cdot \left(\frac{1}{1 + (\lambda/\kappa_\lambda)^{h_\lambda}} \right) \ ... \\ ... \ - b_g \cdot g \cdot p_k + u_g \cdot K - (\lambda + \delta_g) \cdot g \tag{17}$$

$$\frac{dp_k}{dt} = T_L(c_k,\ e) - b_g \cdot g \cdot p_k + u_g \cdot K \ ... \\ ... \ - (\lambda + \delta_{p_K}) \cdot p_K \tag{18}$$

$$\frac{dK}{dt} = b_g \cdot g \cdot p_k - u_g \cdot K - (\lambda + \delta_{p_K}) \cdot K \tag{19}$$

The action of the controller is through inhibition of the circuit birth rate:

$$\mathcal{R}_i(K) = \frac{1}{1 + (K/\kappa_k)^{h_k}} \tag{20}$$

where $i = \{1, 2\}$.

H. ppGpp-based feedforward controller (ffspo)

To model the ppGpp feedforward controller developed by [12], we introduce equations describing the dynamics of the new SpoT protein's mRNA m_s, translation complex c_s and protein p_s. These follow the same dynamics as circuit proteins depicted in Eq. 4-6. We assume that ppGpp is directly proportional to p_s concentration and so do not model its creation and action explicitly, using p_s as a proxy. ppGpp has systemic effects through its action on the host's RNA polymerase therefore we model the action of ppGpp through modification phenomenologically by updating T_X from [10]:

$$T_X(e,\ p_s) = \frac{e}{\theta_X + e} \cdot \phi_{s'} \quad \text{for } E,\ H,\ S,\ 1,\ 2 \tag{21}$$

$$T_X(e,\ p_s) = \frac{e}{\theta_X + e} \cdot \phi_s \quad \text{for } r,\ R \tag{22}$$

where the regulatory expressions are:

$$\phi_{s'} = 1 - v_s \cdot \left(\frac{(p_S/\kappa_s)^{h_s}}{1 + (p_S/\kappa_s)^{h_s}} \right) \tag{23}$$

$$\phi_s = 1 + v_s \cdot \left(\frac{(p_S/\kappa_s)^{h_s}}{1 + (p_S/\kappa_s)^{h_s}} \right) \tag{24}$$

We choose this formalism so that $p_s = 0$, $\phi_s = \phi'_s = 1$. The maximal level change in regulation is given by the scaling factor v_s. We also modify the circuit gene transcription rate T_X with Eq. 21.

Authorized licensed use limited to: University of Warwick. Downloaded on April 12,2023 at 09:31:25 UTC from IEEE Xplore. Restrictions apply.

III. Performance trade-offs

We assess the performance of each controller with respect to three objectives: (i) decoupling potential (increasing modularity), (ii) final gene expression level and (iii) growth. Here we consider how these different strategies improve the performance of the circuit shown in Fig. 1. Process 1 is constitutively (constantly) active producing protein p_1. Process 2 is induced (activated) at a given time t_{ind}. We allow the system to evolve until $t = t_{max}$ (here equivalent to 48 hours). We assess the ability of the control strategy to improve modularity (i.e. reduce coupling) through the metric Δ:

$$\Delta = \Big(p_1(t = t_{max}) - p_1(t = t_{ind})\Big)/p_1(t = t_{ind}) \quad (25)$$

Δ measures the fall in p_1 in response to the activation of p_2 (a measure of modularity failure). The second and third objectives have implications for the use of the system for real-world applications. Industrial applications often require high levels of protein production. For example, pathway enzymes need to be expressed at a sufficiently high level to ensure good flux to metabolites of interest. High levels of expression also reduce population heterogeneity due to reduced noise. Industrial applications often require high populations to maximise product production. Therefore high growth rates are desirable to shorten batch culture times. The optimal controller would drive Δ to zero while maximising p_1 and maximising λ.

A. Controller design process

The systems of ordinary differential equations describing the host, a two-gene circuit (producing proteins p_1 and p_2) and controller system were implemented in MATLAB 2019a (The MathWorks Inc., MA, USA) and solved using the inbuilt stiff-solver $ode15s$. The dynamics were simulated by first considering a run-in period (10^6 minutes) with no circuit activation (i.e. $\omega_1 = \omega_2 = 0$) to allow the host-controller system to achieve steady state. At $t = 0$ minutes, ω_1 was raised to 100 mRNAs per minute. The system was simulated until $t = t_{ind}$ (1440 minutes) before ω_2 was raised, also to 100 mRNAs per minute. We parameterised all host, circuit and controller genes, bar those discussed below, as described in [10]. We utilised a multi-objective optimisation approach to identify designs which minimise coupling and maximise final expression level:

$$\sigma_1 = \Delta^2 \qquad \sigma_2 = -p_1(t = t_{max}) \quad (26)$$

Note that we do not specifically select for final growth rate during the optimisation, using it only as a performance measure of the designed controllers. We carried out the optimisation using the inbuilt MATLAB $gamultiobj$ function with a population of 100 individuals, a maximum number of generations of 500 and a stall distance n to 50. We assessed the convergence of the front using a custom stopping criteria based on the Average Distance d between the points along the front. When the absolute difference between d over n generations was smaller than a threshold value v the algorithm terminated. For this study we set v to be 0.01. The parameters optimised for each control system are shown in Table I with the bounds shown in Table II. Some parameters were optimised on a log-scale (i.e. the parameter $k = 10^x$ is determined by optimising x).

B. Performance trade-offs

We first assessed the ability of control systems to protect a constitutively (constantly) expressed gene from the disturbance caused by the induction (activation) of a second gene. We first stimulated p_1 production before increasing p_2 produciton at time t_{ind} and calculated coupling metrics (as shown in Eq. 25), p_1 and p_2 steady states and growth rate.

We first considered design schemes as they appeared in the original work with the feedback controllers *fbprot* and *fbrna* acting only on p_1 and the feedforward controller *ffspo* being activated with p_2. Note that the centralised controllers *fboribo*, *fbgrow* and *ffspo* target all circuit genes as proposed in [10], [11], [12], this is crucial for their own action.

We designed the control systems as described above. We also re-designed the open loop circuit by varying the ribosome binding rates of the two circuit gene mRNAs, b_1 and b_2. This demonstrates a hard trade-off between gene expression coupling and final protein steady state with coupling reducing (smaller Δp_1) with reducing expression (p_1 steady state) (Fig. 2a). As the protein levels of the re-design circuit fall, the cell growth rate increases (Fig. 2c). All controller topologies are capable of showing improved performance over simple circuit redesign with at least some designs showing reduced decoupling at higher final protein steady states (Fig. 2a). Note also that the presence of the feedback systems turns the convex trade-off to a concave trade-off – addition of the control systems reduces the hard trade off (where increasing expression *must* significantly increase coupling) to a more concave trade-off where small increases in expression do not significantly increase coupling. Two of the centralised control systems break the trade-off (discussed below). The intra-module feedback control systems (where the circuit output directly fed back to the controller, such as *fbprot* and *fbrna*) show lowest growth rates for a given level of

TABLE I

Designed parameters

Controller	Parameters
fbprot	$\kappa_i,\ h_i$
fbrna	$\omega_{s,i},\ \kappa_i,\ h_i$
fboribo	$\omega_\rho,\ \omega_f,\ b_f,\ \kappa_f,\ h_i$
fbgrow	$\omega_g,\ \kappa_\lambda,\ h_\lambda,\ \omega_k,\ b_k, \kappa_k,\ h_k$
ffspo	$b_s,\ \kappa_s,\ v_s,\ h_s$
(Circuit redesign)	b_i

TABLE II

Optimised parameter bounds

Parameter (function)	Bounds	Scale
ω (Transcription rate)	10 to 1,000 mRNAs per min	log
b (RBS association rate)	10^{-3} to 1 cell/(min·molecules)	log
κ (Hill function threshold)	10^1 to 10^8 molecules	log
h (Hill coefficient)	1 to 4	linear
κ_λ (Threshold for growth)	10^{-4} to 1 (1/min)	log
v_s (ppGpp scaling factor)	0 to 1	lin

Authorized licensed use limited to: University of Warwick. Downloaded on April 12,2023 at 09:31:25 UTC from IEEE Xplore. Restrictions apply.

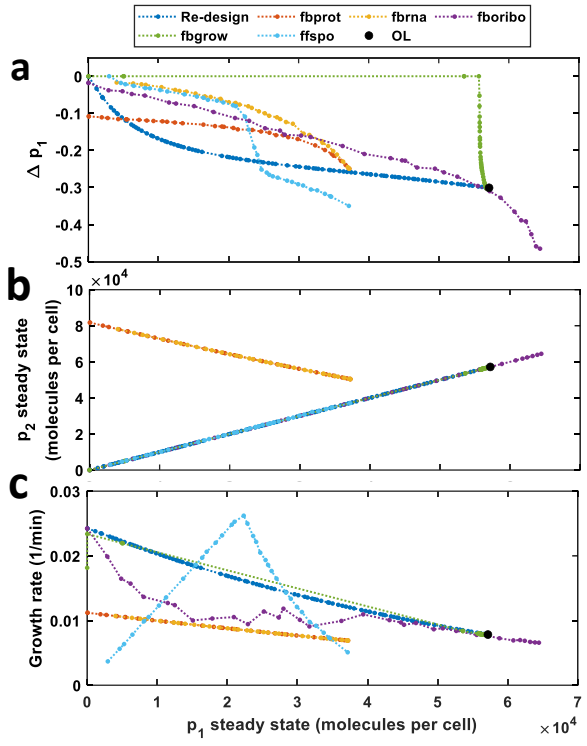

Fig. 2. Comparison of control systems performance to decouple co-expressed genes. Results of the multi-objective optimisation described in the main text. Models were given a run in time $t = [0, 10^6]$ min with $\omega_1 = \omega_2 = 0$ mRNAs per min. At $t = 0$ min., $\omega_1 = 100$ mRNAs per min. At $t = 1440$ min, $\omega_2 = 100$ mRNAs per min. The open loop circuit is shown as a single point, OL. (a) Protein p_1 coupling. Calculated from Eq. 25. (b) Disturbance size. Number of proteins of p_2 produced for the input $\omega_2 = 100$ mRNAs per min. (c) Growth rate at $t = t_{max}$.

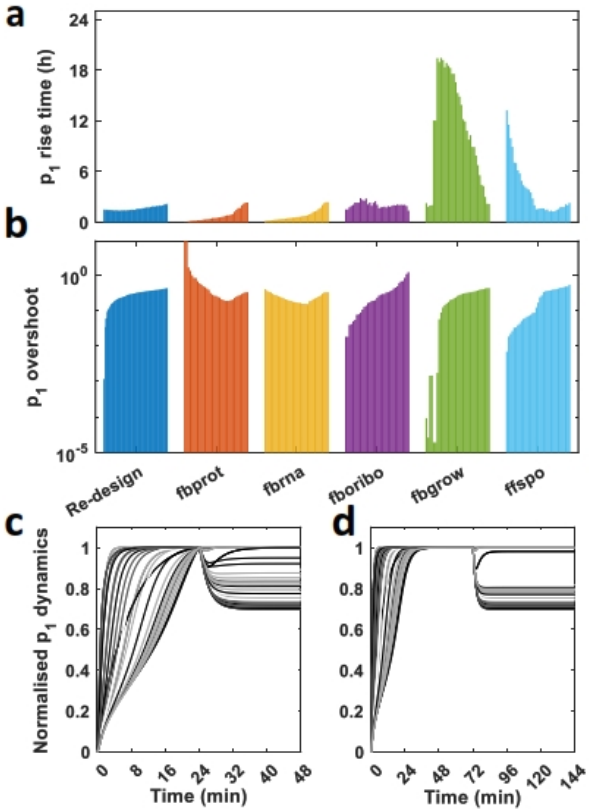

Fig. 3. Dynamic performance of the control systems. All controllers are shown in ascending order of p_1 steady state. (a) The rise time of p_1 from $t = 0$ min when ω_1 is stepped to 100 mRNAs per min. (b) The p_1 overshoot upon p_1 activation. (c) The dynamics of p_1 in the *fbgrow* and *ffspo* controller simulations. All controllers from the Pareto front in Fig. 2. p_1 concentrations were normalised to their level at $t = 24$ h. **(d)** Controllers from (c) where re-simulated with $t_{ind} = 72$ h and $t_{max} = 144$ h.

p_1 expression. This is a consequence of the greater level of p_2 expression which they allow (Fig. 2b). Note that as p_1 expression increases, p_2 expression falls - due to the lack of control of p_2 (Fig. 2b). These controllers also show the fastest response times with most controllers giving response times below hours (Fig. 3a).

The centralised controller *fboribo*, outperforms *fbprot* in terms of decoupling ability. The controller also outperforms both *fbprot* and *fbrna* in terms of growth rate due to its concurrent action on p_2 (Fig. 2b). The optimal designs of this controller show similar rise times to the slowest of the decentralised controllers (Fig. 3a). This delay is due to production of the orthogonal ribosomes and can be improved by designing the orthogonal rRNA decay rate [13]. The centralised growth-based feedback controller *fbgrow* breaks the trade-off in the objectives with some designs showing both high levels of expression and near perfect decoupling. However, this decoupling is due to a prolonged rise time; the disturbance is applied before p_1 reaches steady state (Fig. 3c). If the disturbance is applied after p_1 reaches steady state then poor coupling is observed at high p_1 steady state (Fig. 3d). Optimal designs of this controller produce high p_1 and hence show poorer growth. These controllers show long rise times (Fig. 3a, c). The ribosome-actuation system *ffspo* allows increased growth rate for the same levels of p_1 expression and decoupling at moderate levels of expression compared to all other systems. At these moderate levels of expression it shows best decoupling of all the schemes bar *fbrna*. At $p_1 > 2 \times 10^4$ molecules per cell the system shows poorer levels of decoupling, although it maintains high growth rates. As with *fboribo*, the controllers demonstrate a longer rise time than the module-specific controllers but this is still acceptable (< 12 h).

C. Re-assessing the performance of fbprot and fbrna

We re-assessed the performance of the *fbprot* and *fbrna* control schemes by applying these control schemes to p_2 in addition to p_1 and carrying out the design process again. For this study we assume that the two controllers (i.e. one controller for each module, p_1 and p_2) have the same parametrisations (e.g. $\kappa_1 = \kappa_2$). Here we find that both *fbprot* and *fbrna* outperform open loop redesign of the circuit parameters significantly for moderate levels of gene expression (Fig. 4a). The *fbrna* controller outperforms the *fbprot* controller. The controllers demonstrate a growth defect in comparison to circuit design due to the use of a second controller protein. We find that these control systems, when applied to each gene within the circuit, outperform the centralised control system *fboribo* at low steady state values of p_1.

Authorized licensed use limited to: University of Warwick. Downloaded on April 12,2023 at 09:31:25 UTC from IEEE Xplore. Restrictions apply.

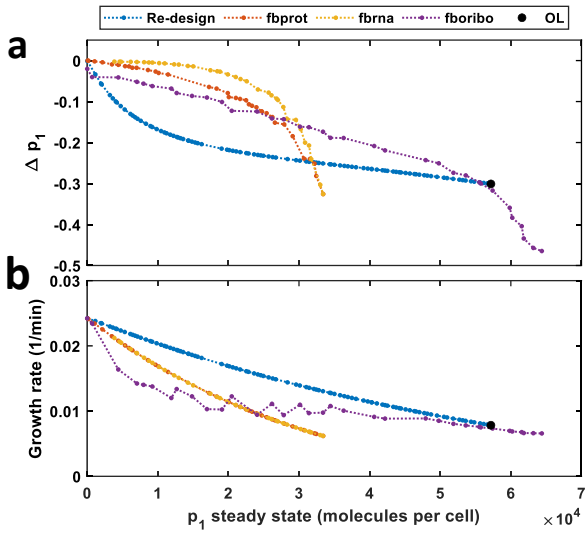

Fig. 4. Decentralised control systems for each circuit gene process. Results of the multi-objective optimisation described in the main text. Models were given a run in time $t = [0, 10^6]$ min with $\omega_1 = \omega_2 = 0$ mRNAs per min. At $t = 0$ min., $\omega_1 = 100$ mRNAs per min. At $t = 1440$ min, $\omega_2 = 100$ mRNAs per min. The open loop circuit is shown as a single point, OL. The centralised control system fboribo is shown for comparison. (a) Protein p_1 coupling. Calculated from Eq. 25. (b) Growth rate at $t = t_{max}$.

IV. Analysis of Controller Robustness

It is not possible to precisely tune controller parameters in experimental implementations of synthetic controllers *in vivo*. We carried out a Monte Carlo parameter sampling analysis to assess the robustness of each controller to parameter uncertainty. For each optimal design of each control strategy, each of the designable parameters was independently varied by drawing a random number $\tilde{\delta}$ from a uniform distribution between $-\delta$ and $+\delta$. For each parameter k_0, the perturbed parameter k is calculated by:

$$k = (1 + \tilde{\delta}) \cdot k_0 \tag{27}$$

For each controller, we assessed 1,000 uncertain parameter sets. We carried out the analysis for $\delta = 0.01, 0.05, 0.1, 0.15$ and 0.2. The uncertainty causes the points to move off the optimal Pareto front with poorer p_1 steady state and/or decreased Δp_1 (example shown in Fig. 5a). Few methods are available in the literature with which to assess the impact of parametric uncertainty on multiobjective performance criteria represented as Pareto optimal fronts. To quantify the impact of the uncertainty, we utilised a geometric approach by calculating the area bounded by all of the 1,000 controller parametrisations (Fig. 5a). Strategies with Pareto fronts which are robust to uncertainty will show small areas as the different controller parametrisations do not cause performance to move significantly. We estimated the bounded area by dividing the p_1 steady state x-axis into discrete bins. For each bin we identified the minimum and maximum Δp_1 y-axis values. This determines the upper and lower bound of the performance of the uncertain controllers. The upper bound corresponds to the original Pareto front. We estimated the total bounded area using MATLAB's *trapz* function. We choose 500 bins which is sufficient to bound 90+% of the points produced by the uncertainty analysis for all controllers

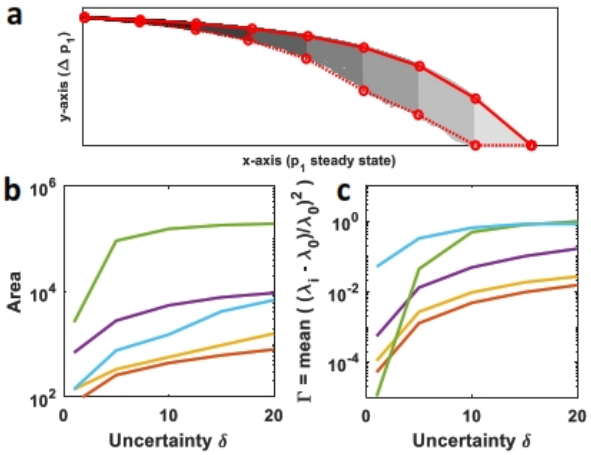

Fig. 5. Sensitivity of the Pareto front to parametric uncertainty. For each optimal controller, N random parameter sets were created as described in the main text. Colours of the controllers in panels (b) and (c) correspond to those used in Fig. 2. (a) Robustness calculation. The points are divided in bins (grey) and the maximum/minimum points identified (red circles). The upper bound of this area is shown as a solid red line and is the Pareto front. The approximate lower edge is shown as a dotted red line. (b) The robustness of the Pareto front quantified as the area of the performance of the uncertain controllers. (N = 1,000) (c) The robustness of the growth rates as measured by the Γ metric described in the main text. (N = 1,000)

(Fig. 5a). To assess the impact of uncertainty on growth, we defined the metric

$$\Gamma = \text{mean}\left(\left((\lambda_i - \lambda_0)/\lambda_0\right)^2\right) \tag{28}$$

This calculates the square of the distance of the perturbed controller growth rate from the optimal controller growth and takes the mean change across all controllers on the tested front. This analysis shows that the *fbprot* and *fbrna* control systems have better robustness across all δ than the global approaches *fboribo*, *fbgrowth* and *ffspo* (Fig. 5b).

V. Robustness to Host Variation

We also assessed the ability of the control strategies designed in Section III to confer robustness to variations in the host parameters. Host parameter variation can result from changes to physiological state (e.g. stress response changing cellular resource allocation) or use of different strains (e.g. industrial strains often have subtle changes in proteome composition due to adaptation to the industrial environment). We varied the parameters in the host model as described in Section IV. We varied key parameters including host metabolic rates (v_T, k_T, v_E, k_E), host transcription rates (ω_T, ω_E, ω_H, ω_R, ω_r, o_X, o_R), peptide elongation rates (γ_{max}, κ_γ) endogenous host feedback (κ_H, h_H). (See [10] for full definitions of parameters).

All controller topologies, bar *fbgrow*, increase robustness of the circuit to host uncertainty when compared to open loop circuit redesign (dark blue). The *fbprot* and *fbrna* control systems show the best robustness to uncertainty with a smaller area of uncertain points than other controllers (Fig. 6a). Uncertainty in the host leads to similar levels of growth changes for all control systems (Fig. 6b). It has been demonstrated experimentally that *fbgrow* maintains function when transferred between different *E. coli* strains but that

Authorized licensed use limited to: University of Warwick. Downloaded on April 12,2023 at 09:31:25 UTC from IEEE Xplore. Restrictions apply.

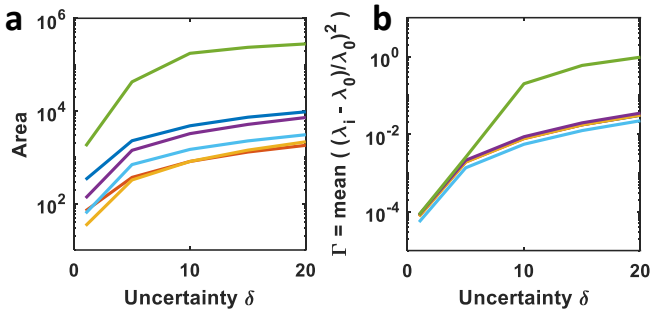

Fig. 6. Robustness to host variation conferred by the controllers. For each optimal controller, N = 1,000 random host parameter sets were created as described in the main text. (a) The robustness of the Pareto front was quantified by the area as described in the main text. (b) The robustness of the growth rates as measured by the Γ metric described in the main text.

there are significant differences in expression of the single gene of interest [11]. Analysis of our simulations shows that here poor robustness is due to the emergence of 'positive' coupling; that is both p_1 and p_2 rise upon activation of p_2 (i.e. $\Delta p_1 > 0$).

VI. Robustness to Environmental Conditions

Metabolic status in the model is determined by nutrient efficiency parameter, ϕ_e (see [10]) which determines the production rate of the anabolic driver e. In Section III, the controllers were designed assuming $\phi_e = 0.5$, equivalent to growth of *E. coli* on rich glucose and amino acid medium [10], [14]. However, in industrial fermentation cost limitations may require poorer quality/heterogeneous culture media to be used (equivalent to $\phi_e < 0.5$). We assessed the change in performance of the control strategies to poor nutrients by simulating the optimal controllers with $\phi_e < 0.5$ (Fig. 7). The *fbrna* strategy shows most robustness to changes in ϕ_e with the performance changing negligibly. The global *fboribo* shows good robustness at high levels of decoupling/low levels of expression ($\Delta p_1 > -0.25$) but is sensitive to change in ϕ_e at higher levels of expression. The global controllers *fbgrow* and *ffspo* show good robustness to changes in ϕ_e, but these effects are nonlinear. For *fbgrow*, at high p_1 increasing ϕ_e decreases p_1 but at low p_1 increasing ϕ_e leads to 'positive' coupling (i.e. $\Delta p_1 > 0$). For *ffspo*, decreasing ϕ_e shifts the front to steepen the association betwen Δp_1 and p_1 steady state.

VII. Conclusions

Our analysis shows that all control systems show a trade-off between decoupling ability and protein output with both decentralised and global control systems showing similar performance at low to moderate protein outputs. Our analysis shows that global control systems (*fboribo* or *ffspo*) enable systems to reach higher growth rates. At high protein production all systems show poor performance. Our initial robustness analysis suggests that the distributed systems show better robustness to uncertainty in the host and environmental variation, whilst also being easier to build (i.e. high robustness to circuit uncertainty). Future work needs to consider these objectives (of maximising growth while minimising coupling) concurrently.

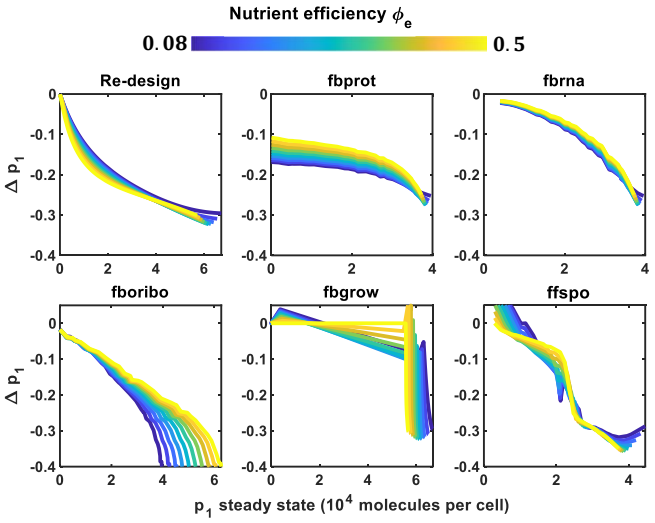

Fig. 7. Performance change upon nutrient variation. The performance of the optimal controllers designed in Section III was reassessed by re-simulating the model with a range of nutrient values $\phi_e = [0.08...0.5]$. Note the different x-axis scales.

References

[1] A. S. Khalil and J. J. Collins, "Synthetic biology: applications come of age." *Nat. Rev. Genet.*, vol. 11, no. 5, pp. 367–379, 2010.

[2] S. Cardinale and A. P. Arkin, "Contextualizing context for synthetic biology - identifying causes of failure of synthetic biological systems," *Biotechnol. J.*, vol. 7, no. 7, pp. 856–866, 2012.

[3] Y. Qian, H.-h. Huang, J. Jiménez, D. Del Vecchio, "Resource Competition Shapes the Response of Genetic Circuits," *ACS Synth. Biol.*, vol. 6, no. 7, pp. 1263–1272, 2017.

[4] J. Kim, A. P. S. Darlington, D. G. Bates, and J. I. Jimenez, "The interplay between growth rate and nutrient quality defines gene expression capacity," *bioRxiv* preprint, doi. 10.1101/2021.04.02.438188, 2021.

[5] M. Mori, Z. Zhang, A. B. Esfahani, J.-B. Lalanne, B. C. Collins, A. Schmidt, O. T. Schubert, D.-S. Lee, G. W. Li, T. Hwa, and C. Ludwig, "From coarse to fine: The absolute Escherichia coli proteome under diverse growth conditions," *Mol. Syst. Biol.*, vol. 17, e9536, 2021.

[6] F. Ceroni, R. Algar, G.-B. Stan, and T. Ellis, "Quantifying cellular capacity identifies gene expression designs with reduced burden," *Nat. Methods*, vol. 12, no. 5, pp. 415–423, 2015.

[7] O. Borkowski, F. Ceroni, G.-B. Stan, and T. Ellis, "Overloaded and stressed : whole-cell considerations for bacterial synthetic biology," *Curr. Opin. Microbiol.*, vol. 33, pp. 123–130, 2016.

[8] T. Shopera, L. He, T. Oyetunde, Y. J. Tang, and T. S. Moon, "Decoupling resource-coupled gene expression in living cells," *ACS Synth. Biol.*, vol. 6, no. 8, pp. 1596–1604, 2017.

[9] H.-H. Huang, Y. Qian, and D. D. Vecchio, "A quasi-integral controller for adaptation of genetic modules to variable ribosome demand," *Nat. Commun.*, vol. 9, e5415, 2018.

[10] A. Darlington, J. Kim, J. Jiménez, and D. Bates, "Dynamic allocation of orthogonal ribosomes facilitates uncoupling of co-expressed genes," *Nat. Commun.*, vol. 9, e695, 2018.

[11] F. Ceroni, A. Boo, S. Furini, T. E. Gorochowski, O. Borkowski, Y. N. Ladak, A. R. Awan, C. Gilbert, G.-B. Stan, and T. Ellis, "Burden-driven feedback control of gene expression," *Nat. Methods*, vol. 15, no. 5, pp. 387–393, 2018.

[12] C. Barajas, J. Gibson, L. Sandoval, and D. Del Vecchio, "Ribosome actuator via the SpoT/ppGpp pathway to mitigate gene overexpression burden," *bioRxiv* preprint, doi. 10.1101/2021.02.11.430724, 2021.

[13] A. P. S. Darlington, J. Kim, J. I. Jimenez, and D. G. Bates, "Engineering translational resource allocation controllers: Mechanistic models, design guidelines and potential biological implementations," *ACS Synth. Biol*, vol. 7, no. 11, pp. 2485–2496, 2018.

[14] A. .Y. Weiße, D. A. Oyarzún, V. Danos, and P. S. Swain, "Mechanistic links between cellular trade-offs, gene expression, and growth," *Proc. Natl. Acad. Sci.*, vol. 112, no. 9, pp. E1038–E1047, 2015.

Authorized licensed use limited to: University of Warwick. Downloaded on April 12,2023 at 09:31:25 UTC from IEEE Xplore. Restrictions apply.