UNIVERSITA DEGLI STUDI DI NAPOLI FEDERICO II
Scuola Politecnica e delle Scienze di Base
Dipartimento di Ingegneria Elettrica e Tecnologie dell' Informazione

DOCTORAL THESIS

Development and validation of synthetically engineered microbial consortia

Author:
Davide Salzano

Tutor:
Prof. Mario di Bernardo

Submitted in fulfilment of the requirements for the degree of Doctor of Philosophy in Information and Communication Technologies for Health, XXXV Ciclo. Coordinator: Prof. Daniele Riccio.

31 October 2022

A Rosaria, Ciro e Irene, che mi hanno sempre sostenuto

Abstract

Microbial consortia constitute a promising solution to expand the complexity of the synthetic gene circuits that can be engineered in living cells. However, being able to construct robust, modular communities is a challenging task. In addition, due to the imbalances in the growth rates between different members of the consortium, some populations in the consortium can outgrow the others, leading to undesired phenomena, such as extinction, undermining the correct operation of the system.

In this work we address both the problem of realizing reliable microbial consortia and the challenge of regulating the relative numbers between subpopulations in the community to their ensure stable, and long term coexistence. Specifically, we realize and validate biomolecular feedback controllers distributed across populations within a microbial consortium to achieve modular and robust regulation of a cell's phenotype. These results are complemented by the development of control architectures to shape the composition of a microbial consortium.

The realisation of consortia with robust and consistent phenotype where it is possible to decide an maintain a specified composition would enable engineering of complex functionalities in living cells. Potential applications include personalized cell therapies and on demand production of chemicals, drugs and bio-fuels for pharmaceutical or commercial use.

Acknowledgments

This study and research journey has been a motivating, inspiring and enthralling experience, especially thanks to the many wonderful people that supported it.

A special thanks goes to my main supervisor Mario di Bernardo, both for supporting the research with great technical feedback and insightful ideas, and for motivating me along the path, gently pushing me to become the best researcher I could. Equally, I wish to acknowledge my supervisors from the University of Bristol: Lucia Marucci for her encouraging technical support and caring words, Nigel Savery for the immense opportunity he gave me hosting me in D40 and his outstanding guide towards the development of wet lab skills, and Claire Grierson for all her valuable feedback. I am also particularly grateful to Davide Fiore for his collaboration in most of the matters in this thesis. His comments and feedback were precious, his guidance motivated me greatly, and I hope to have the pleasure to continue working with him in the future. Furthermore, I want to thank all the members of the SINCRO group in Naples and of D40 in Bristol for making the workplace a fun place. Among those, a special thanks goes to Eugene, Harry, Hattie and Arthur for being the most caring friends I could have ever hoped for. I wish to thank all the collaborators that contributed to the research: prof. Fabio della Rossa, prof. Diego di Bernardo, prof. Joseph Olm, prof. Giovanni Russo, prof. Ludovic Renson, dr. Barbara Shannon, dr. Davide Liuzza, dr. Criseida Zamora, Virginia Fusco, Vittoria Martinelli, Agostino Guarino, Irene de Cesare, Iacopo Ruolo and Sara Napolitano. Lastly, I want to thank all the dear friends for their loving support, my partner Irene for helping me overcome each and every obstacle and for being the most caring, supporting person on the planet, and to my family for always believing in me.

Contents

Chapter 1

Introduction

Synthetic biology aims at engineering biomolecular systems to achieve new useful functionalities [1]. Potential applications range from designing bacteria that can produce biofuels or sense and degrade pollutants in the environment (like hydrocarbons and plastics), to immune cells that can track and kill cancer cells, or that can release drugs at specific points and conditions to avoid side effects (see [1] for references). This is made possible by designing synthetic gene circuits with programmed functionalities and embedding them into living cells. However, the level of complexity and the functions of such engineered genetic circuits are limited by intrinsic factors in the host cells, such as excessive metabolic burden, competition for limited resources and incompatible chemical reactions [2]. To overcome these limitations, a promising strategy is to distribute the required functionalities among multiple cell populations forming a microbial consortium, so that each cell strain embeds a smaller subset of engineered gene networks [3, 4, 5, 6]. In this way, each cell population carries out a specialised function and, by dividing labour with the others in the consortium, contributes more efficiently to the achievement of the overall final goal.

However, engineering reliable and robust cellular communities remains a challenging task [2]. Control Engineering, which aims at improving the stability, robustness, and performance of physical systems, may provide an invaluable contribution towards the realization of reliable and robust modules. Specifically, using mathematical models to predict the interaction between parts and applying concepts from control theory to guarantee desired static and dynamic performance to the circuits, it is possible to drastically decrease the development time of new multicellular systems and to increase their flexibility and modularity. In recent years, research efforts have been focused on model-based design and implementation of biochemical *controllers* to engineer robust and modular synthetic circuits carrying out desired functions [7, 8, 9].

In addition, the development of microbial consortia also presents a unique challenge; cells expressing different genes might also grow and divide at different rates. In particular, cells in the consortium, whose function is associated with a lower metabolic burden, will grow faster, eventually becoming dominant over the other populations, thus compromising the overall function of the consortium and giving rise to undesired spatiotemporal dynamics, such as oscillations, or even extinction [10, 11]. Therefore, it is crucial to develop methods to guarantee the stable coexistence between cell populations in a consortium by regulating and maintaining their relative numbers to some desired level, adjusting them to the requirements of the specific application of interest. This is possible, as suggested in [12], by using *feedback control algorithms* able to sense the relative size of all the populations involved and respond by applying appropriate stimuli to the cells to regulate their relative numbers. This problem is known as *ratiometric control* of cell populations [12] as its overall goal is to achieve and maintain a certain desired ratio between the size of the populations in the consortium, despite differences in their growth rates, noise and perturbations.

This Thesis addresses the two main open problems limiting the expansion of synthetic biology to a consortium level. Specifically, we engineer and validate *in vivo* a *multicellular* controller for the development of robust, reliable consortia. Also, we propose two different approaches to the problem of regulating numbers in bacterial populations, providing possible solutions to the *ratiometric* control problem.

1.1 Outline of the Thesis

- In Chapter 2 we give a brief introduction on synthetic microbial consortia, describing their constitutive components and the main approaches used for their development. In addition, we discuss a promising approach to speed up their design and improve the reliability of the designed consortia;

- In Chapter 3 we describe the three main approaches used to interface control theory and synthetic biology, providing examples for each of them. For each approach, the main advantages, limitations, and applications are expounded. The results of the literature review appeared in a paper in *Current Opinions in Systems Biology* [13], written in collaboration with prof. Diego di Bernardo[1], Iacopo Ruolo[1] and Sara Napolitano[1];

- In Chapter 4 we propose a possible biological implementation of a multicellular biomolecular feedback controller, where the control functionalities are distributed across different populations comprising a microbial consortium. Furthermore, we engineer and

---

[1] Telethon Institute of Genetics and Medicine (TIGEM), Pozzuoli, Italy

test *in vivo* the design, assessing the performance, reliability and robustness of the architecture;

- In Chapter 5 we discuss possible applications and extensions of the *multicellular feedback control* described in Chapter 4. Firstly, we use the multicellular paradigm to design and validate *in silico* a three strain microbial consortium where two *controllers* decide the phenotype expressed by a reversible bistable memory element. Secondly, we design and test *in silico* an extension of the multicellular controller, adding a third population in the consortium that implements a Proportional action to improve the dynamical control performances. The results were reported in two journal papers [5, 14]. The former was written in collaboration with prof. Josep M. Olm[2] and Davide Fiore[3], and the latter in collaboration with Davide Fiore[3] and Vittoria Martinelli[4];
- In Chapter 6 we address the problem of regulating the relative numbers of populations within microbial consortia. We describe two alternative solutions: one uses external stimuli to balance relative numbers in a mono-strain microbial consortium; the other uses an *embedded* controller to regulate the growth rate of a single cellular population. Both solutions appeared in journal papers published on the *Journal of the Royal Society Interface* [15] and on the *International Journal of Robust and Nonlinear Control* [16], respectively. The former was written in collaboration with Davide Fiore[3], and the latter with Davide Fiore[3] and Virginia Fusco[4];
- In Chapter 7 we summarize all the results elaborated in the Thesis and discuss open problems and possible future work.

We complete the Thesis with three appendices. In the first we insert all the additional information related to chapters 4-6. The second appendix contains an accurate description of BSim, an agent based simulator we used to validate *in silico* the strategies presented in this Thesis. In addition, we present three case studies where the simulation environment was used to validate *in silico* the effectiveness of control and continuation algorithms. The results appeared in *ACS Synthetic Biology* [17, 18, 19]. [17] was written in collaboration with Agostino Guarino[4] and Davide Fiore[3], [18] was written in collaboration with Dr.Barbara Shannon[5], Dr. Criseida Zamora[6], Prof. Nigel J. Savery[5], Prof. Lucia Marucci[6] and Prof.

---

[2] Institute of Industrial and Control Engineering, Universitat Politècnica de Catalunya, Barcelona, Spain

[3] Department of Mathematics and Applications 'R. Caccioppoli', University of Naples Federico II, Naples, Italy

[4] Department of Electrical Engineering and Information Technology, University of Naples Federico II, Naples, Italy

Claire S. Grierson[7], and [19] was written in collaboration with Irene de Cesare[6], Prof. Lucia Marucci[6] and Prof. Ludovic Renson[8]. Lastly, the third appendix includes the description of an *ad-hoc* identification procedure developed to parametrize a model describing the spread od Covid-19 in Italy, used to design distributed non-pharmaceutical interventions to limit the spread of the pandemics. This research appeared in a journal paper in *Nature Communications* [20]. It was carried out with all the members of the SINCRO group[4], with Prof. Fabio della Rossa[9], Dr. Davide Liuzza[10] and Prof. Giovanni Russo[11].

1.2 Publications

The research done during the duration of the PhD was published in the following papers:

1. Agostino Guarino, Davide Fiore, Davide Salzano, and Mario di Bernardo. Balancing cell populations endowed with a synthetic toggle switch via adaptive pulsatile feedback control. ACS synthetic biology, 9(4):793–803, 2020.

2. Barbara Shannon, Criseida G Zamora-Chimal, Lorena Postiglione, Davide Salzano, Claire S Grierson, Lucia Marucci, Nigel J Savery, and Mario di Bernardo. In vivo feedback control of an antithetic molecular-titration motif in escherichia coli using microfluidics. ACS Synthetic Biology, 9(10):2617–2624, 2020.

3. Fabio Della Rossa*, Davide Salzano*, Anna Di Meglio*, Francesco De Lellis*, Marco Coraggio, Carmela Calabrese, Agostino Guarino, Ricardo Cardona-Rivera, Pietro De Lellis, Davide Liuzza, Francesco Lo Iudice, Giovanni Russo, Mario di Bernardo. A network model of Italy shows that intermittent regional strategies can alleviate the COVID-19 epidemic. Nature communications, 11(1), 1-9, 2020.

4. Davide Fiore, Davide Salzano, Enric Cristòbal-Cóppulo, Joseph M. Olm, Mario di Bernardo. Multicellular feedback control of a genetic toggle-switch in microbial consortia. IEEE Control Systems Letters, 5(1), 151-156, 2021.

---

[5] School of Biochemistry, University of Bristol, United Kingdom.

[6] Department of Engineering Mathematics, University of Bristol, United Kingdom.

[7] School of Biological Sciences, University of Bristol, United Kingdom.

[8] Department of Mechanical Engineering, Imperial College London, United Kingdom.

[9] Department of Electronic, Information and Biomedical Engineering, Politecnico di Milano, Milan, Italy

[10] ENEA, Fusion and Nuclear Safety Department, Frascati, Rome, Italy

[11] Department of Information and Electrical Engineering and Applied Mathematics, University of Salerno, Fisciano, Italy

5. Iacopo Ruolo*, Sara Napolitano*, Davide Salzano, Mario di Bernardo, and Diego di Bernardo. Control engineering meets synthetic biology: Foundations and applications. Current Opinion in Systems Biology, 8:100397, 2021.

6. Virginia Fusco*, Davide Salzano*, Davide Fiore, and Mario di Bernardo. Embedded control of cell growth using tunable genetic systems. International Journal of Robust and Nonlinear Control, 2021.

7. Vittoria Martinelli*, Davide Salzano*, Davide Fiore, and Mario di Bernardo. Multicellular PI control for gene regulation in microbial consortia. IEEE Control Systems Letters, 6, 3373-3378, 2022.

8. Davide Salzano, Davide Fiore, and Mario di Bernardo. Ratiometric control of cell phenotypes in monostrain microbial consortia. Journal of the Royal Society Interface, 19(192):20220335, 2022.

9. Irene de Cesare, Davide Salzano, Mario di Bernardo, Ludovic Renson, Lucia Marucci. Control-Based Continuation: A New Approach to Prototype Synthetic Gene Networks. ACS Synthetic Biology 2022 11 (7), 2300-2313, 2022.

---

* These authors contributed equally

Chapter 2

From engineered to Controlled consortia

2.1 Key ingredients of microbial consortia

Microbial consortia require multiple populations to coexist and to coordinate their behavior in order to achieve a common goal. Their design and implementation calls for engineering and combining new gene regulatory networks to achieve cell-to-cell communication and regulate the composition of the communities, as well as their collective behaviour [21].

2.1.1 Gene regulatory networks as building blocks

Synthetic gene regulatory networks (GRNs) are engineered genetic circuits that can alter the phenotype of a target host [22]. They constitute the basic building blocks of any synthetic biological system, as they are the mean to either modify or add functionalities in living cells (Figure 2-1).

The earliest examples of Synthesized GRNs were a reversible memory mechanism [23] and a genetic oscillator [24], dating from the early 2000's. Since then, the functional complexity of the systems that have been built increased significantly, going from simple circuits consisting of a single module to more complex designs made of many interacting parts [25]. Specifically, the increasing effort put in the characterization of biological components constituting the GRNs and the advancements in technologies for their construction allowed for more and more sophisticated architectures to be built, including communication elements that enabled both different cells and strains to coordinate and to self organize in well shaped compositions.

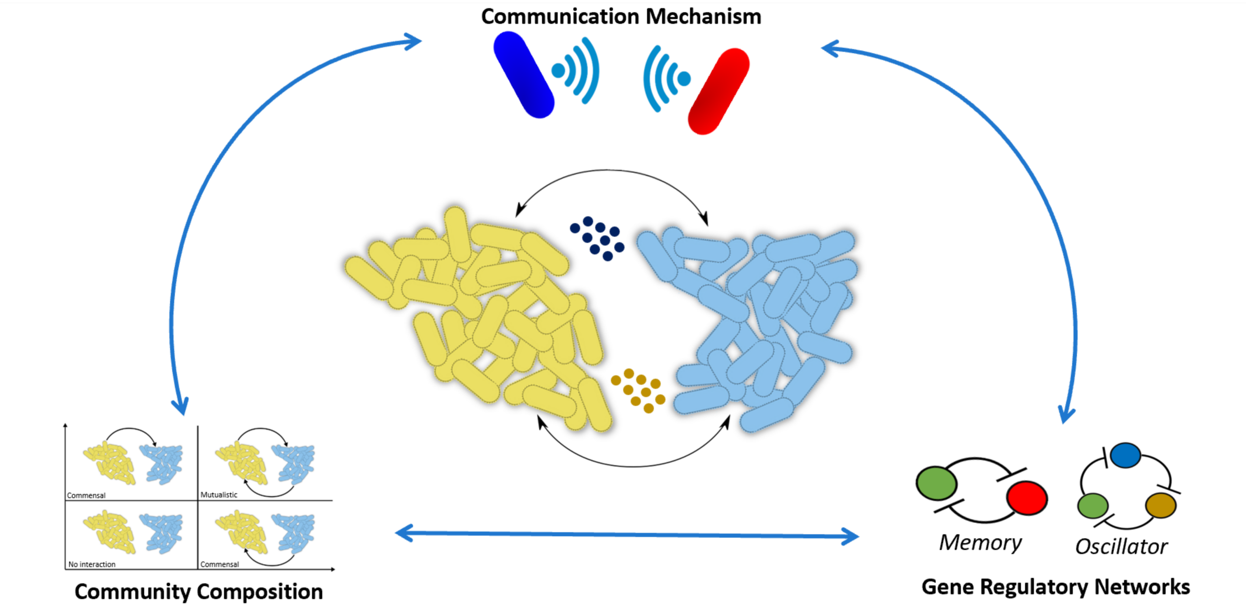


Figure 2-1: Fundamental elements of synthetic microbial consortia. The solid blue arrows represent functional relationships/dependencies between different core elements. Specifically, Gene regulatory networks are the mean either to build communication mechanisms or to construct ecological relationships, shaping the community composition. The Communication mechanisms are the functional modules required for the implementation of ecological relationships among different cellular populations and can also be beneficial in the exchange of information necessary to coordinate the community. Finally, Community composition influences both the effectiveness of the communication protocols and the reliability of the gene regulatory networks synthesized in the consortium.

2.1.2 Cell-to-Cell Communication

The passage from modules to systems was enabled by the rational engineering of synthetic inter and intra-sepcies communication mechanisms. Cell-to-cell communication is what distinguishes single cell modules from multicellular systems. It enables the exchange of information within and between cellular populations, as well as enabling the creation of metabolic dependencies across populations in the consortium. Common communication mechanisms exploited in synthetic biology are cross-feeding, quorum sensing, conjugation and toxin-antitoxin systems [26].

Metabolic cross-feeding is based on the creation of metabolic cross-dependencies between different cellular populations. An early example of a synthetic cross-feeding system was presented by Shou et al. in [27]. Here, two yeast populations were made mutually dependent, allowing each strain to produce a nutrient necessary for the other strain's survival. It was shown that mutual dependency between the populations guarantees coexistence of both strains. A similar effect can be obtained by using toxin-antitoxin systems. Specifically, it is possible to entangle the faith of different species, allowing their survival only when both the toxin and antitoxin are produced. In [28], a feedback control based approach, employing a toxin-antitoxin mechanism, was designed to balance the numbers of individuals within a consortium.

Instead, conjugation and quorum sensing are usually exploited when information about the current state of one or more populations needs to be shared. Specifically, conjugation allows for the exchange of specific information, in the form of DNA [29]. Quorum sensing, on the other hand, enables sharing of population density-based information in the form of small molecules. Among these two communication mechanisms, the most leveraged sensing mechanism is quorum sensing. Using this mechanism, each cell broadcasts information about its population density/state via the production of small molecules, eventually triggering a response in other cells/populations. This mechanism can be used to orchestrate the collective behaviour of a consortium, as done in [30]. Here, an oscillator was realized using two *E. Coli* strains exchanging information about each other state using quorum sensing.

The development of communication mechanisms is crucial for the coordination of different modules. In addition, the composition of multiple modules may require the engineering of multiple orthogonal communication channels. However, the creation of such channels may be challenging due to phenomena such as pathway cross-reactivity, the slowness of diffusion, the need of very high cell densities for them to be effective, and the highly non-linear responses of the protein sensors used to monitor these signals. Some works in the literature addressed these key challenges. For example, in [31] the interaction between multiple quorum sensing molecules was characterized and in [32] the spatial effects on quorum sensing signaling were analysed in microfluidics.

2.1.3 Community Composition

Communication itself does not guarantee the formation of a consortium. The second fundamental ingredient is the possibility to regulate the composition of the communities, granting coexistence, as well as the correct operation of the information exchange mechanism. This can be achieved either by designing specific interaction patterns between different species in the consortium or by manipulating the growth environment [26]. For example, membrane separated culture environments can be used to preserve spatial separation between different populations, allowing at the same time the exchange of small molecules required for the exchange of information. Other common environments apt for the culture of multiple species are microfluidic platforms and chemostats architectures (Table **2-1**). The former provides a controlled growth environment where small cell populations can be grown in exponential phase [33]. Intead, turbidostats provide structures where large populations can be grown [34]. However, none of these technologies can guarantee that the coculture is preserved for a long time, even if some research to adapt chemostats and microfluidic platforms for long term coexistence of microbial populations has been performed [35, 36, 37].

An alternative approach for the achievement of long term coexistence between populations in a consortium is constituted by the implementation of synthetic ecological relationships. Rationally engineering relationships such as prey-predation, commensalism and competition allows communities to self regulate their relative population numbers [38].

2.2 Microbial Consortia orchestrate the collective behaviour of microbial communities

Synthetic biological systems have often limited functional complexity due to technological and biological factors, such as metabolic burden, incompatible reactions and retroactivity [39]. A possible solution to overcome these limitations is to engineer multicellular consortia, whose aim is to orchestrate the collective behaviour of microbial communities. By distributing the functional complexity across multiple species, it is possible steer the communities to a desired phenotype, auto balance population numbers, and achieve specific spatiotemporal patterns (see Table **2-1** for examples).

Splitting the pathways among populations may address the problem of excessive metabolic burden on the host organism, allowing for the construction of more complex synthetic pathways [40, 30, 41, 42]. For example, Chen et. al. [30] developed a multicellular oscillator using two *E. Coli* strains, communicating using two orthogonal quorum sensing pathways. Here, the oscillatory behaviour is obtained using an activator and a repressor strain. The activator strain boosts its expression and the expression of genes in the other strain. Conversely, the repressor strain inhibits both the expression of its characterizing gene and the one of the other strain. It is shown that the combination of a positive and a negative feedback leads to the emergence of oscillations. In this work, the co-culture is obtained by seeding accordingly microfluidic devices with both strains. Note that, even if it is possible to allow for coexistence for a period long enough to observe oscillations, coexistence between the two strains cannot be maintained indefinitely.

Microbial consortia can be also a powerful tool to avoid reactions that are incompatible either with the host organism or with other engineered modules. An example pathway split to avoid incompatible reactions, as well as optimize the production of chemical compounds can be found in [43]. Here the production of Oxygenated taxanes is carried out using an *E. Coli* and a *S. Cereviciae* stain. Specifically, the *E. Coli* strain is used to produce taxanes. However, this organism is not able to oxigenate them, therefore an S. Cereviciae strain is used to further process the taxanes, oxygenating them. In this architecture, each step of the process hosted in the most suitable host, allowing for an increased production rate and a modular optimization.

In both of the illustrated examples, coexistence between species is a fundamental requirement for a correct operation of the consortium. This can be achieved by designing specific interaction patterns among the populations constituting the consortium [26, 44, 45, 46, 47, 48, 16]. For example, in [49] a prey predator interaction pathway was engineered, showing how it can ensure the cohabitation of two species. Similar strategies have been also proposed in [50] and [45]. Alternatively, it is possible to operate on the growth environment of the populations to guarantee coexistence [12, 37, 51]. For example, in [37] the dilution rate of a chemostat was regulated to shape the composition of a two population consortium.

For some applications, in addition to coexistence, it is desirable that cells are organized in specific spatial patterns. This behaviour can emerge by designing *ad-hoc* interaction between populations. One of the first examples where spatial patterns have been engineered is proposed in [52], where a sender and a receiver strain communicate so as activation of the receiver is possible only at intermediate distance from the senders, generating ring-like spatial patterns. In [53] a review on the recent advancements on pattern formation achieved using multicellular synthetic systems can be found.

2.3 Engineering microbial consortia by *ad hoc* biological selection

Microbial consortia are complex systems, whose design, in the vast majority of the applications, involves the careful selection of the most suitable biological components and their combination in an often *ad hoc* procedure.

The usual pipeline for the development of a microbial consortium starts with the construction of the intended genetic circuit by using available characterized biological parts. This process requires a refinement procedure aimed at characterizing the interacting pathways and optimizing the circuit being realized (Figure **2-2**a). The refinement can be necessary due to the insurgence of unexpected phenomena, revealed during the characterization of the circuits. This process is iterated until no further re-engineering of the consortium is needed

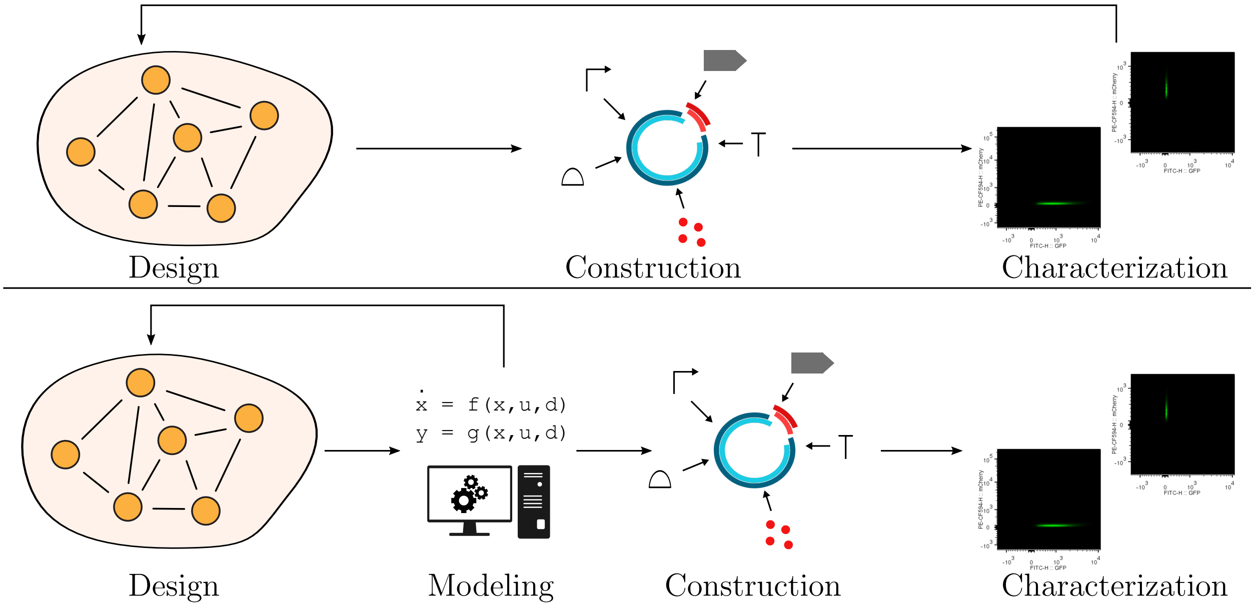


Figure 2-2: Design approaches for the the development of a microbial consortium. The top panel represents the typical approach followed for the construction of a microbial consortium. Starting form the biological design, the consortium is then constructed and testad in vivo. If the in vivo calidation phase fails, then a re-deign of part or the entire consortium is needed and the contruction and validation phases need to be performed again. Instead, the bottom panel shows the design approach for a controlled consotium. First, a design based on control theory is developed. Then a mathematical model is developed and the functionlaities are tested in silico. If the design behaves as expected then the consotium is built and tested in vivo. If not, a redesign of the scheme is performed and the model is adapted to the new design.

for its optimal operation. The examples illustrated in section 2.2 follow this paradigm. Indeed, in [43] a S. Cereviciae strain is engineered to oxigenate taxadiene and an *E. Coli* strain responsible for taxadiene production and observed that the taxadiene production efficiency was low. Hence, they refined the pathways by optimizing the interaction between species, which resulted in an increased production of the compound.

The design approach described defines a simple and direct procedure for the *in vivo* implementation and validation of synthetic microbial consortia. However, the process described is highly specific, and even minor modifications of the required behaviour may require a full re-engineering of the consortium.

2.4 From engineered to controlled microbial consortia

The core limitation of the current design methodology for the development of microbial consortia is the ability to predict the behaviour of interacting characterized components and to engineer reliable and flexible modules. This issue can be addressed with the combined use of mathematical models and control theory-based circuit designs. Specifically, when combining biological parts together, it is usually hard to predict the response of the entire network. This is because factors such as network interactions, feedback loops and local DNA sequence contact can lead to the emergence of unexpected behaviours. For example, closing a loop using a chain of three repressors results in oscillations in gene expressions [24]. To tackle this problem it is possible to develop a mathematical framework and a set of dynamical models to predict the response and the interactions of different characterized parts or modules. In [54], a mathematical model was used to capture all the possible phenotype expressed at a population level, aiding the authors to design experimental conditions without extensive *in vivo* testing. The predictions devised culture conditions that guaranteed coexistence between species and the correct operation of the consortium. The same could have been achieved heuristically, without the use of a mathematical model. However, the design process would have been slower, and it would not have been possible to predict the behaviour in conditions different from the one tested.

The biggest challenge in the design of mathematical models for genetic networks lays in their high complexity, in the inherent stochastic nature of the processes, and the spatial distribution of the biological components [1]. Currently, the vast majority of the available models rely on Ordinary Differential Equations (ODEs), and fewer of those models take into account either stochasticity of the processes or the spatial dynamics of the biochemical compounds involved in the protein synthesis (for more details see Table **2-1**). The development of mathematical models represent also an opportunity to test quickly the viability of some designs, isolating the conceptual problems from the ones caused by an actual biological construction. In addition, performing sensitivity analyses on the parameterized models offers a way to estimate the robustness of the circuit design proposed, accounting for possible mismatches between the predicted dynamics and the biological response of the system. For example, mathematical models have been used to develop and validate proof of concept designs for the control of gene expression in a Genetic Toggle Switch either using a single cell [55] and a multicellular [5] design.

The development of mathematical models and of a general mathematical framework to deal with interacting cultures and gene regulatory networks can be invaluable to design predictable systems and to speed up the design process of both single cell modules and microbial consortia. However, for the construction of reliable modules that can be used and combined in complex architectures mathematical models alone do not suffice. We need some guiding principle to devise structures robust to changes in internal and external perturbations.

Feedback control systems allowed the construction of reliable and robust systems in many application field [56] and many efforts have been put in place in the recent years into interfacing control theory with synthetic biology [57], leading to the emergence of a novel field denoted as cybergenetics. With respect to the *ad hoc* selection of biological components, the communities constructed using feedback controllers are more robust and reliable, making them perfect modules for combination in larger systems. Also, the presence of a reference signal to change the operating condition of the circuit increases the flexiblity of the designed consortia.

The combination of control theory to find network topologies and of mathematical modeling to predict the behavior of interacting circuits can drastically speed up the process of developing microbial consortia and reduce the extension of modifications needed in case of changes in the functional requirements for the consortium.

2.5 Discussion

Microbial consortia represent an outstanding tool to fully realize the potential of synthetic biology [58]. However, their design is still complex and often an *ad hoc* process. A possible solution to optimize the reliability and the speed of microbial consortia design process resides in the use mathematical models and feedback control.

Mathematical models allow predict the behaviour of a community without the need of building it. This can be translated in less refinement steps once the gene regulatory networks have been assembled. Feedback control on the other hand can address the problem of the lack of modularity and the problem of retroactivity [39]. Indeed, self-regulating genetic circuits can constitute new building blocks thanks to their reliability and robustness, which can turn invaluable for the implementation of complex, large scale synthetic microbial consortia. Moreover, the ability of feedback control to reject disturbances and unmodelled dynamics is perfect to tackle phenomena such as corss-reactivity and retroactivity, always present to a certain extent in the development of synthetic communities.

The refinement of the mathematical framework and the implementation of feedback control strategies inside living cells open to a new paradigm in the design of microbial consortia, paving the way for a faster construction of more robust and reliable communities. This in turn will allow for the development of functionally more complex consortia, allowing their employment in the bioproduction of materials at an industrial level and, possibly, in living therapeutics at a clinical level.

Description	*In vivo* validation	Modeling
A prey predator synthetic ecosystem [49]	Chemostat	ODE+white noise
AND gate with 2 populations [40]	Plate	ODE
cross-feeding in yeast populations [27]	Plate	No
Oscillator between 2 populations [30]	Microfluidics	ODE
Toggle switch over 2 populations [42]	Microfluidics	ODE
Stable coculture using lysis [54]	Microfluidics	ODE+PDE
Bioproduction with 2 hosts [43]	Chemostat	No
Resistance to mutation using 3 strains [59]	Microfluidics	ODE
Communication in Gram+ hosts [60]	Plate	No
XOR gate construction [41]	Plate	No
Pattern formation [52]	Plate	ODE
Methil Hilades production [61]	liquid culture	No
Isobuthanol production [62]	Plate	ODE
Population Regulation via dual feedback control [63]	No	ODE
Multicellular control strategy [64]	No	ODE+PDE(Diffusion)
Population composition Control [12]	No	SDE+PDE
Optimal differentiation for bioproduction [65]	No	ODE
Growth control using external inducers [46]	Liquid culture	ODE
Producton of ehanol [66]	Batch	ODE
Multicellular control strategy [67]	No	ODE
Growth control using graph theory [28]	No	ODE+PDE(Diffusion)
Differentiation control [68]	No	ODE
Oscillator between 2 populations [35]	Microfluidics	ODE
Coexistence regimes [69]	No	PDE
Synchronization of genetic clocks [70]	Microfluidics	ODE+PDE
Ecological relationships in microbial communities [38]	Batch	No
Environment control for co-cultures [71]	Batch	ODE
Multicellular XOR Gate [29]	No	SDE

Table 2-1: Examples of developed synthetic microbial consortia. The table highlights the key finding of each consortium, the methods used for its in vivo characterization and the mathematical models used to design or describe the behaviour of the community

Chapter 3

Cybergenetics: where control theory meets synthetic biology

3.1 An introduction on cybergenetics

Control engineering aims to improve the stability, robustness, and performance of physical systems. Its application to synthetic biology focuses on the design and implementation of controllers to engineer robust and modular biomolecular circuits, carrying out desired functions in the cell. The discipline at the interface between control engineering and synthetic biology has been termed as *cybergenetics* [7]. The central tenet of cybergenetics is feedback control, consisting of a *plant* (i.e. the biological process to be regulated) and a controller that regulates its behavior, see Figure **3-1**a. Specifically, the controller receives measures of the plant output from the sensors, compares it with a reference signal, and computes the input to be applied to the plant (via actuators) to decrease the mismatch between the sensed output and the reference signal. The biological process (the *plant*) is usually abstracted as an input-output dynamical system, whose dynamics is parametrised using experimentally observed data (Figure **3-1**b). The model is then used to design and numerically validate an appropriate control strategy to steer its behaviour. The candidate control law is then implemented (as biomolecular processes or a computer software) and experimentally tested. Three categories of control strategies have been so far reported in synthetic biology: embedded, external and multicellular controllers.

In the embedded strategy, the control action is exerted by a synthetic gene regulatory network embedded within the same cell hosting the process to be controlled, as shown in Figure **3-1**c. In this way, cells can self regulate their phenotype. However, extra metabolic burden may be introduced. In addition, when the control strategy is encoded within the same cell as the controlled process, the system lacks modularity as any change in the controller design requires a complete re-engineering of the synthetic gene regulatory network.

A different approach is that of external control, as shown in Figure **3-1**d, where a computer is interfaced to the biological process in the cell with actuators (pumps, optoelectronic devices, or syringes) to provide the control inputs to the cells. Here, with a minimal additional load on the host organism it is possible to achieve accurate and fast regulation of gene expression, at a cost of the necessity for a tightly controlled growth environment.

A promising approach to release metabolic burden on host cells and make the controller modular is represented by multicellular control, Figure **3-1**e. In this architecture, two different populations within the same cellular consortium co-exist, with one population encoding the controller and the other the biological process to be controlled.

3.2 Embedded Control

The embedded controllers are characterized by having both the process and the controller present within the same cell (Figure **3-2**a). The controller is implemented as a synthetic gene network in the cell, somewhat resembling the endogenous regulatory and signalling pathways that have naturally evolved to control cell behaviour [7]. Embedded controllers can be used to enhance the performance and robustness of synthetic genetic circuits to different intrinsic and extrinsic perturbations (noise, parameter uncertainty, loading, fluctuations in nutrients) [75]. Cells engineered with embedded controllers can be easily deployed also in large-scale bioreactors, for example, to improve the production of molecules in industry. The biggest hurdle in designing an embedded controller however is its practical realisability, that its biological implementation, as there is currently no formal approach to go from an abstract model of the controller to the set of biochemical reactions required to make it work. Recent reviews on this control architecture can be found in [7, 75, 76].

Integral feedback control is among the most popular network topologies used for embedded controllers. It belongs to the class of proportional integral derivative controllers (PID). PID controllers are the most widely adopted controllers in all engineering fields because they couple simplicity in implementation with a robust performance. In this type of controller, the control error, that is the deviation of the output (e.g. protein level) from its desired reference value is measured and summed over time, so that even small but persistent deviations can be sensed and used to deliver an appropriate control input to reestablish the correct output value. As such, integral controllers ensure that the control error becomes zero at steady-state, that is the output will be exactly equal to the reference value. This property is referred to in biology as perfect adaptation. Integral controllers can thus be used in a cell

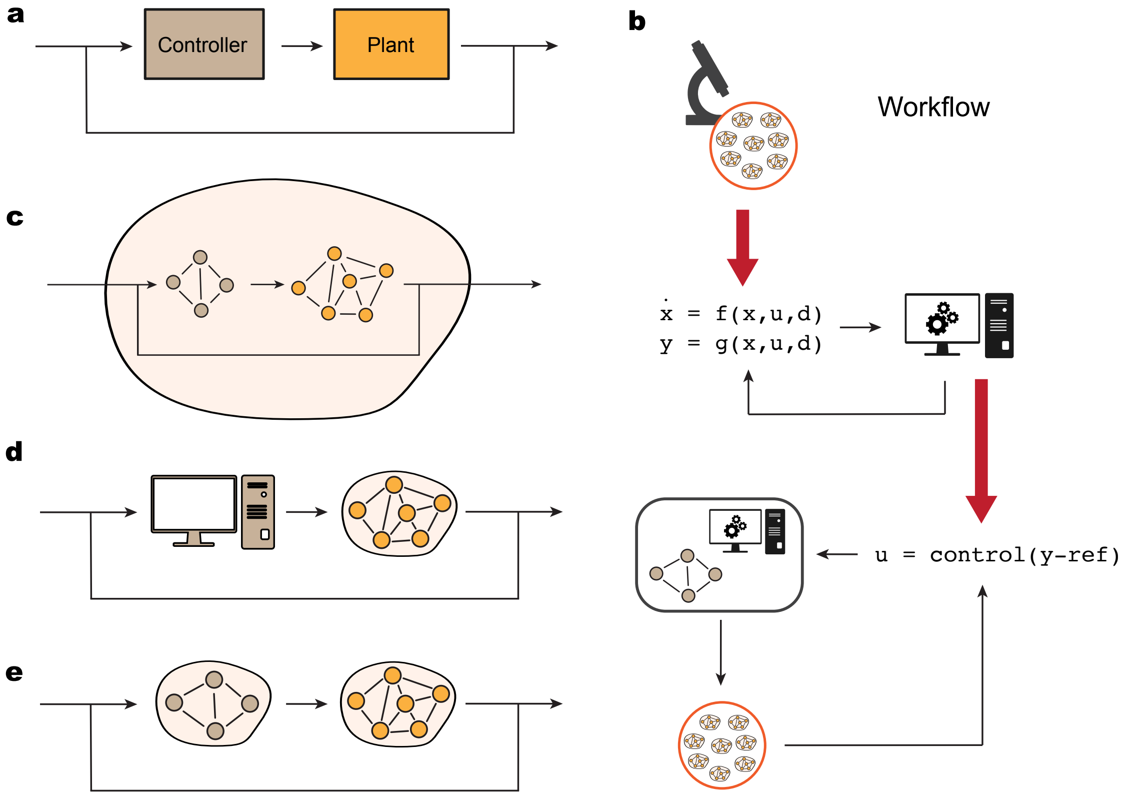


Figure 3-1: control engineering in synthetic biology. (**a**) General feedback control architecture where a controller steers the behaviour of a plant adapting the control input to the plant response. (**b**) Workflow for application of control theory to synthetic biology. (i) The biological process is experimentally characterised in open-loop experiments. (ii) A dynamical model is inferred from the acquired experimental data and used to test *in silico* different control strategies. (iii) The control law is implemented (as biomolecular processes or a computer software) and experimentally tested. (**c**) Embedded control: both the controller and the process to be controlled are present within the same cell. (**d**) External control: the control algorithm is implemented as a computer software and interfaced with cells by dedicated actuators. (**e**) Multicellular control: two cell populations coexist in a consortium, with one population embedded with a controller driving the second population harbouring the biological process to be controlled.

to maintain the homeostasis of a biological process rendering it robust against intrinsic and extrinsic perturbations.

Implementation of input-output integral controllers in living cells is currently the subject of intense investigation. A genetic network for integral control was theoretically proposed in [72] and named the 'antithetic' controller (Figure 3-2b). The authors showed that under a set of assumptions, including no degradation of the molecular species making up the controller, this biomolecular circuit is capable to achieve perfect adaptation. The antithetic

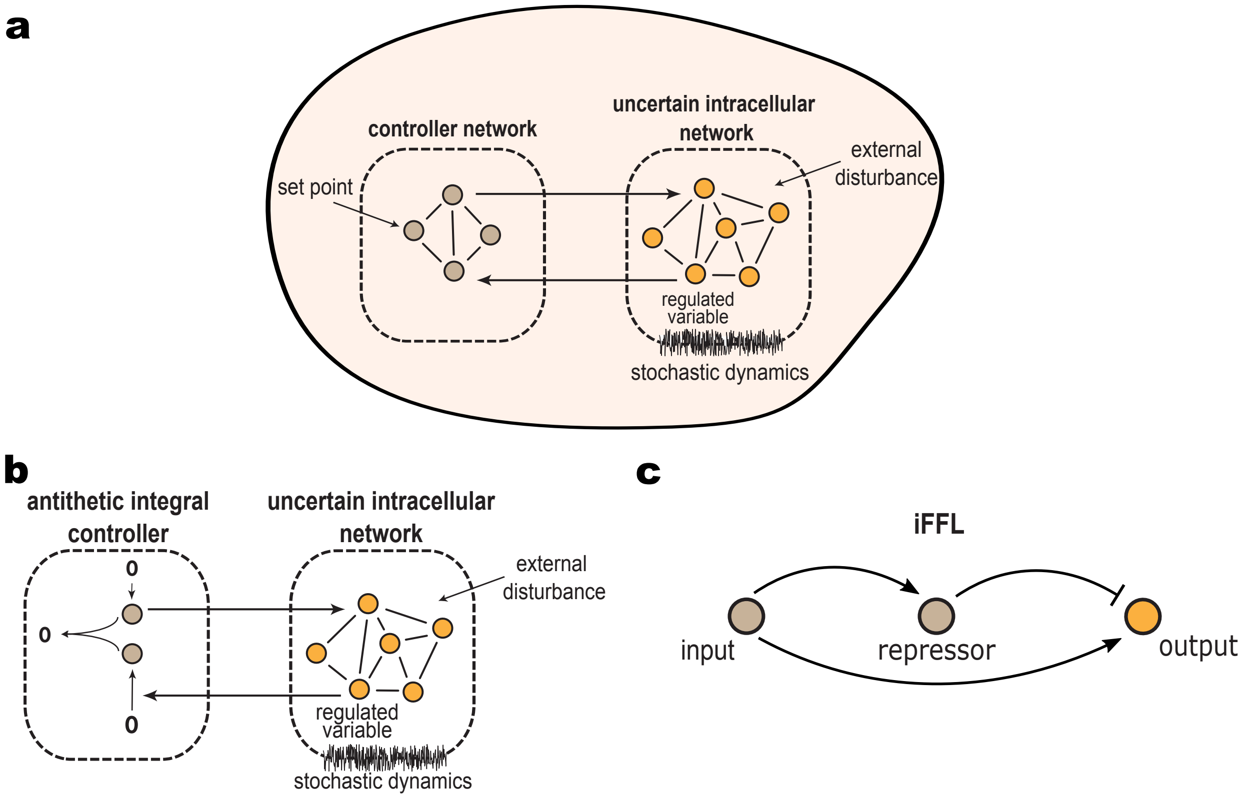


Figure 3-2: Embedded controllers. (**a**) Both the biological process and the controller are present within the same cell as genetic networks. (**b**) The 'antithetic' integral feedback controller [72] is an embedded controller requiring two species that mutually inactivate each other through sequestration or other means. This controller can achieve perfect adaptation. (**c**) Incoherent feedforward loop (iFFL) motif. This is a feedforward controller able to achieve perfect adaptation and has been applied to control burden in mammalian cells. In both the iFFLs, the input is represented by the cellular resources and the output **by the mRNA or protein level of the gene of interest**. The repressor species is represented respectively by either a microRNA [73] or an endoribonuclease [74].

controller has been implemented in *E. coli* [77, 8] by means of an highly affine sigma/anti-sigma protein pair. More recently, the first synthetic implementation of integral feedback in mammalian cells has been presented in a pre-print publication [78], where a sense-antisense RNA pair is used in place of the sigma/anti-sigma pair used in bacteria. The operation of this synthetic biomolecular closed-loop integral controller was further analysed [79, 80] and a set of sufficient conditions under which such a biomolecular feedback controller has the desired asymptotic performance was found to involve time-scale separation. This theoretical analysis was validated in bacterial cells [81] to accommodate physical constraints on the implementation of integral control in cellular systems. Finally, Chevalier *et al.* [9] proposed modifications of the antithetic controller to obtain also proportional and derivative controllers.

The ability of any integral controller to sum over time output deviations from its target value in living cells is challenged by the fact that all biomolecules dilute as the cell grows, resulting in a leaky integration thus causing the controller's performance to degrade. However, it has been mathematically proven that if the reactions implementing the integral controller are all much faster than dilution, then the adaptation error due to integration leakiness becomes negligible [82].

Embedded controllers could be game-changing in synthetic biology, but two main problems have to be solved before they can become the staple food of biomolecular engineers: modularity and burden.

Modularity has been an unsolved issue in synthetic biology preventing the scaling up of synthetic circuits because of undesired retroactivity effects. The concept of retroactivity has been defined in the context of synthetic circuits that can be abstracted as an input-output system: when the output is driving a downstream load, the behaviour of the upstream system should be unchanged. However, this is usually not the case. For example, transiently expressed genes can compete for limited transcriptional and translational resources promoting the coupling of otherwise independent exogenous and endogenous genes [**73**]. Recently, incoherent feedforward loop (iFFL) circuits were identified and engineered to mitigate gene expression burden in mammalian cells using either natural and synthetic miRNA-based [73] or an endoribonuclease-based feedforward loop [74] to adapt the expression level of a gene of interest to significant resource loading (Figure **3-2**c).

Further examples of synthetic circuits able to decouple the synthetic circuit from the rest of the cell and thus offer some sort of modularity have been described: a family of gene expression control systems was developed in Chinese Hamster Ovary (CHO) cells and in human-iPS cells [83]. These include: (i) an incoherent feedforward circuit that exhibits output tunability and robustness to plasmid take-up variation; (ii) a negative feedback circuit that reduces the burden and provides robustness to transactivator dosage variability; and (iii) a hybrid circuit that combines the previous two. A dCas9-based feedback-regulation system that automatically adjusts the expression of a synthetic construct in response to burden was also implemented in *E. coli* [84] ensuring robust growth. A regulator that adjusts dCas9 concentration based on sgRNAs' demand, thus mitigating competition in CRISPRi-based logic gates, was developed in *E. coli* [85].

Despite current progress and the notable successes here reported, some open challenges remain. Embedded control must be customised to the cellular environment, which is a dynamic network among DNA, proteins, and metabolites. Such environment poses limitations

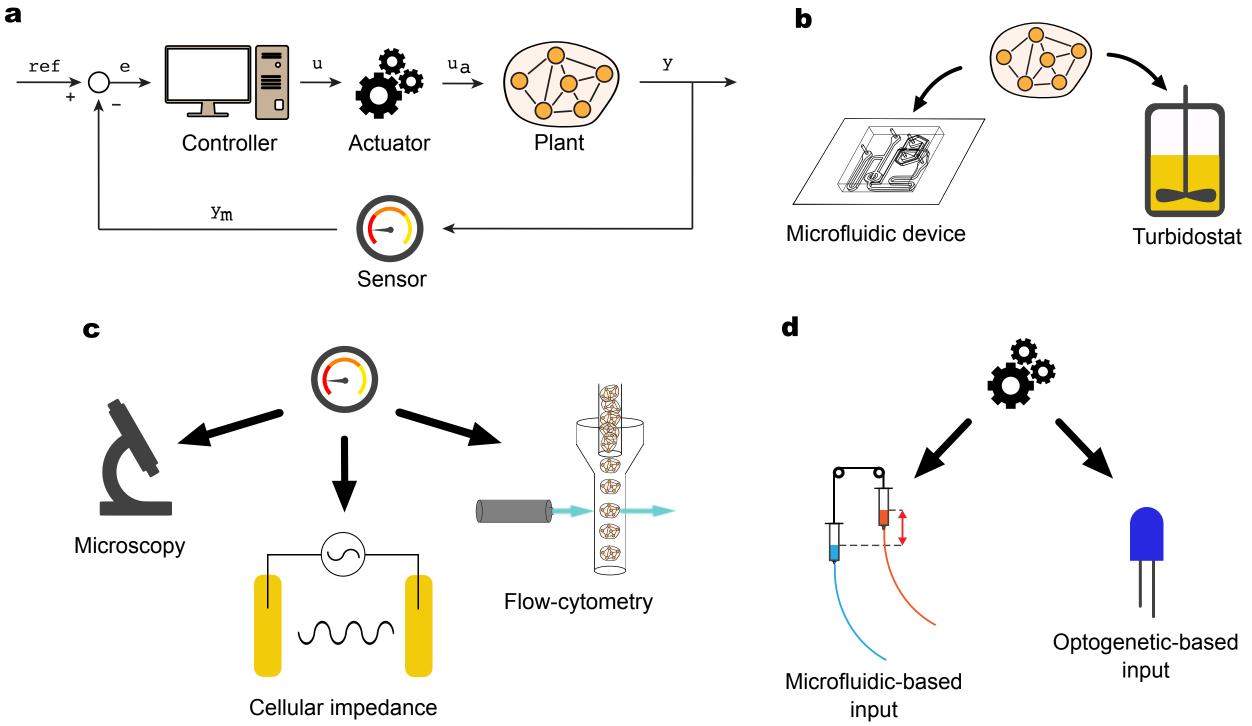


Figure 3-3: Experimental platforms for the application of external control. (**a**) External control platforms consist of four basic modules: **(i) the controller, i.e. a computer algorithm; (ii) the actuation system to provide the control input; (iii) cells, i.e. the plant; and (iv) the sensor to measure the output.** Each module can be implemented in a variety of ways. (**b**) Cells need a suitable environment to grow. Either microfluidic devices or turbidostats have been used for long term culture of cells in control applications. (**c**) A variety of sensors can be used to measure the system output, which is usually proportional to a fluorescent reporter protein. Usually microscopy is used in combination with a microfluidic device, whereas flow-cytometry is preferred with turbidostats. Recently, in [86], microelectrods were used to measure the cellular impedance, considered as proxy of colony growth. (**d**) Actuation includes changing the concentration of a specific inducer (microfluidic-based input) or exposing the cells to light of a specific wavelength (optogenetic-based input). While the first strategy can be used only on cells cultured in a microfluidic device, light can be used both in the case of microfluidics and turbidostats.

on the level of sophistication that controllers can take, limitations that could be addressed by overcoming the current main problems: limited resources, cell-cell heterogeneity, intrinsic stochastic noise, context dependence, unknown or uncertain network topologies, metabolic burden, and the latency of controlled variables.

3.3 External Control

To overcome the current limitations of *embedded* control and to speed up the testing of different control architectures, it is possible to perform *in silico* or *external* control. External control is based on interfacing the biological process to be regulated with a computer in which the control algorithm is implemented as a computer software. *In silico* control has been applied over the years by means of different platforms, as summarized in Figure **3-3** [87, 88, 89, 51, 90, 91].

To achieve external control of the cellular process, i.e. the plant, four basic modules are required: controller, actuator, plant, and sensor, as shown in detail in Figure **3-3**a. Cells grow in a regulated environment such as a microfluidic device or a turbidostat (Figure **3-3**b), whereas the measurement system is usually either a microscope (Figure **3-3**c) to image a fluorescent reporter protein in cells grown in a microfluidic device [87, 89, 51, 90, 91], or a flow-cytometer when cells are grown in turbidostats [88]. Alternatively, Din *et al.* [86], in place of a fluorescent protein, used cellular impedance as system output to measure bacterial colony growth. To this end, cells, grown both in turbidostat and in a customized microfluidic device, are put in contact with microelectrodes able to measure cellular impedance, which is strictly correlated with colony size. The actuation systems necessary to deliver the control input to the cells (Figure **3-3**d) are varied; the most common methods include syringes and pumps to change the concentration of a chemical or a metabolite (microfluidic-based input) [87, 89, 91] or lasers and LEDs to expose cells to light of specific wavelengths (optogenetic-based input) [88, 51, 90]. Chemical inputs have the advantage of being simple to implement and can be biologically relevant (e.g. drugs, hormones, etc.). Optogenetic inputs require extensive engineering of the cells to respond to light and may induce phototoxicity, but they offer precise spatio-temporal modulation allowing input delivery at a single cell scale [90] through a Digital Micromirror Device (DMD) projector. Moreover, in biotechnological applications, an optogenetic input may be the only option as changing the concentration of a chemical input may be very expensive or practically impossible in large biofermenters. Therefore, a growing effort in developing robust optogenetics circuits is ongoing and it has been systematically reviewed in [92, 93].

Unlike embedded controllers, external controllers can be very easily implemented as software in a computer. The most utilized controllers for biomolecular processes are Relay controllers [91, 94, 95, 96], proportional integral derivative (PID) controllers [87, 88, 89, 94, 17] and Model Predictive Control (MPC) [87, 88, 51, 90, 91, 17, 97, 98] to regulate gene expression in bacteria, yeast, and mammalian cells. Alternative model-based controllers have also been presented, such as Zero Average Dynamics (ZAD) control [87] to control gene expression in yeast and reinforcement learning to control microbial co-cultures in bioreactors, optimizing bioproduction despite the resource competition between the two species [99, 100].

External control has been successfully applied to steer gene expression, cell signaling, and endogenous activity in increasingly complex cell systems, such as different types of mammalian cells among which neurons and embryonic stem cells [91, 94, 95, 101], and to key cellular processes, such as the cell-cycle [96] and the maintaining of an undifferentiated state [17]. Indeed, Perrino *et al.* [96] were able to synchronize the cell cycle over a population of budding yeast engineered to respond to an external stimulus in a certain phase. Guarino *et al.* [17], instead, proposed and tested *in silico* different control strategies able to keep a toggle switch in its unstable equilibrium point, demonstrating the feasibility of maintaining cells in an undifferentiated state through cybergenetics tools. Finally, external control was used also as a test-bed for multicellular control (see next section). With cell-in-the-loop, indeed, Perkins *et al.* [102] emulated cell-to-cell signals calculated *in silico* from real-time measurements using light.

External control has also been applied to study unknown biological mechanisms. For example, Harrigan *et al.* [97] applied closed-loop optogenetic compensation (CLOC) to investigate the dynamics of endogenous cellular feedback loops, replacing them with their synthetic light-inducible version and closing the loop externally with an MPC. With this tool, the authors elucidated the time scales of the yeast pheromone response MAPK pathway, which cannot be fully interrogated by compensation with static genetic alleles. Another application of external control is the application of microfluidics-based feedback to precisely control human α-synuclein expression in yeast cells [98], allowing the authors to quantitatively compare the effect of pathogenic mutations on the concentration-dependent formation of α-synuclein aggregates, responsible for Parkinson Disease.

External control overcomes some of the difficulties encountered in embedded control, but some obstacles still need to be bypassed. Here, the output measurement process remains the key issue. Most of the time control and measurement of the same protein exploit fluorescent reporters. Anyway, in most practical applications, an indirect measurement approach is required, resulting in delayed, noisy, and uncertain measurements. Moreover, the problem of scalability of external control to large bioreactors for industrial application is a major hurdle.

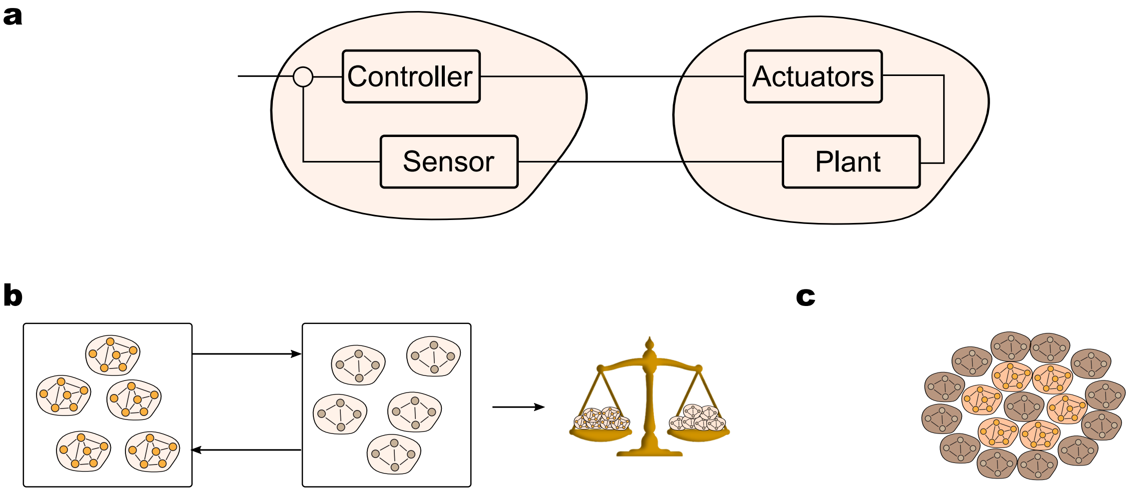


Figure 3-4: Multicellular control strategies can be invaluable to drive the behaviour of microbial communities. They can be used to regulate the phenotipic expression in a target population (**a**), to allow long term coexistence within microbial consortia (**b**) and to induce a desirable collective behaviour for a community. (**a**) The feedback control functionalities are split across two populations. The first, called controller, can sense the the state of a target population and, comparing the measured state to a reference is able to provide a control signal to steer the gene expression in the second population. The second population, called target, implements an actuation system responsive to the controllers output and embeds a plant to be controlled, whose state is signalled to the target population. (**b**) Feedback control implemented across populations is capable of granting a balanced coexistence between different populations inside a microbial consortium. (**c**) Multicellular control can be used to induce a collective behaviour in microbial communities, such as patterning or control of cell cycle.

3.4 Multicellular Control

In embedded control, the controller and the process to be controlled are in the same cell, whereas in external control, the controller resides in a computer. An alternative approach, which is recently gaining ground, is to move the controller to a different cell so that the controller and the process to be controlled reside in two distinct cell populations. This approach, called multicellular control, overcomes some of the limitations of embedded controllers: for example, it reduces the metabolic burden introduced in the host organism and increases the modularity of the developed architecture.

One of the recent steps in this direction was described in [103], where a proof of concept design for a feedback control scheme split among two bacterial populations was developed (Figure 3-4a). Here, a controller population is engineered to steer the behaviour of a target population. The controllers embed synthetic circuits to sense and control in real-time the state of a process in the targets. The controller cells can also receive an external signal (e.g. an inducer molecule) to set the desired reference level of the process output to be regulated in the target cells. The communication between the two subgroups is realised by means of the so-called quorum sensing mechanism, in which the cells produce and release in the growth medium a molecule that serves as a proxy of their current state and that is sensed by the other cells [104]. A mathematical model of the consortium was developed, showing *in silico* the potential of the architecture for a reliable and robust regulation of gene expression. More precisely, it was proved that the target population can be robustly controlled to a desired state, without the need of engineering in it a complex and metabolically burdensome gene regulatory network. Following this work, in [18] and [105] a biological implementation of the controller population has been developed and characterised *in vivo*.

A critical requirement for a correct operation of a multicellular control architecture is keeping the ratio between the two populations within an acceptable range (Figure **3-4**b). Strategies to tackle this problem have been recently proposed. For example, in [106] a lysis mechanism is leveraged to ensure coexistence between three *E. coli* strains. A similar goal was achieved in [44, 107] where a toxin-antitoxin based feedback control system is implemented across two populations aiming at controlling the population size and composition of a microbial community. A follow up work also investigates the role of spatial configuration in the operation of this feedback control scheme [50]. A dual approach to achieve the same goal was presented in the Ecolibrium project (2016 iGEM entry from Imperial College London; URL: `http://2016.igem.org/Team:Imperial_College`) where the regulation is achieved via a growth regulation protein. Control architectures distributed among members of the same population have been also engineered to orchestrate the collective behaviour of a microbial consortium (Figure **3-4**c). As an example, in [108] a quorum sensing mechanism is used to share knowledge across the population, allowing for a synchronized cell division across members of the same population. Furthermore, synthetic consortia can be engineered to implement a majority or minority sensing mechanism [48] or to exhibit oscillations across members of the same [109] or different [110] populations. Finally, communication and group knowledge can also be beneficial to control spatial patterns in microbial communities allowing cells to self-organize in specific spatial patterns [102, 111, 32].

Multicellular control is limited by the complexity of its design and the lack of an *in vivo* realisation of a consortium implementing this paradigm. Moreover, an open challenge is how to ensure stable and long term co-existence of the two populations (controllers and targets).

3.5 Discussion

We reviewed three classes of control strategies currently being researched in cybergenetics, the emerging field combining synthetic biology with Control Theory. Specifically, we reviewed the key features, possible applications, advantages, and disadvantages of external, embedded, and multicellular control strategies.

We showed that implementing biomolecular controllers it is possible to embed living cells with the ability to autonomously regulate their phenotype in response to a reference signal. However, the realisation of an entire feedback control loop inside a single cells can result in an excessive metabolic load on the host organisms. A promising solution to overcome this limitation is called multicellular control, where the desired functionalities are distributed across populations comprising a microbial consortium.

The realisation of a multicellular control strategies could lead to the realisation of controlled microbial communities, where each population robustly expresses a desired phenotype. This would enable the consortia to carry out complex tasks without overloading each organism with burdensome gene networks.

Chapter 4

In-vivo validation of a multicellular control architecture

The realization and validation of a multicellular control architecture *in vivo* requires splitting the sensing, control and actuation functions across multiple cell populations by engineering a closed-loop consortium able to exhibit a desired phenotype (e.g. fluorescence level) in a robust manner. Inspired by the design proposed in [103], we designed a first population, denoted as controllers, whose role is to sense the expression of the controlled gene in another population, the targets, and produce, according to some control logic, a stimulus steering the gene expression towards a desired level. The targets sense the controllers' output and in response modify the expression of the controlled gene, feeding back its level to the controllers (Figure **4-1** a). Note that such architecture requires engineering two synthetic gene pathways and also embedding an interface for the communication between the populations. We implemented the communication interface using quorum sensing, a mechanism used by bacteria to sense and respond to changes in their local population density.

4.1 Biological Implementation

The control logic, embedded in the controllers, is derived from the error computation module (comparator) based on molecular titration proposed in [105] and characterized dynamically in [18]. The system is a multi-input, single output device that uses two independent signals to regulate the expression level of its output protein. It is constructed around an orthogonal pair of σ/anti-σ (σ_a) factors [112], which are molecules able to specifically regulate the transcriptional activity of the *p20_992* promoter through an antithetic titration motif. The *p20_992* promoter is placed upstream of the gene coding for LasI, responsible for the production of the quorum sensing molecule 3-O-C12-HSL. The production of σ is controlled

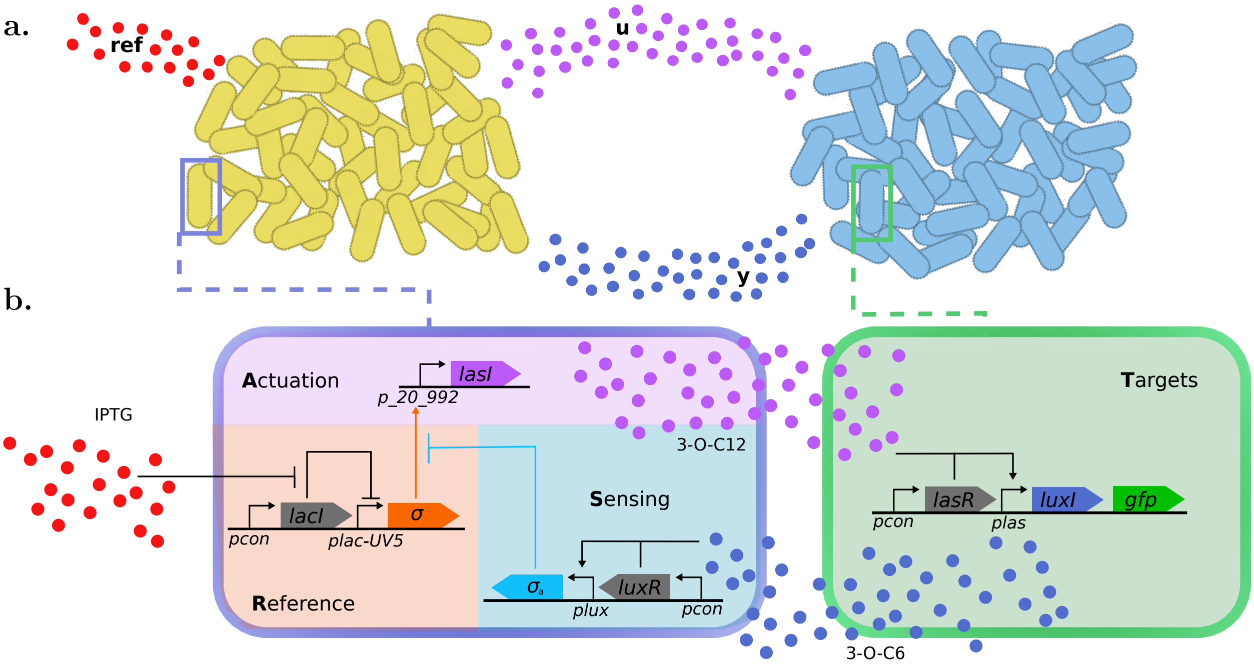


Figure 4-1: **a** Microbial consortium implementing a multicellular control architecture. The two populations, i.e. controllers and targets, share information using orthogonal quorum sensing channels (u and y), playing the role of the control input and the output of the process, respectively. In addition, the controllers sense the external stimulus ref, used to set up the reference of the feedback loop. **b.** Biological implementation of the multicellular control architecture. Standard notation is used to denote promoters, genes, promotion/inhibition relationships and quorum sensing molecules. The shaded area identify different functional modules within each population.

by the IPTG inducible promoter *plac-UV5*, while the production of anti-σ is controlled by the 3-O-C6-HSL inducible promoter *plux*. The *plux* promoter, together with the constitutively expressed LuxR plays the role of a sensor, able to act as a proxy for the abundance of the controlled protein in the targets. The 3-O-C-6-HSL, which is produced by the targets proportionally to the amount of the controlled gene and used as its proxy, activates the *plux* promoter with the aid of the constitutively expressed LuxR (Figure 4-1b). All proteins are fused to a degradation tag (ssrA tag) [113], ensuring fast dynamics of expression. The proposed control logic embeds an antithetic motif that has been shown to implement an integral feedback controller with the ability to guarantee *robust perfect adaptation* [72, 8].

The targets can express a green fluorescent protein (GFP), whose expression level we aim to regulate. The GFP coding gene is transcribed together with the *luxI* gene, responsible for the production of the sensor molecule 3-O-C-6-HSL. The actuation module, designed to regulate *luxI* and *gfp* expression levels, is realized by constitutively expressing LasR that, binding to 3-O-C12-HSL (produced as output by the controllers), activates the *plas* promoter, placed upstream *gfp* and *luxI*. All the details on the plasmids and strains used in this chapter can be found in section A.1.1.

The goal of the design is to make the two engineered populations cooperate so as to steer and maintain the GFP expression in the targets to a desired setpoint level, which can be changed by modifying the concentration of IPTG. Specifically, when GFP is overexpressed, the excessive production of 3-O-C6-HSL should upregulate the production of anti-σ. In turn, a reduction of the free available σ, should downregulate the production of LasI and 3-O-C12-HSL production. As a consequence, the activity of *pLas* should be reduced, knocking down the GFP expression levels. A similar mechanism should be triggered when GFP falls below the desired level.

4.2 Controllers and targets can successfully communicate

A fundamental requirement for the correct operation of the engineered microbial consortium is for the interfaces between controllers and targets to be compatible. More precisely, we need the output range of each population to be compatible with the input range of the other. To this aim we first characterized the steady state response of both populations using flow cytometry, which allows collection of morphological and fluorescence data at a single cell resolution [114]. The controllers can be induced using IPTG and 3-O-C6-HSL, which are used in the architecture as a reference signal and a proxy for the targets state, respectively. We characterised their response to IPTG (3-O-C6-HSL) concentrations ranging in the interval $[0\,\mu M, 100\,\mu M]$ ($[0\,nM, 100\,nM]$). Note that, to measure the output level of the controllers, we substituted LasI with a green fluorescent protein (see Figure **A-2**). Increasing the concentration of the 3-O-C6-HSL had an inhibitory effect on the average fluorescence of the controllers' population, presumably due to the increase of the production rate of anti-σ (Figure **4-2**a). Specifically, we observed a decrease between 10 to 100 fold in fluorescence, depending on the IPTG concentration. Conversely, IPTG promoted σ production, causing up to a 300 fold increase in average fluorescence when $100\,\mu M$ IPTG is present, with respect to the case where no IPTG is added to the growth environment. Similarly, we induced the targets using different 3-O-C12-HSL concentrations, analysing their average fluorescence output after 3 h to derive their working range (for details on the protocol used to induce the cells, see Section A.1.5). As expected, the inducer molecule promotes *plas* activity, increasing the average fluorescence levels of the targets (Figure **4-2**

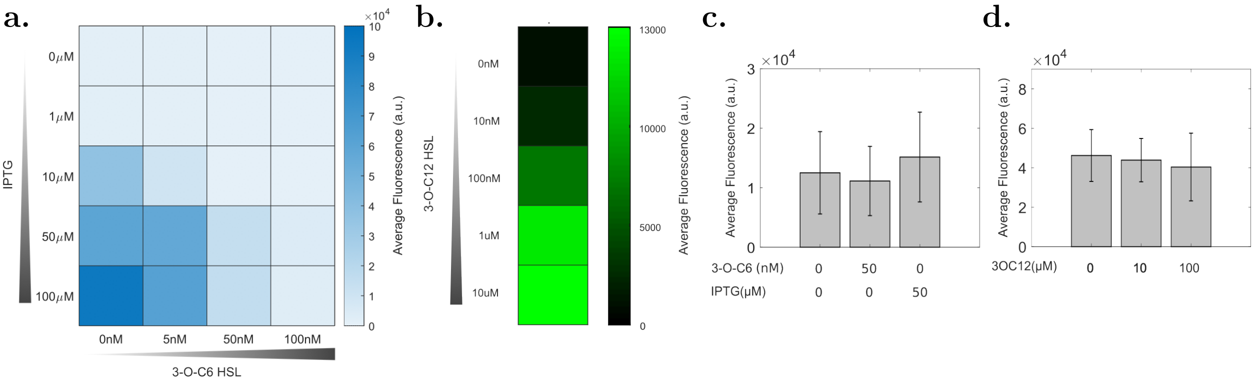


Figure 4-2: **a**: Average fluorescence of the controller population induced using different concentrations of 3-O-C6-HSL and IPTG. **b**: Average fluorescence levels of the target population induced using increasing levels of 3-O-C12-HSL. **c**: Average fluorescence levels of the targets grown in the controllers supernatant. **d**: Average fluorescence levels of the controllers grown in the targets supernatant when no IPTG was provided in the culture medium. The error bars in panels c and d are representative of the standard deviation among $n = 3$ biological replicas.

b). More precisely, it produces up to a 14 fold increase in average fluorescence when 10 μM 3-O-C12 HSL is added to the growth medium.

After we characterized the ranges where controllers and targets are sensitive to their inputs, we tested their ability to induce one another by growing each population in the supernatant obtained from a culture of the other strain. First, we observed the response of the targets when grown in the medium where we previously cultured the controllers (see Appendix A.1.6 for details on the protocol used). The controller population was able to activate GFP production in the targets (Figure 4-2c). Moreover, inducing the controllers with IPTG further increased the average fluorescence of the targets, while the addition of 3-O-C6-HSL in the controllers culture caused a reduction of the fluorescent reporter expression level, confirming that the concentration of 3-O-C12-HSL produced by the controllers is compatible with the target's dynamic range. Analogously, we induced the controllers with the supernatant of the targets. Figure 4-2d highlights that the targets are producing 3-O-C6-HSL at levels compatible with the controllers dynamical range. In detail, increasing the concentration of 3-O-C12-HSL in the target cultures reduced the average GFP fluorescence levels of the controllers. Note that in both communication tests, the replicates showed a high variability in the amount of quorum sensing molecules, potentially as a result of their different growth rates (data not shown) during the first incubation step. Such variability caused a failure in proving statistical difference between the different conditions. However, the average behavior and the qualitative response of each replica consistently suggested that the communication architecture enabled exchange of information between the populations.

4.3 Robust gene regulation is possible via a multicellular feedback control architecture

As all the functional requirements for the architecture to work were met, we tested its ability to regulate the GFP expression levels in the targets. The analysis has been conducted using flow cytometry, where we can select and analyse the targets in the consortium by gating cells using the appropriate fluorochrome (see A.1.4 for details on the gating strategy used). The separation of the populations was possible thanks to the lack of overlap between the controllers (not fluorescent) and targets (fluorescent) profiles (see Section A.1.4 for details on the gating strategy).

We tested the regulation capability of the controllers by comparing the closed loop system response with its open loop counterpart, where the sensing module (see Figure **4-1**) was removed to break the feedback loop. This was obtained by synthesizing alternative controllers, denoted as *open loop Controllers*, missing the anti-σ plasmid (denoted as sensing in Figure **4-1**). The open loop controllers retain the components needed to modify the GFP expression levels in the targets, however they miss the module required to adjust their output in response to changes in the targets' state. By mixing controllers and targets we observed that the open loop architecture does not stabilize GFP during a 6 h time-course, as opposed to the closed loop control architecture which is able regulate GFP expression levels in about 3 h. Details on the protocol used can be found in Section A.1.7.

The impossibility of regulating the fluorescence of the targets in open loop is due to the lack of robustness of such architecture. Specifically, since the two strains bear different metabolic loads, their growth rates are different and their relative numbers in the community change continuously. Thus, since the open loop controllers are not able to counterbalance the imbalances in sensing molecules created by the changes in the population densities, the targets' GFP levels will not settle to a steady state value. We further investigated this phenomenon by analysing the 6 h response of the targets in open loop and closed loop when different percentages of targets are present in the consortium. Those conditions were obtained by inoculating cultures with different initial controllers/open loop controllers over targets ratios and measuring the targets percentage and fluorescence levels in the consortium after 6 h (more details in section A.1.8). We observed the controllers in closed loop to consistently stabilize the targets around a constant desired GFP steady state expression level, independently of the percentage of targets in the consortium at steady state. Instead, in open loop the targets' average fluorescence level decreased as the percentage of targets increased (Figure **4-3** b). This is likely due to the level of 3-O-C12-HSL being produced,

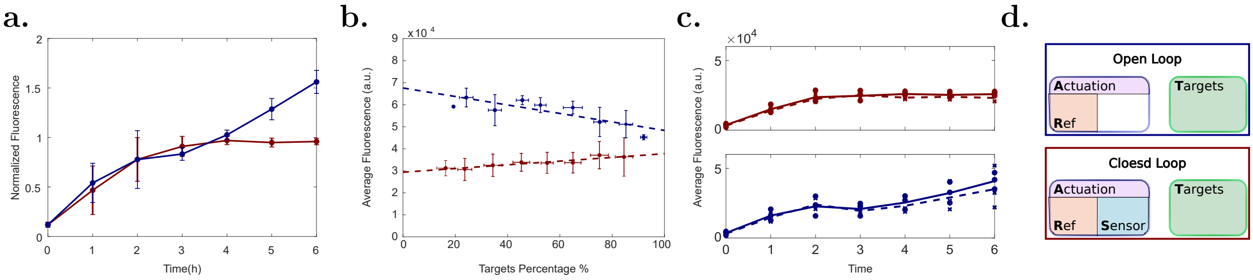


Figure 4-3: **a.** Normalized average fluorescence of the target population in open (blue) and closed (red) loop when no IPTG is provided in the media. The values are normalized on the steady state value reached in closed loop. **b.** Average fluorescence of the targets in open (blue) and closed (red) loop when no IPTG is present in the media, plotted against the targets percentage in the consortium. The data, collected after 6 h, are grouped according to the targets percentage in 10 bins spanning from 0% to 100%. **c.** Average fluorescence of targets in closed (top panel) and Open (bottom panel) loop in presence of additional metabolic burden in the targets. Dashed lines and crosses represent average trajectories and data points of experiments conducted using targets bearing additional burden, while solid lines and dots are average trajectories and points obtained from targets without any additional metabolic load. **d.** closed loop and open loop abstract representation. The open loop system was obtained mixing controllers missing the anti-σ plasmid with the targets. Instead, the closed loop system was assembled mixing full controllers with the targets. In panels a. and b. the error bars represent the standard deviation of the distribution over $n = 3$ biological replicas.

which is proportional also to the density of the controllers in the consortium, hence to their percentage. We quantified the difference between open and closed loop steady states by fitting the points in Figure 4-3b using a first order polynomial. The angular coefficient of the line which best fitted the open loop data (blue dotted line in figure 4-3b) is -192, which is 2.3 times bigger that the slope of the line fitting the closed loop steady states (blue dotted line in figure 4-3b). This highlights that the percentage of targets in the consortium is twice more influential on their average fluorescence in open loop then it is in closed loop. Hence, this analysis highlights the necessity of a closed loop feedback control scheme to achieve a robust regulation of gene expression.

We also tested robustness of the architecture to metabolic burden, which was shown to be detrimental for the performance of bio-molecular controllers in [2]. The introduction of an highly expressed gene can indirectly interact with the designed gene network by taking up some of the shared transcriptional and translational resources of the cell. Specifically, we increased the metabolic load on the targets by embedding an additional plasmid in the target population, which constitutively expressed GP55.1, a protein involved in the inhibition of DNA repair [115] (Figure A-3a) . This modification caused a 30% drop in the average

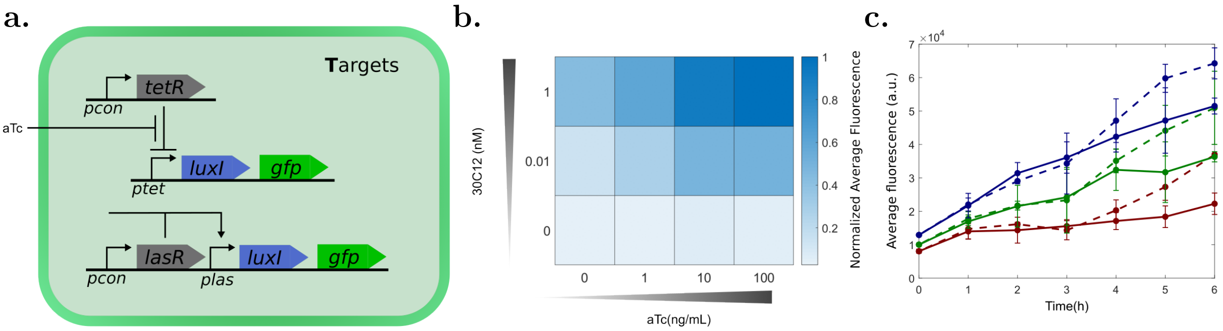


Figure 4-4: **a.** Implementation of a tunable output disturbance in the target population. **b.** Normalized steady state average fluorescence of the circuit described in panel a in presence of different levels of 3-O-C12 and aTc. The steady state fluorescence was collected after 3 h. The values are normalized with respect to the condition of maximal fluorescence corresponding to 10 μM 3-O-C12 + 100 ng/mL aTc **c.** Average fluorescence level of the targets described in panel a in closed loop (solid lines) and open loop(dashed lines) in a 6 h timecourse experiment. The red, green and blue curve correspond to aTc concentrations of 0 ng/mL, 1 ng/mL and 10 ng/mL, respectively.

fluorescence in presence of 10 μM 3-O-C12 with respect to targets not bearing the additional load (Figure A-3b).

Adding the controller population we were able to reduce consistently the effect of metabolic burden (Figure 4-3c). The average reduction of GFP expression levels at steady state with respect to cells where no extra metabolic burden was added went from 30% to almost 10%. In addition, the standard deviation between the perturbed and unperturbed scenarios were comparable. Instead, in open loop we observed a 2.5 fold increase in the variability of the 6 h response of the targets when the burden was introduced. However, the Average fluorescence level had a similar relative shift with respect to the closed loop counterpart. This was due to inevitable interference between the effect of metabolic burden and the fluorescence change induced by the differences in the relative numbers between controllers and targets.

As an additional test of the robustness properties of the architecture, we implemented a tunable disturbance in the output fed back by the target population (Figure 4-4a). Specifically, we added another copy of the *lasI* and GFP coding genes downstream of the inducible promoter *ptet*, whose activity is repressed by the presence of TetR, which is constitutively expressed. The repression exerted by TetR can be modulated by sequestration adding anhydrotetracycline (aTc) in the culture environment [89]. Note that this system works as a disturbance in the signal fed back to the controllers, as the fluorescence signal and the corresponding 3-O-C6 HSL production cannot be regulated by the action of the controller. We tested the engineered disturbance system, confirming that by varying aTc levels we could modulate the GFP over-expression at all levels of 3-O-C12 used, causing up to a two-fold increase in GFP fluorescence when TetR inhibition was fully released (Figure 4-4b). However, when we mixed targets and controllers, the feedback loop was not able to reject the perturbation as we expected. By comparing the response of the closed loop system when aTc 10 ng/mL was added to the environment to the case where the perturbation was not present (0 ng/mL aTc), we noticed that the increase in fluorescent signal induced in open and closed loop was comparable (Figure **4-4**c). Hence, we concluded that the amplitude of the perturbation was excessively high for our controllers to reject it.

4.4 Testing the tunability of the architecture

The final step towards the *in vivo* validation of our architecture was to assess the tunability of the target fluorescence in response to changes in the reference value. Hence, we compared the closed and open loop dynamic response providing 50 μM IPTG in the culture media. This concentration was chosen as the controllers produced significantly more σ with respect to the uninduced scenario (100 fold increase), and we had evidence that the communication between controllers and targets still worked (Figure **4-2**c). However, there was no statistical difference between the open and closed loop target response when the consortium was grown in presence of 50 μM IPTG (Figure **4-5**a, top panel). To identify the causes of the regulation failure, we analysed the morphology of the cells in the consortium. Specifically, we gated the cells analysed with flow cytometry based on their size and complexity (see section A.1.4 for more details on the gating). We found out that the controllers morphology was altered in presence of IPTG after 6 h of growth and that most of the population had abnormal size and internal complexity greatly, which can be a sign of cells bearing an excessive metabolic load [116] (Figure **4-5**a, bottom panel). In addition, by analysing the response of controller cells exhibiting such unhealthy morphology, we confirmed that they were exhibiting a different phenotype (Figure **A-4**). We hypothesized that this toxic effect on the controllers was induced by the excessive production of σ and anti-σ factors when cells reached a stationary growth regime, as shown in [112].

We addressed this problem by growing cells so that they never transitioned to a stationary growth phase, as in [112] it was shown that σ and anti-σ factors were not toxic during exponential growth. Specifically, we used Chi.Bio, an open-source platform for continuous cultures of microbial populations [117]. It allows a precise control of the culture density, allowing cells to grow indefinitely at an exponential rate. We repeated the Timecourse experiment using this platform and observed a much healthier morphology in the controllers (Figure **4-5**b, bottom panel). Details of the protocol used can be found in section A.1.9. In this assay we normalized the data as the open and closed loop experiments were done on different days using different colonies due to technical constraints. Indeed, when comparing

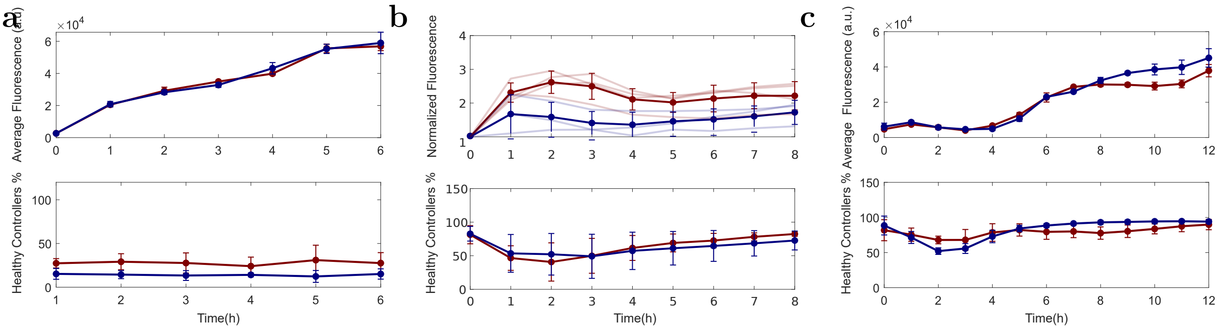


Figure 4-5: **a.** Timecourse of the open (blue) and closed (red) loop architecture in a test tube, using LB as growth medium. *Top panel*: Average fluorescence level of the targets. *Bottom panel*: Percentage of healthy controllers in the consortium (i.e. controllers without abnormal morphology). **b.** Timecourse of the open (blue) and closed (red) loop architecture in continuous culture (Chi.Bio), using LB as growth medium. *Top panel*: Normalized Average fluorescence level of the targets. The normalization was done with respect to the targets fluorescence at time 0 h. *Bottom panel*: Percentage of healthy controllers in the consortium (i.e. controllers without abnormal morphology). **c.** Timecourse of the open (blue) and closed (red) loop architecture in a test tube, using M9 as growth medium. *Top panel*: Average fluorescence level of the targets. *Bottom panel*: Percentage of healthy controllers in the consortium (i.e. controllers without abnormal morphology). In all panels vertical bars indicate the standard deviation between $n = 3$ biological replicas.

the targets normalized average fluorescence in open loop and closed loop, we noticed that the open loop settled to a lower value with respect to the closed loop. This was an artifact of the normalization process. In detail, the normalization was carried out on the initial fluorescence levels of the target population. However, as can be seen in Figure **A-5**a, the fluorescence of the targets at time 0 h, corresponding to the expression level of an overnight culture, was comparable with the fluorescence of targets induced with 10 μM 3-O-C12 HSL for 3 h. This phenomenon was due to the accumulation of GFP due to the absence of an active degradation mechanism for the fluorescent reporter. We confirmed this hypothesis by culturing the targets alone in the chemostat. We noticed that their fluorescence decreased over time thanks to cell division which slowly allowed the population to get rid of the accumulated protein (Figure **A-5**b). A possible solution to this accumulation phenomenon could have been to grow targets and controllers overnight in a chemostat, so that the continous division of cells avoids the excessive accumulation of stable proteins. However, this was not possible due to the lack of additional chemostats.

As an alternative route to slow down cell growth and avoid them entering stationary phase, we grew cells in a different medium, where the availability of nutrients was more limited with respect to LB broth. More precisely, we cultured the cells in minimal media (M9) (see section A.1.2 for details on the preparation of the media). In addition, to avoid overgrowth and retain the small molecules produced by both controllers and targets, we diluted the consortium in its own supernatant after 6 h (the detailed protocol can be found in section A.1.10). This adjustment revealed a difference in the response of the closed and open system were statistically different after 8 h (p-value $7.9\,10^{-7}$). However, the relative difference we observed between the open and closed trajectories was very small (30%).

The results from these assays suggested that with the current implementation of the multicellular architecture we cannot tune the GFP expression levels in the targets in response to IPTG. Possible causes for this phenomenon could be ascribed to the excessive overproduction of σ in the controllers due to the basal expression of the *plac-UV5* promoter and to the accumulation of GFP in the targets due to the absence of any active degradation mechanism.

4.5 Improvements to the Controller and target populations

Failing to prove that the architecture was capable of regulating GFP expression to different levels, we modified the genetic circuits to improve the performance of the consortium. First, we added a degradation tag to the fluorescent reporter expressed in the targets to avoid accumulation of GFP overnight. Specifically, we fused an ssrA tagged *lasI* with GFP (Figure **4-6**a). We chose this option so that the enzyme and the reporter were transcribed and translated together, hence their expression levels were similar. This new version of the targets proved to be inducible by 3-O-C-12 HSL, showing up to a 36 fold increase in average fluorescence when induced with $10\,\mu M$ 3-O-C12 HSL for 6 h (Figure **A-6**a). In addition, there was no significant difference between uninduced cells after 3 h and 6 h and cells not expressing GFP, confirming the effectiveness of the degradation tag in avoiding GFP accumulation (Figure **A-6**b). The insertion of the tag also drastically decreased the leakiness of GFP, creating an overlap between the controllers and targets histograms using the FITC filter (see section A.1.4 for more details on the filters used). For this reason, we engineered a constitutively expressed RFP in the controllers (see section A.1.1 for the details of the plasmid constructed), which enabled the distinction between the two populations within the consortium using the PE-CF594 filter (see section A.1.4 for more details on the filters used).

We tested the performance of the new targets with the controllers in open and closed loop. While the addition of the degradation tag completely avoided the accumulation of GFP in the targets, we were still unable to see any difference between the two configurations in presence of IPTG (Figure **4-6**c). In addition, the difference between the fluorescence levels reached at steady state in closed loop with and without IPTG were not statistically different

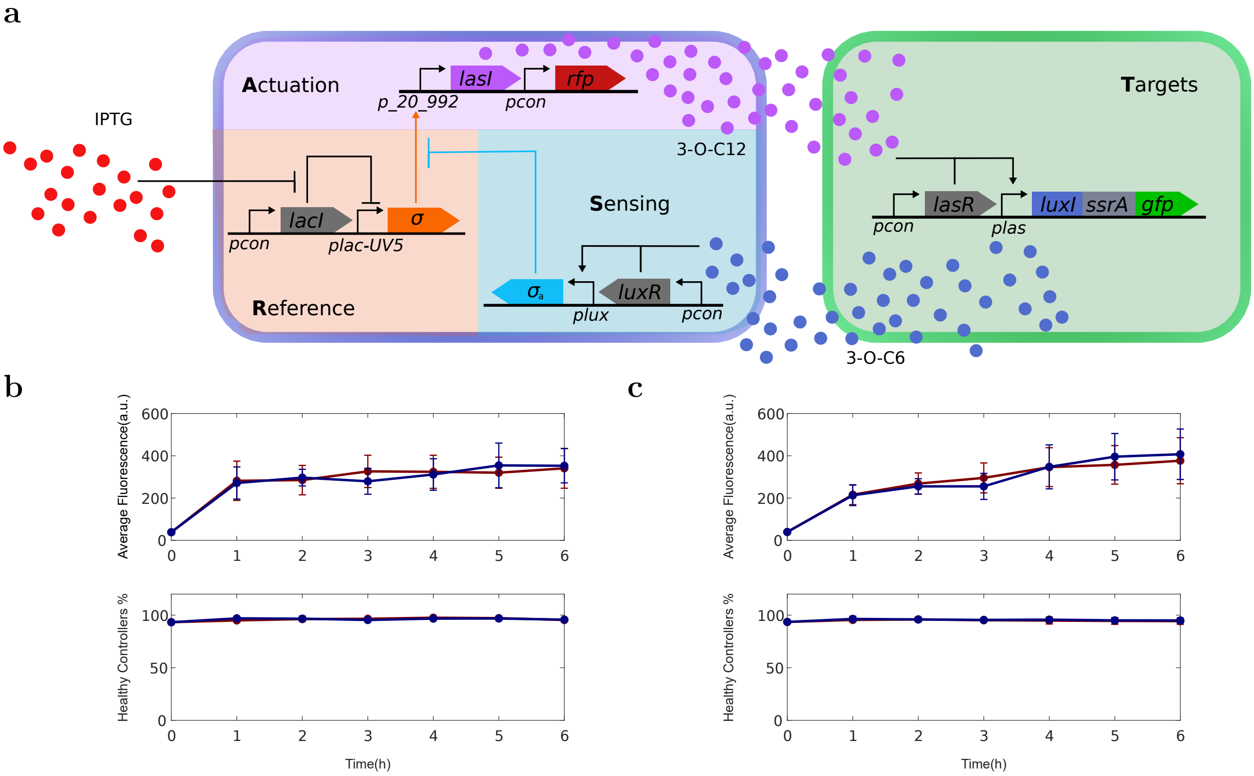


Figure 4-6: **a.** Improved Multicellular control architecture. The controllers were modified by adding a constitutively expressed red fluorescent protein, enabling a reliable distinction between the controllers and the targets. The targets were modified by fusing the ssrA tagged *luxI* gene with the green fluorescent reporter. **b.** Timecourse of the open (blue) and closed (red) loop architecture in absence of IPTG. *Top panel*: Average fluorescence level of the targets. *Bottom panel*: Percentage of healthy controllers in the consortium. **c.** Timecourse of the open (blue) and closed (red) loop architecture when $20\,\mu M$ IPTG were present in the culture media. *Top panel*: Average fluorescence level of the targets. *Bottom panel*: Percentage of healthy controllers in the consortium. In all panels vertical bars indicate the standard deviation between $n = 3$ biological replicas.

(p-value 0.86), suggesting that the targets were expressing saturating levels of the fluorescent reporter, making it hard to capture the differences between the open loop and closed loop response. Therefore, we re-engineered the controllers to reduce the leakiness of the *plac-UV5* promoter by adding an extra lac operator, which increased the repression efficiency of LacI [118]. Mixing this version of the controllers with the original targets (where GFP and *luxI* were not fused in a single protein), we found out that the targets both in closed and open loop were able to settle to a steady state after 6 h, with the closed loop system settling down faster than the open loop counterpart when $7\,\mu M$ IPTG were present in the growth media (Figure 4-7f). In addition, targets fluorescence without any induction was comparable to the steady state response of the targets mixed with controllers when no IPTG was present in the culture media (Figure A-7), proving that the addition of the extra operator reduced almost to zero the leakiness of the *plac-UV5* promoter in the controllers.

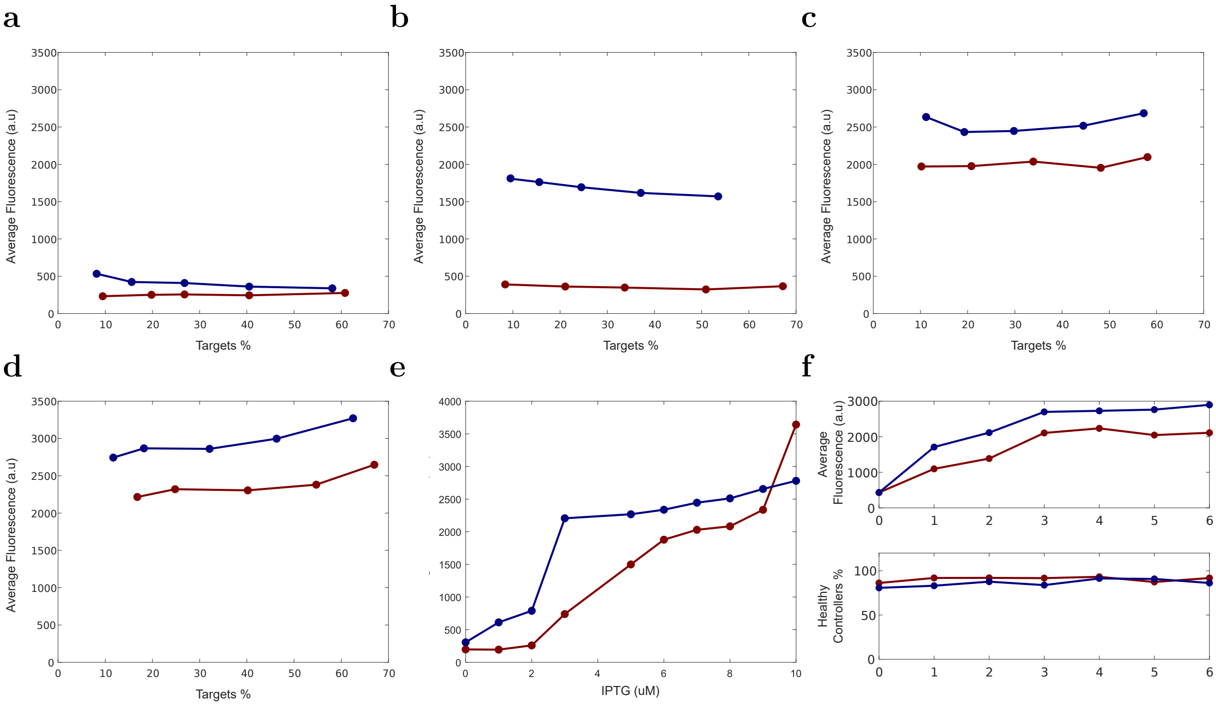


Figure 4-7: **a-d.** Steady state target fluorescence in open (blue) and closed (red) loop when in the culture media was present $0\,\mu M$ (panel a), $3\,\mu M$(panel b), $5\,\mu M$ (panel c) and $7\,\mu M$ (panel d) IPTG, respectively. All steady state data were collected after 6 h growth in LB broth at $37°C$. **e.** Steady state comparison of open loop (blue) and closed loop (red) steady states varying IPTG concentration in the culture environment. Steady state fluorescence was collected after 6 h incubation in LB. **f.** Timecourse of the open (blue) and closed (red) loop architecture when $7\,\mu M$ IPTG were present in the culture media. *Top panel*: Average fluorescence level of the targets. *Bottom panel*: Percentage of healthy controllers in the consortium. All panels represent results of a single experiment.

We tested the tunability of the response by mixing targets and controllers in LB with increasing levels of IPTG. Different GFP steady states were reached between open and closed loop architectures at all IPTG concentrations, with substantial differences when IPTG concentration was lower than $8\,\mu M$ (Figure 4-7e). The closed loop system showed a linear increase in GFP expression levels for $IPTG \in [2\,\mu M, 8\,\mu M]$, while targets in open loop exhibited a sharp increase in fluorescence levels happening between $2\,\mu M$ and $3\,\mu M$ IPTG, followed by a plateau in steady state fluorescence reached. The performance of the closed loop control system degraded fast above $8\,\mu M$ IPTG, exhibiting an unexpected burst in GFP steady state levels when $10\,\mu M$ IPTG were added to the culture media. This is likely due to a saturation in the ability of IPTG to promote further the activity of *plac-UV5*. Finally, we analysed the robustness of the architecture showing that the closed loop control system exhibited a reduced sensitivity to variations in the targets percentage in the consortium. However, as we increased IPTG concentration, the differences in 3-O-C12 levels caused by imbalances in the relative numbers between the populations became less influential on the steady state fluorescence due to saturation. This was particularly evident in open loop for IPTG concentrations higher than $3\,\mu M$ (Figure **4-7**a-d).

These assays suggest that the new architecture can reliably and robustly regulate GFP expression levels in the targets and that such regulation is tunable modulating the IPTG concentration in the growth environment. However, further biological replicas are needed to confirm the reproducibility of our findings.

4.6 Discussion

We presented a possible biological implementation of a multicellular control architecture where the control functionalities are distributed across members of a microbial consortium. After validating the individual constituents of the community, we tested by means of flow cytometry the performance and the robustness of the architecture, confirming the ability of the distributed feedback controller to robustly regulate gene expression in the targets to a desired level. However, due to the toxic effect of the σ factor when overexpressed in a stationary growth regime, we were unable to prove that such regulation was tunable modulating IPTG concentration. Hence, we redesigned the circuit and collected preliminary evidence that the new design could realize a reliable, robust and tunable regulation of gene expression levels in the target population.

The validation of a multicellular feedback biomolecular controller would enable the implementation of modules robustly expressing a desired phenotype that can be integrated in larger systems to realize complex functionalities. In addition, the realisation of multicellular controllers could pave the way for the control of cellular populations that cannot be engineered, given the ability to realize appropriate interfaces between the wild type and engineered organisms.

Chapter 5

Applications and extension of Multicellular control

The realisation of a multicellular biomolecular controller paves the way for the realisation of microbial communities carrying out complex functionalities. For example, in this chapter we present a synthetic microbial consortium leveraging a multicellular control paradigm to set the state of a bistable memory mechanism implemented using a genetic toggle switch. In addition, aiming at improving further the dynamical performances achievable using multicellular controllers, we design and validate *in silico* a consortium implementing a modular PI controller, where each bacterial population implements a different contribution to the control signal.

5.1 Multicellular feedback control of a genetic Toggle Switch in microbial consortia

In this section we present a possible application of the multicellular feedback control strategy validated in chapter 4 for the control of a bistable memory element. The architecture involves a consortium consisting of three cellular populations, in which the activity of one of them is governed by the other two. Specifically, the state of a genetic toggle-switch endowed in one of the populations, the *targets*, can be controlled by providing or removing a reference input to the other two, the *togglers*, which communicate with the targets via orthogonal quorum sensing molecules. In this way, additional functionalities can be toggled in the targets, as required in a number of applications, e.g. production and secretion of some desired molecule or drug in the environment [119, 120].

The relationship between the three cell populations in the consortium and their molecular signals can be schematically represented (see Fig. **5-1**) as a sequential logic circuit [121]. The

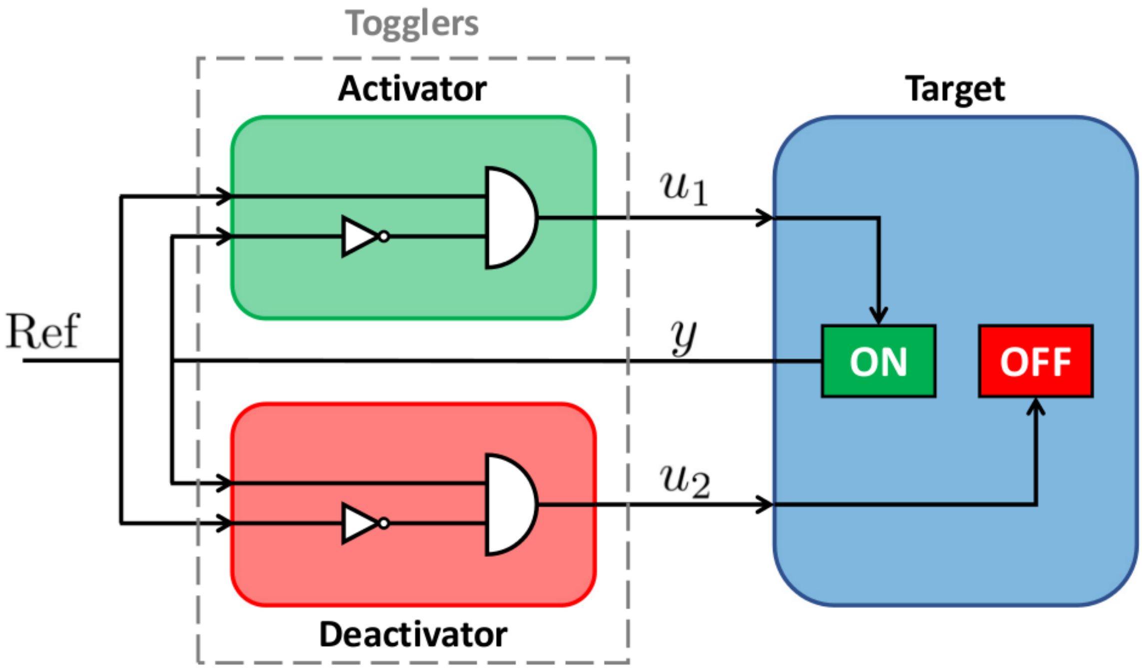


Figure 5-1: Representation as a sequential logic circuit of the relationship between the cell populations and their molecular signals. The target receives ON commands ($u_1 = 1$) only when Ref $= 1$ AND $y = 0$, and OFF commands ($u_2 = 1$) only when Ref $= 0$ AND $y = 1$. In this way, the input signals u_1 and u_2 are equal to 1 only when there is disagreement between Ref and y.

two controller cells sense the concentration in the environment of the reference signal, Ref, and of the targets' output, y, which is high ($y = 1$) only when the targets are active. The controllers then generate two control signals, u_1 and u_2, according to the following logic functions

$$u_1 = \text{Ref AND } (\text{NOT}\, y), \tag{5-1}$$

$$u_2 = (\text{NOT}\,\text{Ref}) \text{ AND } y, \tag{5-2}$$

so that the reference signal, can be used to toggle the switch in the targets.

In particular, a controller population, the *activators*, commands the activation of the targets when, at the same time, it perceives the presence of a specific *reference* chemical signal in the environment and the targets are inactive. The other controller population, the *deactivators*, inhibits the activity of the targets when they are active and the reference signal is no longer present in the environment. In this way, the targets are active only when the reference signal is present in the environment.

The crucial challenge we address in this section is the abstract biological implementation of this novel multicellular control scheme. After proposing a possible realization of all

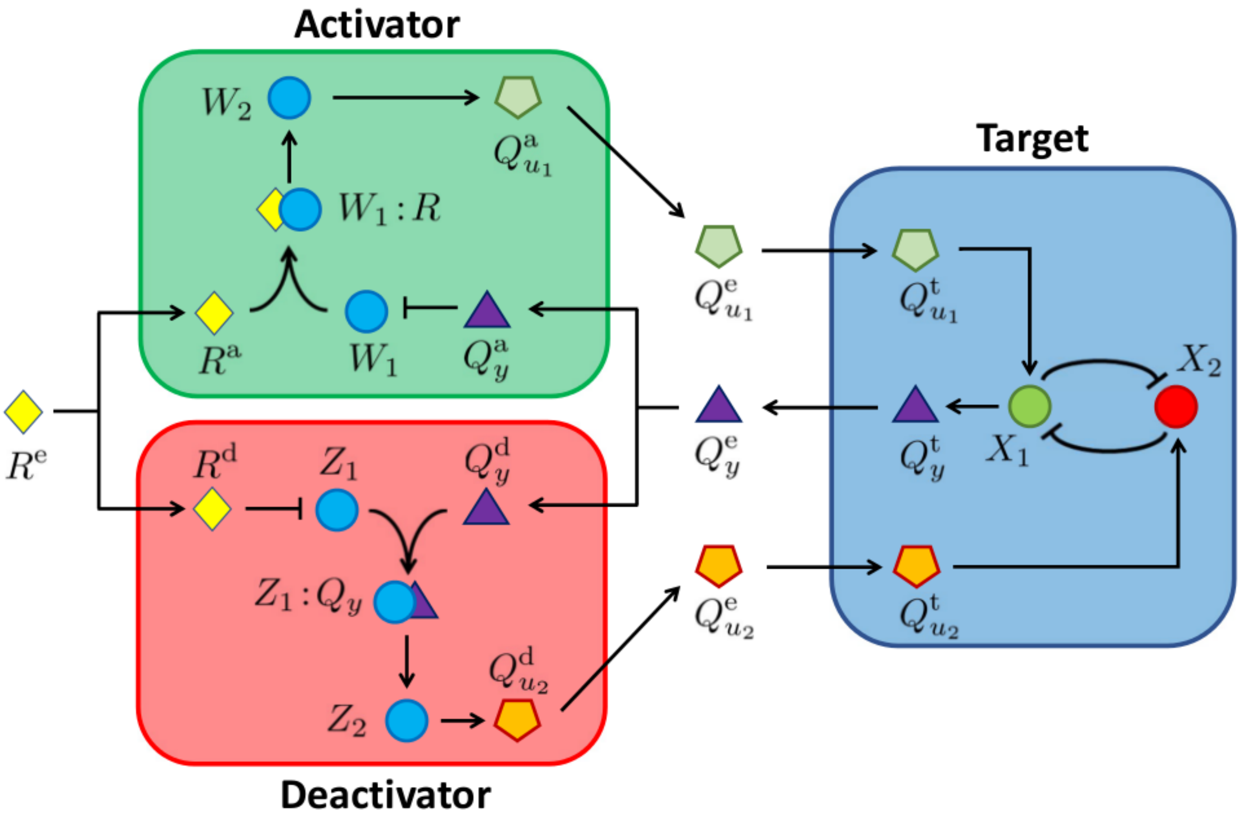


Figure 5-2: Abstract biological implementation of the multicellular control of a genetic toggle-switch. The togglers compare the concentrations of the signaling molecules R and Q_y using an antithetic motif and produce Q_{u_i} according to the logic functions (5-1)-(5-2). This, in turn, diffuses inside the target and promotes the activation of X_i, which makes the target change its state. Circles represent internal molecular species and polygons represent signaling molecules diffusing in the cells.

the functions required, we model the three populations and investigate analytically how to engineer the consortium parameters so as to guarantee its desired operation. We then provide in-silico experiments in BSim [122, 123], a realistic agent-based simulator of bacterial populations, confirming the viability of the approach.

5.1.1 Multicellular control system

A schematic biological implementation of the multicellular control strategy we propose is illustrated in Fig. **5-2**. The superscripts e, t, a, d are used in the rest of the section to denote quantities in the environment, in the target cells, in the activator cells or in the deactivator cells, respectively.

As done in [4, 89], activation and repression of each species is governed by Hill functions with dissociation coefficient θ and exponent n. Moreover, we denote with α_X^0 and α_X the basal and maximal expression rates of species X, and with γ^v the degradation rate of a species within a domain v.

Target population

We assume that the bistable memory regulating the activation of the target cells is implemented by an inducible genetic toggle-switch [124, 89]. This genetic network consists of two proteins, X_1 and X_2, each repressing the expression of the other, so that at steady state only one is fully expressed. Without loss of generality, we assume full expression of X_1 corresponds to the "active" state of the cell where some desired functionalities are turned on, while full expression of X_2 corresponds to its inactive state.

Here we focus on the problem of toggling the target population between the two states. Other works in the literature have considered the alternative control problem of stabilizing the toggle-switch about some intermediate expression levels of X_1 and X_2, see e.g. [125, 126, 127, 128]; a problem we do not address in this section.

The dynamical model of the toggle-switch can be given as

$$\dot{x}_1 = \alpha_{x_1}^0 + \frac{\alpha_{x_1}}{1 + \left(\frac{x_2}{\theta_{x_2}}\right)^{n_{x_2}}} - \gamma^{\mathrm{t}} x_1 + u_1 \tag{5-3}$$

$$\dot{x}_2 = \alpha_{x_2}^0 + \frac{\alpha_{x_2}}{1 + \left(\frac{x_1}{\theta_{x_1}}\right)^{n_{x_1}}} - \gamma^{\mathrm{t}} x_2 + u_2 \tag{5-4}$$

where the state variables x_1 and x_2 denote concentrations of molecules X_1 and X_2 inside the cell and we assume u_1 and u_2 capture the effect of two inputs that can be used to toggle the switch between one state and the other.

We assume that the parameters of the toggle-switch are chosen such that, in the absence of external inputs, i.e. $u_1 = u_2 = 0$, the system is bistable [129], with well separated equilibria and sufficient transversality of the nullclines [130]. Specifically, system (5-3)-(5-4) admits two stable equilibria, $x_1^{\mathrm{eq}} = [\bar{x}_1, \underline{x}_2]$ and $x_2^{\mathrm{eq}} = [\underline{x}_1, \bar{x}_2]$, associated to high expression of species X_1 or X_2, respectively. We also assume that there exists some positive value $\widehat{u}_1$ ($\widehat{u}_2$) such that, when $u_1 > \widehat{u}_1$ ($u_2 > \widehat{u}_2$) and $u_2 = 0$ ($u_1 = 0$), system (5-3)-(5-4) converges to a unique equilibrium point corresponding to high expression of X_1 (X_2) and remains therein when inputs are switched off.

As shown in Fig. **5-2**, we associate each of the inputs of the toggle-switch (5-3)-(5-4) in the targets to the concentration of a quorum sensing molecule coming from the activator and deactivator cells. Specifically, we capture the promoting action of the signaling molecule Q_{u_i} on the expression of X_i by setting

$$u_i = \beta_i \cdot \frac{(q_{u_i}^{\mathrm{t}})^{n_{u_i}}}{\theta_{u_i}^{n_{u_i}} + (q_{u_i}^{\mathrm{t}})^{n_{u_i}}}, \qquad i = 1, 2, \tag{5-5}$$

where $q_{u_i}^{\mathrm{t}}$ denotes the concentration of molecule Q_{u_i} inside the target cell, and β_i, θ_{u_i} and n_{u_i} are the maximal promoter activity, activation and Hill coefficients, respectively.

In our design, target cells can signal their state to the other cells by means of another, orthogonal, quorum sensing molecule Q_y that is produced at a rate f_y^{t}, assumed to be proportional to X_1 [4], that is

$$f_y^{\mathrm{t}} = k_y\, x_1, \qquad k_y > 0. \tag{5-6}$$

Hence, at steady state, when the cell is active, $f_y^{\mathrm{t,ON}} = k_y\, \bar{x}_1$.

toggler populations

The two controllers implement the same logic circuits (Fig. **5-1**) and therefore they also share similar biological implementations. However, since the reference molecule R and signaling molecule Q_y have inverted roles, the biochemical reactions describing the activator and deactivator cells are in general different. For the sake of brevity, we next describe only the biological implementation of the deactivator cells in Fig. **5-2**, which is directly taken from [4].

The logic function (5-2) is realized in the deactivators by means of an antithetic motif. Specifically, the expression of Z_1 is regulated by two independent and competing species, R and Q_y; R represses Z_1, while Q_y activates Z_1 and reacts with it forming the complex $Z_1\!:\!Q_y$. The control signal molecule Q_{u_2} is produced through a synthesis process catalyzed by Z_2, which is promoted only by the active compound $Z_1\!:\!Q_y$. As a result, Q_{u_2} is produced and released only when the concentration of R inside the deactivator cells is low while that of Q_y is high.

By denoting with z_1 and z_2 the concentrations of the species $Z_1 : Q_y$ and Z_2 in the deactivators, their dynamics can be written as

$$\begin{aligned}
\dot{z}_1 = &\left(\alpha_{z_1,r}^0 + \alpha_{z_1,r}\,\frac{\theta_{r,z_1}^{n_{r,z}}}{\theta_{r,z_1}^{n_{r,z}} + (r^{\mathrm{d}})^{n_{r,z}}}\right)\\
&\cdot\left(\alpha_{z_1,q}^0 + \alpha_{z_1,q}\,\frac{(q_y^{\mathrm{d}})^{n_{q,z}}}{\theta_{q,z_1}^{n_{q,z}} + (q_y^{\mathrm{d}})^{n_{q,z}}}\right) - \gamma^{\mathrm{d}}\, z_1\\
\dot{z}_2 = &\,\alpha_{z_2}^0 + \alpha_{z_2}\cdot\frac{z_1^{n_z}}{\theta_{z_1}^{n_z} + z_1^{n_z}} - \gamma^{\mathrm{d}}\, z_2
\end{aligned}$$

The output signaling molecule Q_{u_2} is produced through a synthesis process catalyzed by Z_2, and so at a rate $f^{\mathrm{d}}_{u_2}$ proportional to the concentration of Z_2, that is

$$f^{\mathrm{d}}_{u_2} = k_{u_2}\, z_2, \qquad k_{u_2} > 0. \tag{5-7}$$

Moreover, the substrates required to synthesize Q_{u_2} are assumed to be in excess and therefore this process does not directly affect Z_2 [4].

Similarly, another antithetic motif is embedded in the activators so that, by denoting with w_1 and w_2 the concentrations of the species $W_1 : R$ and W_2 therein, their dynamics can be written as

$$\begin{aligned}
\dot{w}_1 &= \left(\alpha^0_{w_1,q} + \alpha_{w_1,q} \frac{\theta^{n_{q,w}}_{q,w_1}}{\theta^{n_{q,w}}_{q,w_1} + (q^{\mathrm{a}}_y)^{n_{q,w}}} \right) \\
&\quad \cdot \left(\alpha^0_{w_1,r} + \alpha_{w_1,r} \frac{(r^{\mathrm{a}})^{n_{r,w}}}{\theta^{n_{r,w}}_{r,w_1} + (r^{\mathrm{a}})^{n_{r,w}}} \right) - \gamma^{\mathrm{a}}\, w_1 \\
\dot{w}_2 &= \alpha^0_{w_2} + \alpha_{w_2} \cdot \frac{w_1^{n_w}}{\theta^{n_w}_{w_1} + w_1^{n_w}} - \gamma^{\mathrm{a}}\, w_2
\end{aligned}$$

The activators will then generate a quorum sensing molecule Q_{u_1} at a rate $f^{\mathrm{a}}_{u_1}$ proportional to the concentration of W_2, that is

$$f^{\mathrm{a}}_{u_1} = k_{u_1}\, w_2, \qquad k_{u_1} > 0. \tag{5-8}$$

Intercellular communication

Intercellular communication between the three populations is realized by means of three pairwise orthogonal quorum sensing molecules, namely Q_{u_1}, Q_{u_2} and Q_y, which are produced by activators, deactivators and targets, respectively. For the sake of brevity, in what follows we use the placeholder superscript j to denote concentrations of signaling molecules in a generic cell type, where $\mathrm{j} = \mathrm{a}$ for activators, $\mathrm{j} = \mathrm{d}$ for deactivators and $\mathrm{j} = \mathrm{t}$ for targets. The quorum sensing molecules and the reference signal molecule R diffuse across the cell membrane of the genetic cell of type j with diffusion rate η^{j}. The evolution of the concentrations of the signaling molecules *inside* the generic cell of type j can then be given as

$$\begin{aligned}
\dot{r}^{\mathrm{j}} &= \eta^{\mathrm{j}}\,(r^{\mathrm{e}} - r^{\mathrm{j}}) - \gamma^{\mathrm{j}}\, r^{\mathrm{j}} \\
\dot{q}^{\mathrm{j}}_{u_1} &= f^{\mathrm{j}}_{u_1} + \eta^{\mathrm{j}}\,(q^{\mathrm{e}}_{u_1} - q^{\mathrm{j}}_{u_1}) - \gamma^{\mathrm{j}}\, q^{\mathrm{j}}_{u_1} \\
\dot{q}^{\mathrm{j}}_{u_2} &= f^{\mathrm{j}}_{u_2} + \eta^{\mathrm{j}}\,(q^{\mathrm{e}}_{u_2} - q^{\mathrm{j}}_{u_2}) - \gamma^{\mathrm{j}}\, q^{\mathrm{j}}_{u_2} \\
\dot{q}^{\mathrm{j}}_{y} &= f^{\mathrm{j}}_{y} + \eta^{\mathrm{j}}\,(q^{\mathrm{e}}_{y} - q^{\mathrm{j}}_{y}) - \gamma^{\mathrm{j}}\, q^{\mathrm{j}}_{y}
\end{aligned}$$

where the production functions f_y^{t}, $f_{u_2}^{\mathrm{d}}$, and $f_{u_1}^{\mathrm{a}}$ are defined in (5-6), (5-7) and (5-8), $f_{u_1}^{\mathrm{t}} = f_{u_1}^{\mathrm{d}} = 0$, $f_{u_2}^{\mathrm{t}} = f_{u_2}^{\mathrm{a}} = 0$, and $f_y^{\mathrm{a}} = f_y^{\mathrm{d}} = 0$.

The concentrations of the reference signal molecule and of the quorum sensing molecules secreted by the three cell populations in the environment can be described by the following set of ODEs:

$$\begin{aligned}
\dot{r}^{\mathrm{e}} &= r_{\mathrm{in}}(t) + \eta^{\mathrm{a}}\,\Delta r^{\mathrm{a,e}} + \eta^{\mathrm{d}}\,\Delta r^{\mathrm{d,e}} + \eta^{\mathrm{t}}\,\Delta r^{\mathrm{t,e}} - \gamma^{\mathrm{e}} r^{\mathrm{e}}\\
\dot{q}_{u_1}^{\mathrm{e}} &= \eta^{\mathrm{a}}\,\Delta q_{u_1}^{\mathrm{a,e}} + \eta^{\mathrm{d}}\,\Delta q_{u_1}^{\mathrm{d,e}} + \eta^{\mathrm{t}}\,\Delta q_{u_1}^{\mathrm{t,e}} - \gamma^{\mathrm{e}}\,q_{u_1}^{\mathrm{e}}\\
\dot{q}_{u_2}^{\mathrm{e}} &= \eta^{\mathrm{a}}\,\Delta q_{u_2}^{\mathrm{a,e}} + \eta^{\mathrm{d}}\,\Delta q_{u_2}^{\mathrm{d,e}} + \eta^{\mathrm{t}}\,\Delta q_{u_2}^{\mathrm{t,e}} - \gamma^{\mathrm{e}}\,q_{u_2}^{\mathrm{e}}\\
\dot{q}_{y}^{\mathrm{e}} &= \eta^{\mathrm{a}}\,\Delta q_{y}^{\mathrm{a,e}} + \eta^{\mathrm{d}}\,\Delta q_{y}^{\mathrm{d,e}} + \eta^{\mathrm{t}}\,\Delta q_{y}^{\mathrm{t,e}} - \gamma^{\mathrm{e}}\,q_{y}^{\mathrm{e}}
\end{aligned}$$

where $\Delta r^{\mathrm{j,e}} := r^{\mathrm{j}} - r^{\mathrm{e}}$, $\mathrm{j} \in \{\mathrm{a},\mathrm{d},\mathrm{t}\}$, and $\Delta q_{u_1}^{\mathrm{j,e}}$, $\Delta q_{u_2}^{\mathrm{j,e}}$, and $\Delta q_{y}^{\mathrm{j,e}}$ are defined similarly, and the function $r_{\mathrm{in}}(t)$ represents the concentration of the reference signal provided externally to influence the cell behavior. In the above equations, γ^{e} and γ^{j} are the degradation rates in the environment and in the generic cell of type j (assumed to be the same for all species for the sake of simplicity).

5.1.2 Consortium engineering

Next, we show that, for the control loop to be effectively closed across the three populations, the parameters characterizing each of the cell populations and the intercellular communication channels must fulfill a set of necessary conditions. In particular, a set of constraints on the parameters can be derived by analyzing the model equations at steady state, assuming that spatial effects are negligible and the number of cells in the three populations are equally balanced, which implies $\eta^{\mathrm{j}} = \eta$, for all $\mathrm{j} \in \{\mathrm{a},\mathrm{d},\mathrm{t}\}$. These assumptions will then be relaxed in the next section where in-silico experiments are carried out also in the presence of cell-to-cell variability and spatio-temporal effects.

Feedback loop pathways We start by making the realistic assumption that $\eta \gg \gamma^{\mathrm{j}}$, $\mathrm{j} \in \{\mathrm{a},\mathrm{d},\mathrm{t}\}$, that is, the signaling molecules diffuse through the cellular membrane faster then they are degraded. Hence, when the reference signal fed to the environment is constant and large enough (i.e. $r_{\mathrm{in}}(t) = r_{\mathrm{in}}^{\mathrm{ON}} = \mathrm{const.}$) and the target cells are not active (i.e. Q_y is not expressed), it is easy to verify that at steady state the concentrations of the signaling molecules R and Q_{u_1} reach the same value in every cells, that is, for all $\mathrm{j} \in \{\mathrm{a},\mathrm{d},\mathrm{t}\}$ we have

$$\bar{r}^{\mathrm{j}} = r_{\mathrm{in}}^{\mathrm{ON}}/\Gamma, \qquad \bar{q}_{u_1}^{\mathrm{j}} = k_{u_1}\,\bar{w}_2/\Gamma, \tag{5-9}$$

where $\Gamma := \gamma^{\mathrm{e}} + \gamma^{\mathrm{t}} + \gamma^{\mathrm{a}} + \gamma^{\mathrm{d}}$, and $\bar{w}_2$ denotes the steady-state value of w_2 when it is fully expressed.

Analogously, for Q_y and Q_{u_2}, when the reference signal r_{in} is absent (i.e. $r_{\mathrm{in}}(t) = r_{\mathrm{in}}^{\mathrm{OFF}} =$ const.) and the target cells are initially active (i.e. Q_y is expressed), at steady state, we have for all $\mathrm{j} \in \{\mathrm{a}, \mathrm{d}, \mathrm{t}\}$

$$\bar{q}_y^{\mathrm{j}} = k_y\, \bar{x}_1/\Gamma, \qquad \bar{q}_{u_2}^{\mathrm{j}} = k_{u_2}\, \bar{z}_2/\Gamma, \tag{5-10}$$

where, similarly, $\bar{x}_1$ and $\bar{z}_2$ denote the steady-state values of x_1 and z_2 when they are fully expressed.

toggler cells To guarantee that the activators and deactivators implement at steady state the logic functions (5-1)-(5-2), it is necessary that only the activators produce their control signal when the concentrations in the cells of the reference molecule R and Q_y are sufficiently high and low, respectively. Therefore, it must hold that

$$\theta_{r,w_1} \ll \bar{r}^{\mathrm{a}} \quad \text{and} \quad \theta_{r,z_1} \ll \bar{r}^{\mathrm{d}}. \tag{5-11}$$

Similarly, when the concentrations of R and Q_y are sufficiently low and high, respectively, then, in order that only the deactivators generate their control signal, it must hold that

$$\theta_{q,w_1} \ll \bar{q}_y^{\mathrm{a}} \quad \text{and} \quad \theta_{q,z_1} \ll \bar{q}_y^{\mathrm{d}}. \tag{5-12}$$

Target cells In order for the signaling molecules coming from the controllers to toggle the switch within the targets, the input functions (5-5) must be such that

$$\beta_i > 2\,\widehat{u}_i, \qquad i = 1, 2, \tag{5-13}$$

and the concentrations of the quorum sensing molecules within the targets must be sufficiently high, i.e.

$$\theta_{u_i} \ll \bar{q}_{u_i}^{\mathrm{t}}, \qquad i = 1, 2, \tag{5-14}$$

so that the control input is strong enough to trigger the transition from one state to the other and render the toggle-switch monostable.

Parameters' constraints Substituting equations (5-9)-(5-10) in conditions (5-11), (5-12) and (5-14), we obtain that, at steady state, the togglers can activate or deactivate the targets in response to the presence or absence of the external reference signal $r_{\text{in}}(t)$, if the system parameters satisfy (5-13) and the following conditions are satisfied:

$$\begin{aligned} \Gamma\,\theta_{r,w_1} &\ll r_{\text{in}}^{\text{ON}}, & \Gamma\,\theta_{r,z_1} &\ll r_{\text{in}}^{\text{ON}}, \\ \Gamma\,\theta_{q,w_1} &\ll k_y\,\bar{x}_1, & \Gamma\,\theta_{q,z_1} &\ll k_y\,\bar{x}_1, \\ \Gamma\,\theta_{u_1} &\ll k_{u_1}\,\bar{w}_2, & \Gamma\,\theta_{u_2} &\ll k_{u_2}\,\bar{z}_2. \end{aligned}$$

Using similar arguments, it is also possible to obtain the lower bounds, yielding

$$r_{\text{in}}^{\text{OFF}} \ll \Gamma\,\theta_{r,w_1} \ll r_{\text{in}}^{\text{ON}}, \qquad r_{\text{in}}^{\text{OFF}} \ll \Gamma\,\theta_{r,z_1} \ll r_{\text{in}}^{\text{ON}} \tag{5-15}$$

$$k_y\,\underline{x}_1 \ll \Gamma\,\theta_{q,w_1} \ll k_y\,\bar{x}_1, \qquad k_y\,\underline{x}_1 \ll \Gamma\,\theta_{q,z_1} \ll k_y\,\bar{x}_1 \tag{5-16}$$

$$k_{u_1}\,\underline{w}_2 \ll \Gamma\,\theta_{u_1} \ll k_{u_1}\,\bar{w}_2, \qquad k_{u_2}\,\underline{z}_2 \ll \Gamma\,\theta_{u_2} \ll k_{u_2}\,\bar{z}_2 \tag{5-17}$$

where $\underline{x}_1$, $\underline{w}_2$ and $\underline{z}_2$ denote the steady-state values of the corresponding species when they are completely repressed.

The previous conditions represent a set of necessary conditions for the consortium to exhibit its desired multicellular control functions.

Remark 1 *Conditions (5-15)-(5-17) depend on steady-state values of x_1, w_2 and z_2, which in general would need to be estimated in-silico or quantified experimentally. However, at the price of relaxing the bounds, conditions independent from these values can be obtained by approximating Hill functions with step functions (i.e. by letting $n \to \infty$) yielding $\underline{x}_1 = \alpha_{x_1}^0/\gamma^{\text{t}}$, $\underline{w}_2 = \alpha_{w_2}^0/\gamma^{\text{a}}$, $\underline{z}_2 = \alpha_{z_2}^0/\gamma^{\text{d}}$, $\bar{x}_1 = (\alpha_{x_1}^0 + \alpha_{x_1})/\gamma^{\text{t}}$, $\bar{w}_2 = (\alpha_{w_2}^0 + \alpha_{w_2})/\gamma^{\text{a}}$ and $\bar{z}_2 = (\alpha_{z_2}^0 + \alpha_{z_2})/\gamma^{\text{d}}$.*

5.1.3 *In silico* experiments

Agent-based simulations

To validate the effectiveness of our multicellular control design, we implemented a set of in-silico experiments via BSim, a realistic agent-based simulator of bacterial populations developed in [122, 123]. In so doing, we modeled a microfluidics chamber of dimensions $13.3\,\mu\text{m} \times 16.6\,\mu\text{m} \times 1\,\mu\text{m}$ and used BSim to take into account cell growth and division, spatial effects, diffusion of the signaling molecules, cell-to-cell variability and geometric constraints (details on the microfluidic chamber simulated can be found in section B.1). The nominal values of the parameters used in simulations are reported in Table **5-1**. They have been

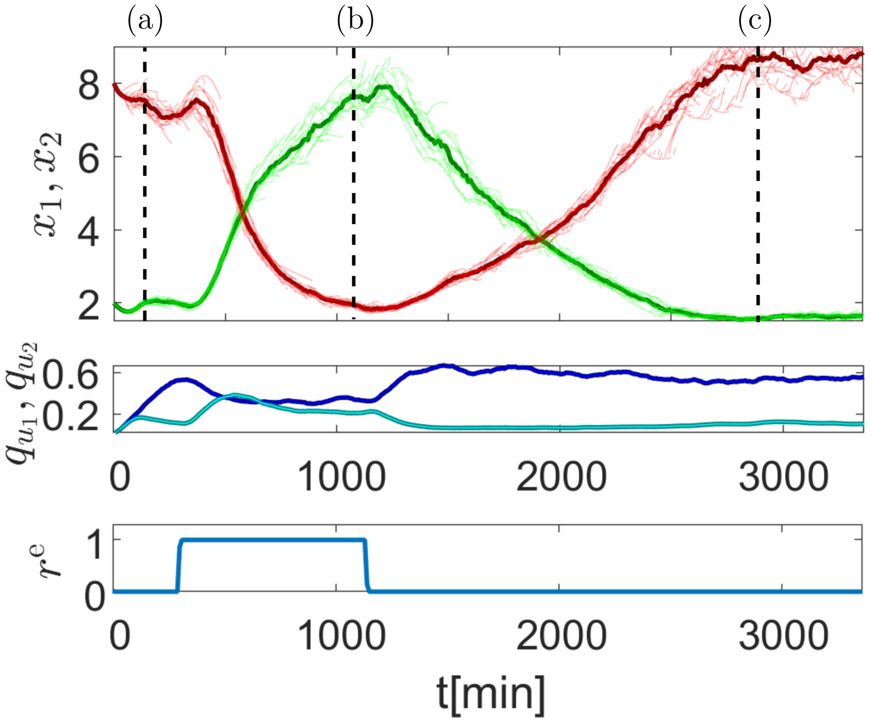


Figure 5-3: Evolution of the average (thick lines) and single cell (thin lines) values of the concentrations of repressor proteins x_1 (green) and x_2 (red) in the target population (top panel) when the reference signal $r_{\text{in}}(t)$ is switched from low to high and vice versa. The middle panel shows the average values of the concentrations of quorum sensing molecules $q^{\text{t}}_{u_1}$ (light blue) and $q^{\text{t}}_{u_2}$ (dark blue) inside the target cells. The bottom panel shows the value of r^{e} at the center of the chamber.

chosen in the ranges reported in [4, 128], and satisfying conditions (5-13) and (5-15)-(5-17). Cell-to-cell variability was modeled by assigning a different set of parameters to daughter cells when they split from their mothers. Namely, each of their parameters, say μ, was drawn independently from a normal distribution centered at its nominal value $\bar{\mu}$ and with coefficient of variation $c_{\text{v}} = 10\%$.

Figs. **5-3** and **5-4** show the results of a typical in-silico experiment where the toggler cells (depicted in red and green) successfully flip the target cells from their active state (depicted in blue) to their inactive state (depicted in black) and vice versa, following changes in the reference signal r_{in}. The amplitude and the duration of the reference pulse r_{in} have been heuristically set to 43 μM and 1140 min, respectively. (Tuning of the pulse features can also be done by means of other methods, e.g. [131].)

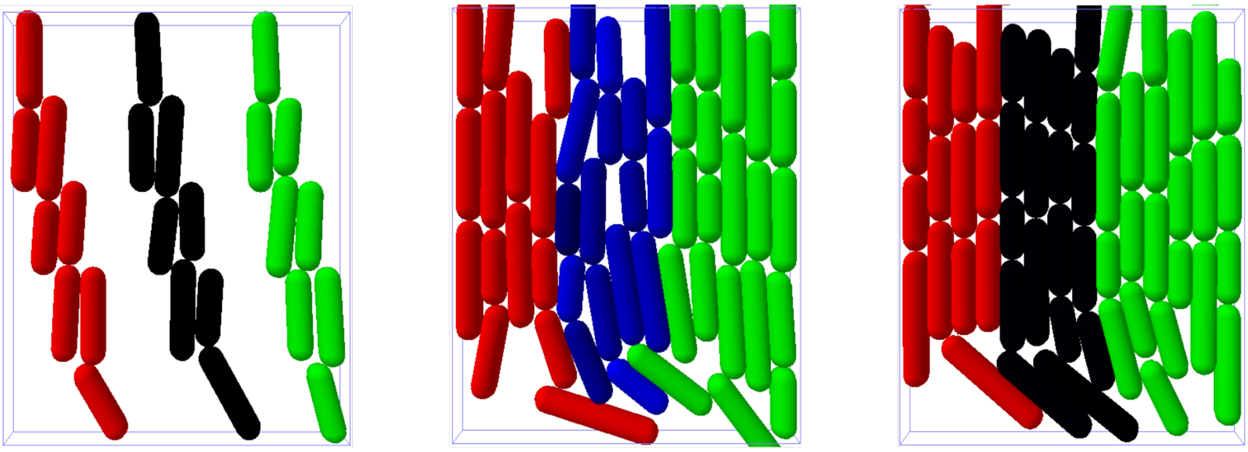

(a) (b) (c)

Figure 5-4: Snapshots of an agent-based simulation at different time instants (highlighted in Fig. 5-3 with dashed vertical lines). Specifically, panel (a) corresponds to $t = 120$ min, panel (b) to $t = 1100$ min and panel (c) to $t = 2800$ min. activator cells are shown in green, deactivator cells in red and target cells are depicted in blue when they are active and in black when they are inactive.

Robustness to parameter variations

Next, we performed numerical analysis in Matlab, in the illustrative case of no population growth, to evaluate (i) how imbalances between populations affect the operation of the consortium due to poor intercellular communication, and (ii) its robustness to parameters' variations.

Fig. 5-5 shows the values at steady state of the ratio x_1/x_2 when the targets are switched OFF (lower triangular panel) and ON (upper triangular panel), respectively, following the application of the corresponding reference signal r_{in}, as the ratios of the cell populations in the consortium are being varied. In this scenario, we see that for a wide range of population densities (black region for deactivators in panel (a), colored for activators in panel (b) of Fig. 5-5), the togglers are effectively able to flip the state of the targets.

Finally, we tested robustness of our design when all parameters of targets, activators and deactivators are perturbed from their nominal values. As shown in Fig. 5-6, even in the presence of a consistent parameter mismatch ($c_{\text{v}} = 0.2$), the togglers are able to activate or deactivate a large fraction of the targets' population. The activators showed better performance than the deactivators due to different values of parameters β_i considered in the in-silico experiments (see Table 5-1).

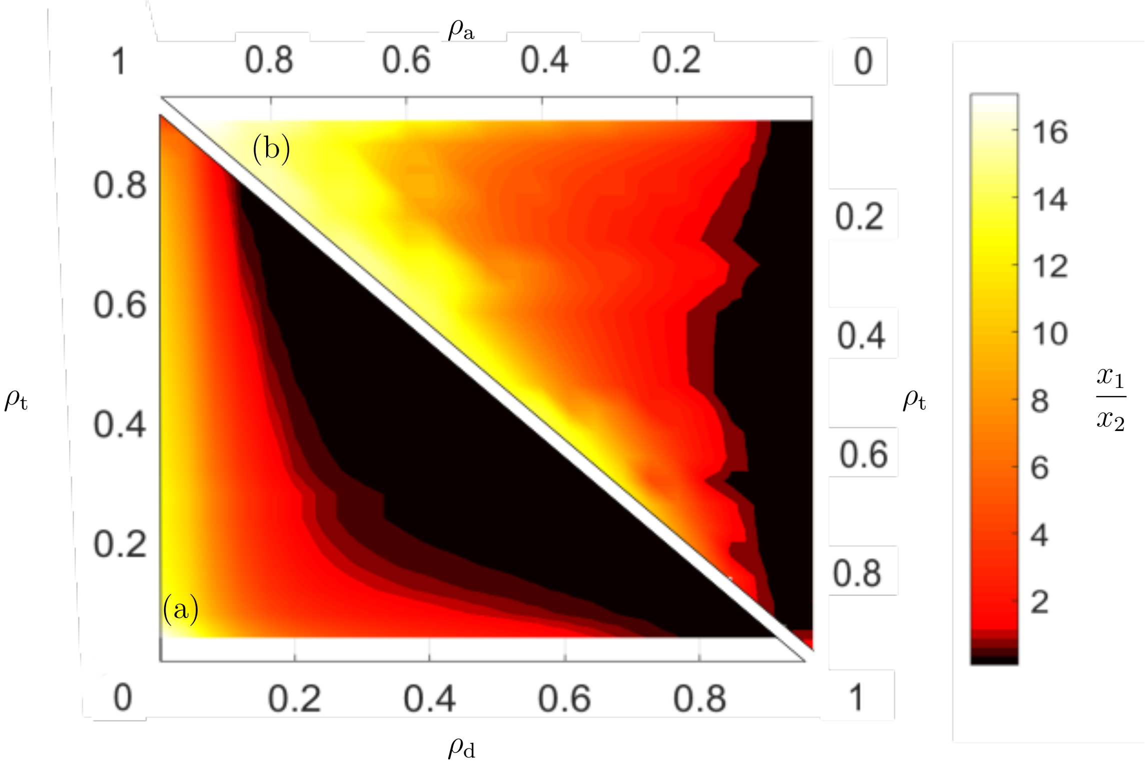


Figure 5-5: Steady-state value of x_1/x_2 in response to switch commands, $r_{\mathrm{in}}^{\mathrm{OFF}}$ (lower triangular panel a) and $r_{\mathrm{in}}^{\mathrm{ON}}$ (upper triangular panel b), as the relative ratios of the three populations are changed. Switch commands are applied with the targets starting from the opposite state. The total cell population in the consortium is set to $N = 50$. For each cell type $\rho_{\mathrm{j}} = N_{\mathrm{j}}/N$, $\mathrm{j} = \{\mathrm{t}, \mathrm{a}, \mathrm{d}\}$, is its relative ratio within the consortium, and such that $\rho_{\mathrm{a}} + \rho_{\mathrm{d}} + \rho_{\mathrm{t}} = 1$.

5.1.4 Discussion

We have presented a multicellular control solution to the problem of toggling a memory mechanism in a target cell population. This confirms, by means of realistic agent-based simulations, that the solution we propose is theoretically viable.

A possible future application for the described design could the controlled delivery of active molecules or drugs synthesized by the targets when they are active. Indeed, the consortium is designed so that the drug or active molecule of interest is only produced and secreted when a specific reference chemical signal is perceived by the controller cells, while the production is stopped when the reference is removed and the targets are deactivated. By using a cancer biomarker as the reference signal, the togglers could then be used to activate the targets to deliver chemotherapy drugs in situ only when the biomarker is detected in the tissue, providing a multicellular feedback control alternative to the open-loop design

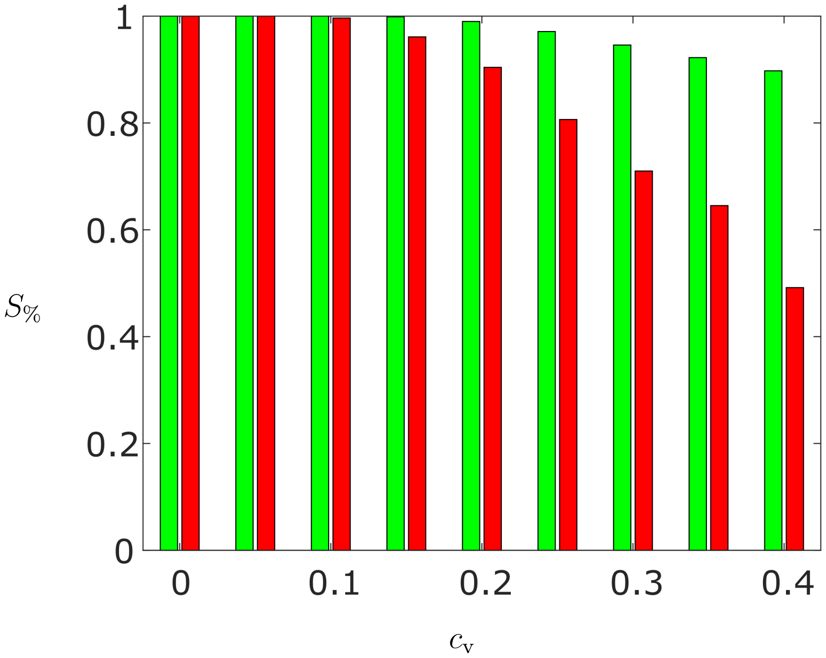


Figure 5-6: Percentage of successfully switched targets ($S_{\%}$) in a balanced consortium ($N_{\mathrm{t}} = N_{\mathrm{a}} = N_{\mathrm{d}} = 17$) as the coefficient of variation c_{v} is varied. The bar plot in green (red) represents the percentage at steady state of targets that, starting from OFF (ON) state, are turned ON (OFF) following the reference input r_{in} being switched to $r_{\mathrm{in}}^{\mathrm{ON}}$ ($r_{\mathrm{in}}^{\mathrm{OFF}}$). For each value of c_{v}, the results of 100 simulations were averaged, each obtained by drawing independently all cells' parameters from normal distributions centered on their nominal values, $\bar{\mu}$, and with standard deviation $\sigma = c_{\mathrm{v}} \cdot \bar{\mu}$.

proposed in [120]. Also, if the reference signal were linked to the presence of some pollutant in the environment, the controlled activation of the targets cells could be used to synthesize active molecules for bioremediation when and where needed.

5.2 Multicellular PI

In chapter 4 we presented a possible *in vivo* implementation of a multicellular implementation of an antithetic integral feedback controller. Although integral control allows achieving perfect robust adaptation, the dynamical performances as well as the accuracy and robustness of the controller can be enhanced by the addition of proportional and derivative actions. In this section, we first present a multicellular scheme of the PID controller inspired from the embedded single-cell solution in [9]. Then, we focus on the implementation of P and PI controllers within a microbial consortium comprising different cell populations communicating through orthogonal quorum sensing molecules. After presenting abstract biological imple-

Table 5-1: Nominal simulation parameters of the multicellular consortium

	[μM/min]		[μM]
$\alpha^0_{x_1}$, $\alpha^0_{x_2}$	0.005	$\theta_{x_1}, \theta_{x_2}$	2
α_{x_1}, α_{x_2}	0.1083	$\theta_{u_1}, \theta_{u_2}$	0.5
$\alpha^0_{z_1,q}$, $\alpha^0_{z_1,r}$	0.0348	$\theta_{r,z1}, \theta_{q,z1}$	0.5
$\alpha_{z_1,q}$, $\alpha_{z_1,r}$	0.1305	$\theta_{r,w1}, \theta_{q,w1}$	0.5
$\alpha^0_{w_1,q}$, $\alpha^0_{w_1,r}$	0.0348	$\theta_{z_1}, \theta_{w_1}$	0.5
$\alpha_{w_1,q}$, $\alpha_{w_1,r}$	0.1305		[1/min]
$\alpha^0_{z_2}$, $\alpha^0_{w_2}$	0.0016	γ^{t}	0.0092
α_{z_2}, α_{w_2}	0.026	$\gamma^{\mathrm{d}}, \gamma^{\mathrm{a}}, \gamma^{\mathrm{e}}$	0.0230
β_1	0.09	η	2
β_2	0.016	k_y	0.03
n	2	k_{u_1}	0.06
		k_{u_2}	0.06

mentations for each of the proposed strategies, we derive analytical conditions for tuning the controller gains, providing insights on the biological parameters that most influence their performance. We complement the theoretical derivations with a set of *in silico* experiments carried out using the realistic agent-based microbial simulator BSim [122, 123].

The results of all the experiments confirm the effectiveness of the proposed multicellular architectures whose *in vivo* implementation is the subject of ongoing research.

5.2.1 Multicellular PID control strategy

We propose to realize a distributed biological PID controller entrusting each action to a different cellular population within a microbial consortium, see Fig. 5-7. Here, three cellular populations, denoted as *controllers*, provide the proportional, integral and derivative actions, respectively, implementing a biomolecular multicellular controller inspired by the classical PID control strategy given by:

$$\begin{aligned} u_{PID}(t) :&= u_P(t) + u_I(t) + u_D(t) \\ &= k_P e(t) + k_I \int_0^t e(\tau) d\tau + k_D \dot{e}(t), \end{aligned} \tag{5-18}$$

where k_P, k_I and k_D are the proportional, integral and derivative gains, and $e(t)$ is the control error, which is a function of the measured output $y(t)$ of the process $\Phi(t)$ under control and the desired value $Y_{\mathrm{d}}(t)$. The overall control signal $u_{PID}(t)$ computed by the controllers is sensed by the target population hosting the process $\Phi(t)$, whose output $y(t)$ is fed back to the controllers closing the control loop.

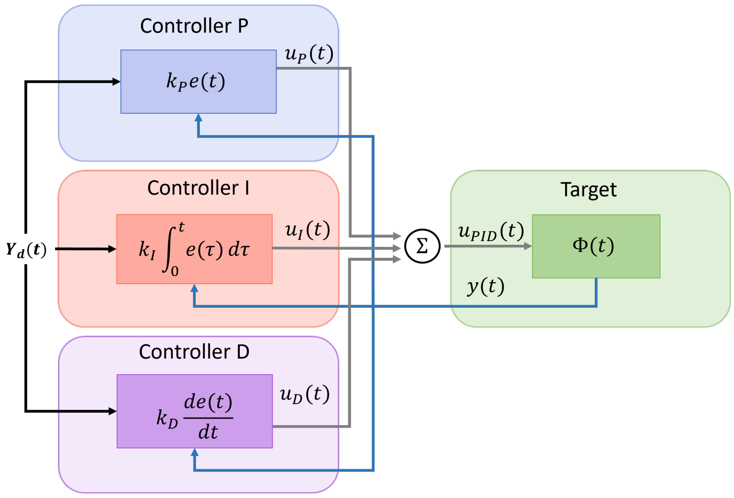


Figure 5-7: Schematic representation of a distributed biological PID controller. The three controllers compare the reference signal $Y_\mathrm{d}(t)$ with the output of the target population $y(t)$, and collectively compute the overall control signal $u_{PID}(t)$, closing the control loop and regulating the process $\Phi(t)$.

Here, we present two biomolecular control architectures stemming from the full schematic in Fig. 5-7; the former composed by one population implementing a Proportional (P) action that controls the targets, and the latter in which an additional population is inserted in the consortium to implement an integral (I) action (Fig. 5-8). For the sake of brevity, we leave the study of other possible configurations for a future work. Obviously, because of the highly nonlinear nature of biomolecular reactions the resulting controller will have a much more cumbersome and nonlinear nature than the classical PID expressed in (5-18).

The communication between controllers and targets is realized by a pair of orthogonal quorum sensing molecules Q_u and Q_x that act as proxies of the control input $u_{PID}(t)$ and the measurement of the process state $y(t)$, respectively, that are produced by the cells and diffuse through their membranes into the environment.

Each controller population senses the control error $e(t)$ by comparing the reference signal $Y_\mathrm{d}(t)$ with measurement of the output carried by the *sensing* quorum sensing molecule Q_x. It then contributes accordingly to the overall production of the quorum sensing molecule Q_u which delivers the control input affecting the target cells. Therein, the process $\Phi(t)$ can be any network of genes that is directly affected by Q_u and whose output is the expression of some gene of interest. To close the loop, the process $\Phi(t)$ activates the production of Q_x that, by diffusing into the environment, can be sensed by the controllers.

Next, we derive the mathematical models of both schemes, describing the *average* dynamics of the populations, that is, the evolution of the concentrations of the chemical species

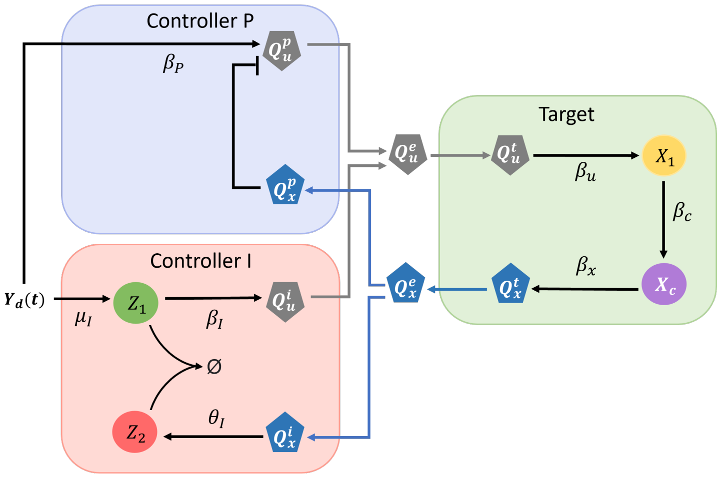


Figure 5-8: Abstract implementation of a distributed biological PI controller. The output of the process $\Phi(t)$ is the quorum sensing molecule Q_x, produced proportionally to the target gene X_c. Its input is the gene X_1, which is actuated by the *control* quorum sensing molecule Q_u. Each controller population evaluates the control error $e(t)$ by comparing the reference signal $Y_{\mathrm{d}}(t)$ and the process output carried by Q_x, thus contributing to the overall production of Q_u. Circles represent internal molecular species, while polygons represent the signaling molecules.

averaged over the entire consortium. We assume that all populations in the consortium are equally balanced and that the total number of cells in the community is fixed. This also implies that the molecules diffuse through the cell membrane with the same diffusion rate η. These assumptions are later relaxed in Section 5.2.3 where *in silico* experiments are carried out to evaluate the impact of cell-to-cell variability and spatio-temporal effects on the control performance. Note that in *in vivo* experiments in microfluidics, after an initial growth phase during which the cells occupy the entire chamber, as newborn cells push out older cells from the chamber, their total number can be assume to be constant. The superscripts e, t, p, i, are used in the rest of this section to refer to quantities in the environment, in the target cells, in the proportional, or in the integral controller cells, respectively.

Mathematical modelling We assume that the output of the process $\Phi(t)$ is the quorum sensing molecule Q_x, produced proportionally to the target gene X_c, and that its input is the gene X_1, which is actuated by Q_u (Fig. 5-8). The dynamics of the network hosted in the target cells can be described by the following set of ODEs:

$$\begin{aligned} \dot{X}_1 &= -\gamma_1 X_1 + \beta_u Q_u^t, \\ \dot{X}_c &= \beta_c X_1 - \gamma_c X_c, \end{aligned} \tag{5-19}$$

where γ_1 and γ_c are the degradation rates of the species X_1 and X_c, respectively, and β_c and β_u are activation rates, modeling the strength of the activation induced by transcription factors X_1 and Q_u [8, 132, 9]. The sensing molecule and process output Q_x is produced by the target cells as a function of the species X_c (for the sake of simplicity we assume the production of Q_x is proportional to X_c). Hence, information about the target state is broadcast to the other cells in the consortium by Q_x diffusing into the environment with a dynamics assumed here to be linear, as often done in literature [110]. Under these assumptions, the dynamics of the sensing molecule in the targets can be modeled as:

$$\dot{Q}_x^t = \beta_x X_c + \eta(Q_x^e - Q_x^t) - \gamma_t Q_x^t, \tag{5-20}$$

where β_x is the activation rate due to X_c, η is the diffusion rate of the molecule Q_x across the cell membrane, and γ_t is the dilution rate of Q_x in the target cells.

The proportional controller is implemented here as in [9] by means of an inhibitory action whose strength is a nonlinear function depending on the output fed back from the targets combined with an activation proportional to the amplitude of the reference signal. Namely, we can describe the dynamics of the intracellular concentration Q_u^p as:

$$\dot{Q}_u^p = \beta_P Y_\mathrm{d} \tfrac{\mu_P Y_\mathrm{d}}{\mu_P Y_\mathrm{d} + \theta_P Q_x^p} + \eta(Q_u^e - Q_u^p) - \gamma_p Q_u^p, \tag{5-21}$$

where γ_p is the dilution rate of Q_u inside the proportional controller cells, β_P represents the maximal production rate of the control molecule and plays the role of the proportional gain, μ_P and θ_P are positive coefficients characterizing the control action [9]. It can be shown (see [9] for details) that the first term in (5-21) embeds a nonlinear action that is function of the control error $e_P(t) = \mu_P Y_\mathrm{d} - \theta_P Q_x^p$. Note that a possible biological implementation of this action is also proposed in [9].

To implement the integral action, an antithetic motif [72] is embedded into the controller population. This module uses a pair of chemical species, say Z_1 and Z_2, produced proportionally to the reference signal Y_d and to the sensing molecule Q_x, respectively, able to annihilate each other with a high affinity. The dynamics of the network embedded in this population can then be described by the set of ODEs:

$$\begin{aligned} \dot{Z}_1 &= \mu_I Y_\mathrm{d} - \gamma_z Z_1 Z_2, \\ \dot{Z}_2 &= \theta_I Q_x^i - \gamma_z Z_1 Z_2, \end{aligned} \tag{5-22}$$

in which $\mu_I Y_\mathrm{d}$ and $\theta_I Q_x^i$ are production rates, γ_z is the annihilation rate between Z_1 and Z_2, which is assumed here for the sake of simplicity to be the only source of degradation for the species Z_1 and Z_2, as also done in [8, 9] (dilution of Z_1 and Z_2 in (5-22) can also be considered

[82] to describe a more realistic, non-ideal integrator). Following similar arguments to those in [9], it can be shown that $\frac{d}{dt}(Z_1 - Z_2)$ is proportional to the control error, defined as $e_I(t) = \mu_I Y_\mathrm{d} - \theta_I Q_x^i$.

The network described in (5-22) is complemented with the control molecule dynamics, described as:

$$\dot{Q}_u^i = \beta_I Z_1 + \eta(Q_u^e - Q_u^i) - \gamma_i Q_u^i, \tag{5-23}$$

where β_I plays the role of the integral gain and all the other parameters have an analogous meaning as those in (5-20).

Finally, the mathematical models are completed by adding ODEs describing the diffusion dynamics of the quorum sensing molecules across the cell membranes. Specifically, under the assumption that the two populations are balanced, the concentration of the molecules into the environment is described by:

$$\begin{aligned} \dot{Q}_x^e &= M\eta \textstyle\sum_{\mathrm{j}\in\mathcal{S}}(Q_x^\mathrm{j} - Q_x^e) - \gamma_e Q_x^e, \\ \dot{Q}_u^e &= M\eta \textstyle\sum_{\mathrm{j}\in\mathcal{S}}(Q_u^\mathrm{j} - Q_u^e) - \gamma_e Q_u^e, \end{aligned} \tag{5-24}$$

where $\mathcal{S} = \{p, t\}$ if the proportional controller population is the only present, while $\mathcal{S} = \{p, i, t\}$ when the integral controller population is also added to the consortium, and M is the number of cells in each population. The dynamics of the concentrations of Q_x and Q_u inside those cells not directly producing them is given by:

$$\begin{aligned} \dot{Q}_x^\mathrm{h} &= \eta(Q_x^e - Q_x^\mathrm{h}) - \gamma_\mathrm{h} Q_x^\mathrm{h}, \quad \mathrm{h} \in \{p, i\}, \\ \dot{Q}_u^t &= \eta(Q_u^e - Q_u^t) - \gamma_t Q_u^t. \end{aligned} \tag{5-25}$$

5.2.2 Circuit Design

We derive some analytical conditions on the parameters of the genetic circuits which guarantee successful regulation of the measured output Q_x to the desired value. Towards this aim, we first derive a reduced order model of the consortium dynamics and then, via a stability analysis, we provide sufficient conditions that the biomolecular parameters must satisfy in order for the consortium to operate correctly.

Assumptions and problem statement To derive simple, yet meaningful, analytical conditions guiding the design of the controller populations, we make some realistic assumptions on the values of the parameters. Namely, we assume that (i) each population divides at the same rate, implying that the dilution rates for all species are identical (i.e. $\gamma_p = \gamma_i = \gamma_t = \gamma_1 = \gamma_c = \gamma$); (ii) the degradation of each quorum sensing molecule in the external growth environment can be neglected, i.e. $\gamma_e = 0$; and (iii) the promoters induced by the reference signal Y_d and the sensing molecule Q_x are the same in the proportional and integral controllers, implying that $\mu_p = \mu_i = \mu$ and $\theta_p = \theta_i = \theta$. Also, defining $\Gamma_P := \gamma_t + \gamma_p + \gamma_e = 2\gamma$ and $\Gamma_{PI} := \gamma_t + \gamma_p + \gamma_i + \gamma_e = 3\gamma$, we assume that:

Assumption 1 *The quorum sensing molecules diffuse faster than they degrade, that is, $\eta \gg \Gamma_P$ when only proportional controller cells are present, or $\eta \gg \Gamma_{PI}$ when both controller populations are present.*

Assumption 2 *The annihilation process between the species Z_1 and Z_2 is fast enough, that is, $\gamma_z \gg \max\{\frac{\alpha^2}{\mu Y}, \frac{\Gamma_{PI}\alpha}{\mu Y}\}$, where α is a function of the model parameters defined as*

$$\alpha = \left(\frac{9\gamma^4}{\beta_x\beta_I\beta_c\beta_u} - \frac{\beta_P\theta}{2\beta_I\mu}\right)^{-1}.$$

Note that in different models parameterized from *in vivo* experiments this assumption holds (e.g., [110, 5]).

Given the assumptions above and defining the control error as:

$$e(t) := \mu Y_\mathrm{d} - \theta y, \tag{5-26}$$

where we defined the controlled output as $y = Q_x^t$, the control problem is that of engineering a microbial consortium comprising one or more controller populations so that at steady state the output of the process in the target cells is robustly regulated to the desired value, that is, $\lim_{t\to\infty} e(t) = 0$, with a transient response characterized by small overshoot and fast convergence.

Multicellular P controller Under the assumptions made in Section 5.2.2 and quasi-steady state assumption on the quorum sensing dynamics, when the targets are regulated by controllers solely implementing a proportional control action, the dynamics of the consortium can be approximated by the following reduced order model:

$$\begin{aligned} \dot{X}_1 &= \beta_u\beta_P Y_\mathrm{d}\frac{\mu Y_\mathrm{d}}{\Gamma_P\mu Y_\mathrm{d}+\theta\beta_x X_c} - \gamma X_1, \\ \dot{X}_c &= \beta_c X_1 - \gamma X_c, \end{aligned} \tag{5-27}$$

where the controlled output is defined as $y = \frac{\beta_x}{\Gamma_P}X_c$ (see Appendix A.2.1 for details). It can be demonstrated that model (5-27) has a unique admissible equilibrium point, which is always locally asymptotically stable. However, for the sake of brevity, the proof is omitted here and reported in [14]. Moreover the steady-state error e_ss is given as:

$$e_\mathrm{ss} = \frac{1}{2}\left(3 - \sqrt{1 + \beta_P\frac{\beta_c\beta_u\beta_x\theta}{\mu\gamma^4}}\right)\mu Y_\mathrm{d}, \tag{5-28}$$

which nonlinearly depends on the value of the proportional gain β_P and that can be made closer to zero by selecting β_P as:

$$\beta_P^* = \frac{8\mu\gamma^4}{\beta_c\beta_u\beta_x\theta}. \tag{5-29}$$

Indeed, if it were possible to select $\beta_P = \beta_P^*$, we could have $Q_x = \frac{\mu Y_\mathrm{d}}{\theta}$ at steady state, which implies $e(t) = 0$. Note that such a choice would require perfect knowledge of the target cells parameters which is unrealistic. We will therefore explore later in Section 5.2.3 how the error varies for values of β_P in a given range of interest and evaluate the effects of β_P on the dynamics of Q_x^t, assessing the sensitivity of the control strategy to parameter mismatches or uncertainties.

Multicellular PI controller A possible solution to overcome model uncertainties and robustify the designed control system is to add a third population implementing an integral control action. Under the same assumptions made in Section 5.2.2, the dynamics of the resulting consortium comprising both the P and I controller populations can be approximated by the following set of ODEs:

$$\begin{aligned} \dot{X}_1 &= \beta_u \left[\beta_P Y_\mathrm{d} \tfrac{\mu Y_\mathrm{d}}{\Gamma_{PI}\mu Y_\mathrm{d} + \theta\beta_x X_c} + \tfrac{\beta_I \zeta_1}{\Gamma_{PI}}\right] - \gamma X_1, \\ \dot{X}_c &= \beta_c X_1 - \gamma X_c, \\ \dot{\zeta}_1 &= \mu Y_\mathrm{d} - \theta Q_x, \end{aligned} \tag{5-30}$$

where $\zeta_1 = Z_1 - Z_2$ and $y = \frac{\beta_x X_c}{\Gamma_{PI}}$. Details on the derivation of equation (5-30) are reported in Appendix A.2.1. This dynamical system has a unique, non-negative equilibrium point if the proportional gain is chosen such that:

$$\beta_P \leq \frac{18\gamma^4\mu}{\beta_c\beta_u\beta_x\theta}. \tag{5-31}$$

Reaching this equilibrium point ensures that $Q_x = \frac{\mu Y_\mathrm{d}}{\theta}$, thus $e(t) = 0$. By carrying out a local stability analysis, we found that the equilibrium point is locally asymptotically stable if the value of the integral gain β_I does not exceed a threshold dependent on the other parameters including the proportional gain β_P, that is:

$$\beta_I < \frac{\beta_P\gamma}{2\mu} + \frac{18\gamma^5}{\beta_c\beta_u\beta_x\theta}. \tag{5-32}$$

Condition (5-32) gives insights on the biological elements that influence the performance of the PI multicellular architecture. Specifically, choosing fast dividing cells (i.e. high values for γ) or reducing of the strength of the promoters induced by the reference Y_d and the sensor molecule Q_x can widen the range of values of β_I that guarantee the correct operation of the control consortium.

5.2.3 In silico control experiments

To validate the proposed multicellular control architectures we carried out *in silico* experiments in BSim [122, 123], an agent-based environment explicitly designed to simulate bacterial populations. BSim allows to keep track of each cell in the consortium, simulating both the dynamical processes hosted in the organism and the bio-mechanics of the cell. In addition, the diffusion of the quorum sensing molecules, cell growth and division, and cell-to-cell variability can be explicitly simulated together with realistic geometric constraints of the host environment. In particular, BSim accurately simulates cells cultured in a micro-environment such as a microfluidic device where nutrients are constantly provided, allowing cells to grow in exponential phase (details on the simulation environment can be found in section B.1). Here, we emulated a scaled version of the device described in [133, 12], that is, a microfluidic chamber of dimensions $17\,\mu\text{m} \times 15\,\mu\text{m} \times 1\,\mu\text{m}$, which can host around 100 cells. These dimensions were selected as a good trade-off between number of cells hosted and computational burden. Unless otherwise stated, the growth and mechanical parameters were selected as in [5], while the nominal values of the parameters in the network were chosen as described in Appendix A.2.2.

Selecting $\beta_P = \beta_P^*$, both control architectures showed good regulation capabilities of the output species Q_x^t over a period of 12000 min at different set-point values (Fig. **5-9**, blue curves), with a settling time of about 500 min for the proportional controller (Fig. **5-9**a) and about 700 min for the PI controller (Fig. **5-9**b). In all simulations β_I was selected according to (5-32). However, when β_P cannot be tuned to match equation (5-29), the proportional controller alone fails to regulate y to the desired value, showing increasingly higher value of the steady-state errors as β_P decreases. Instead, as expected, adding the integral control action, the steady-state error is not sensitive to changes in the value of β_P, as long as (5-31)-(5-32) are satisfied.

Next, we tested the robustness of the control architectures with respect to imbalances in the relative composition of the consortium. Indeed, despite the host cells being identical, unavoidable asymmetries in the metabolic load on each population can cause their relative numbers to change over time. Towards this aim, neglecting cell growth and division, we varied the populations' relative numbers defined as $\rho_j = \frac{N_j}{N}$, $j \in \{p, i, t\}$, where N_j is the number of individuals belonging to the population j, while keeping constant the total number N of cells in the chamber, and evaluating the percentage error at steady state defined as:

$$e_{\%} = \left| \frac{Q_{x,\text{ss}}^t - Q_\text{d}}{Q_\text{d}} \right| \times 100\% = \left| \frac{e_\text{ss}}{\mu Y_\text{d}} \right| \times 100\%, \tag{5-33}$$

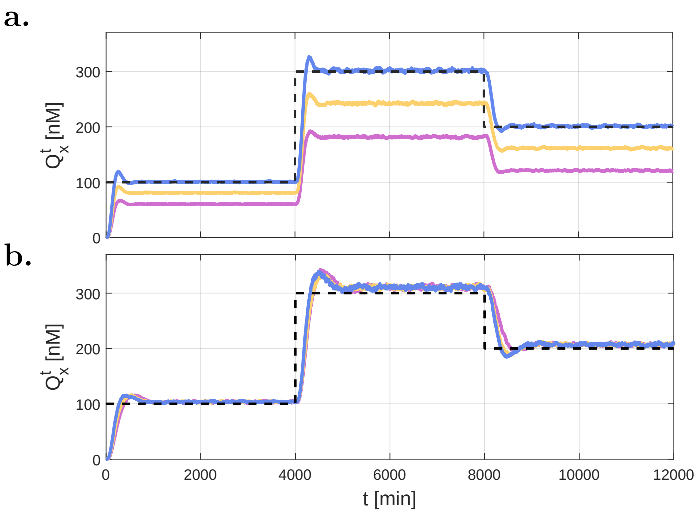


Figure 5-9: Set-point tracking experiments in BSim: evolution in time of the average concentration in the targets of the quorum sensing molecule Q_x^t when they are controlled by the proportional controllers only (panel a) and by a PI control action (panel b). The control gains were selected as $\beta_P = \{0.02, 0.03, 0.0414\}\,\mathrm{min}^{-1}$ (purple, yellow and blue line, respectively) and $\beta_I = 0.0002\,\mathrm{min}^{-1}$. The piece-wise constant reference signal $\mu Y_\mathrm{d}(t)/\theta$ is depicted as a dashed line. At $t = 0\,\mathrm{min}$, $N_p = N_t = 9$ or $N_p = N_i = N_t = 6$ cells depending on whether there is only proportional controller population or both controllers were positioned at the center of the chamber along a horizontal stripe. The initial state for each species was set to 0.

where e_ss is the control error (5-26) at steady state, $Q_{x,\mathrm{ss}}^t$ is the value of Q_x^t at steady state, and $Q_\mathrm{d} = \frac{\mu Y_\mathrm{d}}{\theta}$. Note that keeping N constant implies that $\rho_p + \rho_i + \rho_t = 1$.

Fig. **5-10** shows that the proportional controller alone works best when the populations are close to balance, while exhibiting increasingly higher errors at steady state as imbalance between controllers and targets increases, with a maximum error of 70% when the imbalance is extreme ($\rho_p < 0.1$). Adding an integral contribution to the control action significantly increases the architecture performance and robustness, with the relative error never exceeding 30% even when the imbalance of the controllers and target populations' numbers are consistent.

Finally, we tested the effects of cell-to-cell variability on the overall control performance. This heterogeneity in the response of the cells is mainly due to variations between plasmid copy numbers among different individuals in a population, caused by a possibly uneven distribution of the genetic material between cells after division. To model this effect, at each cell division, the value of all parameters of the daughter cells were drawn from a normal distribution centered at their nominal values μ with standard deviation $\sigma = CV \cdot \mu$, where

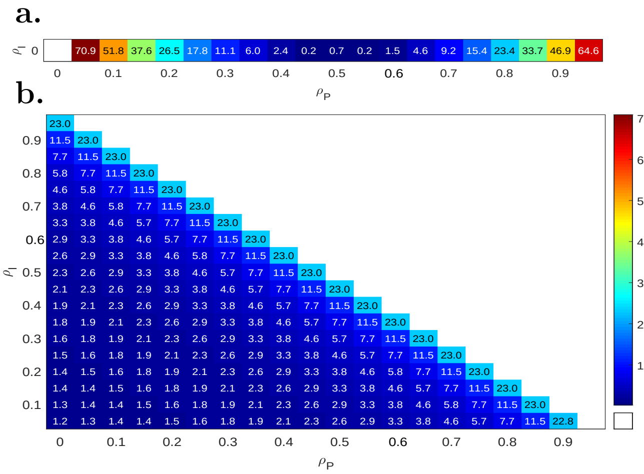


Figure 5-10: Robustness to imbalances in the consortium composition: percentage error at steady state (5-33) as the relative ratios of the three populations are changed. Cells movement, growth and division are neglected in these BSim simulations. The initial population was composed by $N = 20$, and the ratio ρ_j for $j \in \{p, i, t\}$ is constrained to $\rho_p + \rho_i + \rho_t = 1$. The reference signal is fixed to $Y_{\mathrm{d}} = 60\,\mathrm{nM}$, while the gains are chosen as $\beta_P = 0.0414\,\mathrm{min}^{-1}$ and $\beta_I = 0.0002\,\mathrm{min}^{-1}$.

CV is the coefficient of variation. The sensitivity of the control system was evaluated, as CV increases, by computing the relative error with respect to the unperturbed scenario, defined as:

$$\Delta e = \frac{1}{n} \sum_{k=1}^{n} \left| \frac{\bar{e}_k - \bar{e}_0}{\bar{e}_0} \right|, \tag{5-34}$$

where $\bar{e}_k$ and $\bar{e}_0$ are the control errors, averaged over the last 5000 min, of the k-th experiment and of a control experiment where $CV = 0$, respectively, and n is the total number of experiments. We observed that both the P and the PI control architectures guarantee small sensitivity to increasing levels of heterogeneity within the cellular populations with the relative error never exceeding 10% (Fig. 5-11). However, the multicellular PI control architecture shows higher robustness with the steady-state error showing much smaller variations under perturbation.

5.2.4 Discussion

We investigated analytically and numerically two multicellular architectures where a process of interest, hosted in a target cell population, was regulated using a P or a PI control law implemented across other populations in the consortium. We confirmed that robust regulation is possible by means of agent-based simulations. Our evidence show that the presence of an integral action is necessary for the regulation to a desired set point when it is

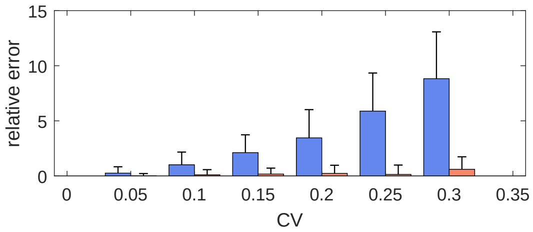


Figure 5-11: Robustness to cell-to-cell variability: relative variation of error at steady-state as the heterogeneity of the parameters increases, when the consortium is regulated by P controllers only (blue bars) and PI controllers (orange bars). For each value of $CV \in \{0.05, 0.1, 0.15, 0.2, 0.25, 0.3\}$ we performed $n = 100$ simulations drawing independently all cells' parameters from a normal distribution centered at their nominal value μ with standard deviation $\sigma = CV \cdot \mu$. The reference signal is fixed to $Y_{\mathrm{d}} = 60\,\mathrm{nM}$, while the gains are chosen as $\beta_P = 0.0414\,\mathrm{min}^{-1}$ and $\beta_I = 0.0002\,\mathrm{min}^{-1}$. All simulations were performed for a total time of $8000\,\mathrm{min}$, using a chamber of dimensions $5.7\,\mu\mathrm{m} \times 15\,\mu\mathrm{m} \times 1\,\mu\mathrm{m}$.

not possible to fine tune the parameters of the model. In addition, we showed that including an integral action improves robustness both with respect to the consortium composition and to parameter mismatches. However, the presence of a proportional action can be beneficial to reduce the settling time of the targets and improving the overall transient response, as shown in [9].

Chapter 6

Control of consortium composition

Regulating the relative numbers between members of different sub-populations within a microbial consortium is fundamental to ensure its correct operation. To this purpose, in this chapter we propose *external controllers* regulating the composition of a class of microbial consortia characterized by the presence of a bistable memory mechanism. However, this solution requires the modification of the growth environment to shape the consortium composition, which might be not feasible in some applications. Hence, we develop and present also an *embedded* controller for the growth rate of a bacterial population, which constitutes a first step towards the implementation of communities able to autonomously regulate their composition in response to external stimuli.

6.1 Ratiometric control of cell phenotypes in monostrain microbial consortia

In this section we consider a microbial consortium composed by *reversible differentiable cells* [134], that is, cells that belong to the same strain and embed a genetic mechanism allowing them to keep memory of past states and adapt their behaviour to external stimuli from the environment, for example by activating/deactivating specific set of genes. Specifically, we consider here the simplest case of cells that can switch between two states, mimicking a flip-flop or binary memory element (Figure **6-1**.a). The state of this bistable memory encodes the current role played by the cells in the consortium, and therefore the set of genes they are expressing at that moment. For example, a cell can use its resources either to produce some molecule or to grow and divide sustaining the cell population number [6] (Figure **6-1**.b, top panel). Also, in the case of a genetic pathway divided into two parts, a cell can switch from one state to the other so as to activate either two depending on the overall production levels in the consortium (Figure 6-1.b, bottom panel). We find that by endowing the cell population with a reversible bistable system, an external control strategy can be used to solve the ratiometric control problem. Specifically, by applying external stimuli to all the cells in the consortium it is possible to switch some cells from one state to the other so as to maintain the desired ratio. We show that this is possible in a number of different ways (namely, by using relay and PI controllers) and provide stability analysis of the resulting closed-loop system and exhaustive *in silico* validation of its performance and robustness. The validation is conducted by means of agent-based simulations in BSim [122, 123], a powerful platform for realistic *in silico* experiments in bacterial populations. As a representative example we consider the realistic agent-based implementation of the proposed ratiometric control strategy to enable cooperative bioproduction in microbial consortia, showing its effectiveness and flexibility when cell growth, cell-to-cell variability and other effects are appropriately modelled.

Although different approaches have been used to address the ratiometric control problem, their results hold under different sets of assumptions than we use here. For example, in [37] the authors considered a microbial consortium already comprising two different cell strains, adjusting the dilution rate in a chemostat to regulate the relative numbers of the populations in the consortium, while the platform developed in [51], although powerful, relies on delivering a different control input to each cell and, in addition, all controlled cells are cultured in spatially distinct environments. Moreover, in [135] a non-reversible, efficient differentiation control mechanism has been proposed for the creation and maintenance of cellular sub-populations in single strain microbial consortia, while a computer-controlled optogenetic platform for the regulation of the ratio of two-strain *E. coli* community has been recently presented in [136].

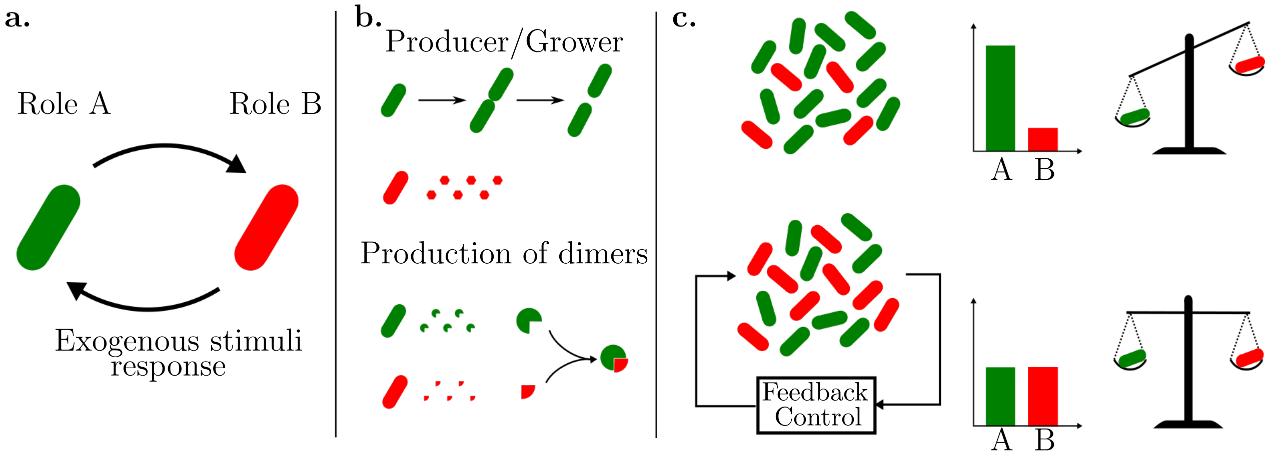


Figure 6-1: **Microbial consortia composed of reversible differentiable cells can be balanced in real-time by means of external feedback controllers to guarantee efficient labor division.** (**a**) Reversible differentiable cells can carry out different roles by activating/deactivating specific set of genes, depending on which state of the internal bistable memory is currently active. Cells can change role in response to exogenous stimuli from the environment, e.g., injection of inducer molecules or light. (**b**) Cells can, for example, either grow and duplicate or produce some desired molecule (red exagons), or they can produce two different molecules that react and produce the desired final bioproduct (green and red circular sectors). (**c**) Cells expressing different genes grow also at different rates, and thus their coexistence can be compromised. Feedback control algorithms can be employed to regulate in real-time the relative number of cells in the two groups, so that a balance in the population numbers and in the expression of desired genes is always guaranteed.

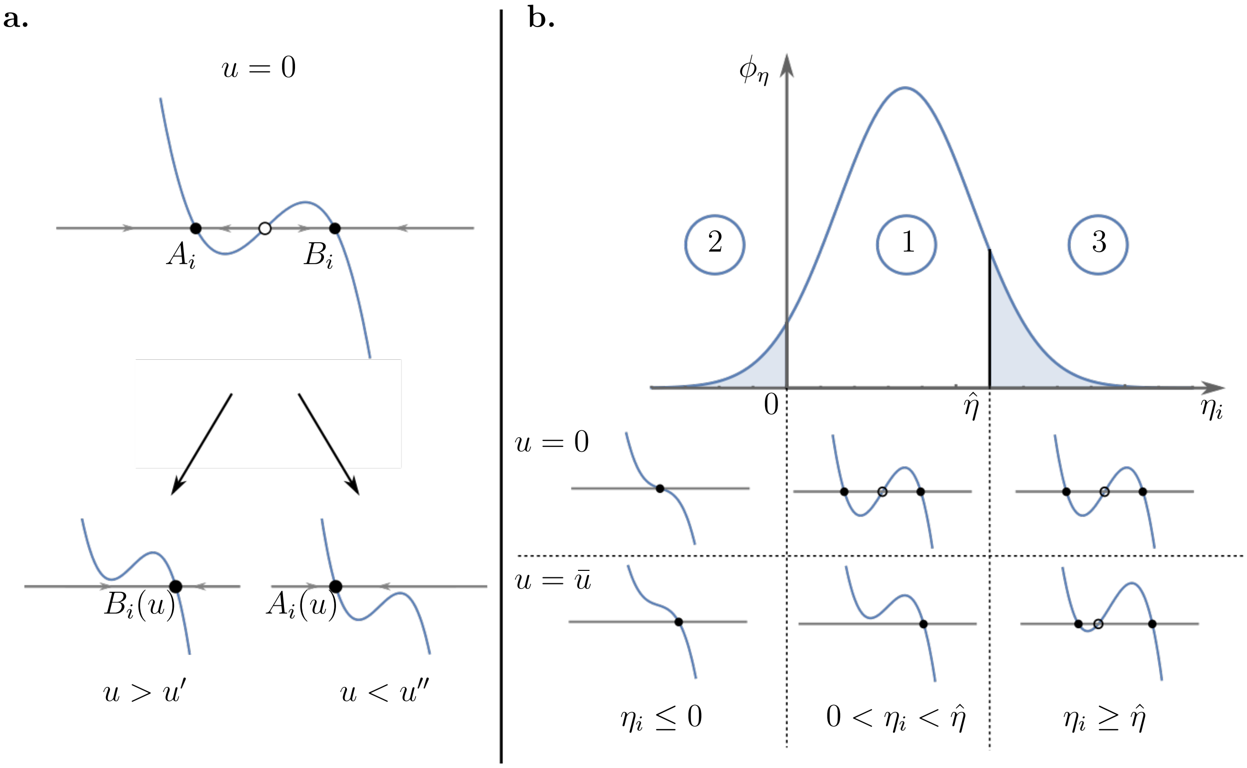


Figure 6-2: **Reversible cells can be switched from state A to B, and vice versa, by means of a common external input.** (**a**) The scalar dynamical system $\dot{x}_i = \eta_i x_i - x_i^3 + u$, with $\eta_i > 0$, has two stable equilibrium points for $u = 0$, namely A_i and B_i, each one corresponding to one of the two possible *roles* the cell can play in the consortium. A cell is *controllable* if by varying the input u in its interval of definition $\mathcal{U}$, it can be moved from one group to the other and vice versa. That is, there exists an admissible value $u' \in \mathcal{U}$ ($u'' \in \mathcal{U}$) such that there is a unique positive (negative) stable solution to the equation $\dot{x}_i = 0$ when $u > u'$ ($u < u''$). Full and empty dots represent stable and saddle equilibria, respectively. (**b**) Not all cells might respond as desired due to their heterogeneity, captured here by different values of the parameter η_i (assumed to be drawn from some probability distribution with density function ϕ_η, here sketched as Gaussian, just for the sake of illustration). Only cells whose value of the parameter η_i is between 0 and $\hat{\eta}$ are *controllable* (Case 1), that is, they have two stable equilibria for $u = 0$ and a unique stable equilibrium for $u = \pm\bar{u}$. Others cells can either be memory-less, or *monostable*, that is, they have only one equilibrium point for all values of u (Case 2), or can be *unswitchable*, having two stable equilibria for every $u \in \mathcal{U}$, and therefore cannot change role in the consortium (Case 3).

6.1.1 Reversible differentiable cells can be controlled to a desired state via a common exogenous input

The cells we consider here are assumed to embed some mechanism that can store the memory of past events. Memory is a fundamental mechanism in cell decision making. Furthermore, it is necessary in case we want to develop systems for living therapeutics [137]. In particular, we suppose that cells can be switched between two different states by appropriate external stimuli. The simplest model having the *memory-like property* (see Section A.3.1 for further details) can be described by the following ordinary differential equation:

$$\dot{x}_i = \eta_i x_i - x_i^3 + u, \tag{6-1}$$

where $u \in \mathcal{U} \subset \mathbb{R}$ is an input signal and is *common* to all cells.

In equation (6-1), $x_i \in \mathcal{X} \subset \mathbb{R}$ represents the macroscopic state, or role, of cell i and the value of parameter $\eta_i \in \mathbb{R}$ is assumed to be different for all cells, accounting for their heterogeneous responses to the common external input signal u. For positive values of η_i, the equation $\dot{x}_i = 0$ with $u = 0$ has two stable, non-trivial solutions, one negative and another positive, that we denote as A_i and B_i, respectively (Figure **6-2**.a). These solutions are the stable equilibrium points of the dynamical system described in (6-1) when no input is applied. We define as $\mathcal{R}_{A_i} = \{x_i : x_i < 0\}$ and $\mathcal{R}_{B_i} = \{x_i : x_i > 0\}$ the regions of attraction [138, Sec. 8.2] of A_i and B_i, respectively. Each cell will asymptotically converge to either A_i or B_i depending on which region of attraction its initial condition belongs to. Moreover, we denote by $\mathcal{N}_t$ the finite set of all cells in the consortium at time t and with $N(t) = |\mathcal{N}_t|$ its cardinality, that is, the number of cells currently under observation (e.g., via a fluorescence microscope). Note that this number may vary in time as a consequence of cell growth or death, or because of their removal (e.g., flush away) from the culture chamber in which they are hosted. We define the sets $\mathcal{A}_t := \{i \in \mathcal{N}_t : x_i(t) \in \mathcal{R}_{A_i}\}$ and $\mathcal{B}_t := \{i \in \mathcal{N}_t : x_i(t) \in \mathcal{R}_{B_i}\}$, such that $\mathcal{A}_t \cup \mathcal{B}_t = \mathcal{N}_t$ and $\mathcal{A}_t \cap \mathcal{B}_t = \varnothing$, and denote with $n_A(t)$ and $n_B(t)$ their cardinality. These two sets represent the group of cells in the consortium that, at time instant t, in the absence of any control input u are expected to asymptotically converge to A_i and B_i, respectively. Note that, as $\mathcal{A}_t$ and $\mathcal{B}_t$ form a partition of $\mathcal{N}_t$, at any time it holds that $n_A(t) + n_B(t) = N(t)$.

We model cell-to-cell variability by assuming that the parameter η_i in (6-1) is drawn randomly for each cell from the real interval $[\underline{\eta}, \overline{\eta}]$ with some probability distribution (Figure **6-2**.b). Also, we assume that the magnitude of the control input u is upper bounded by some maximum value $\bar{u} := \max_{u \in \mathcal{U}} |u|$. In the presence of such a bound on the control signal, only cells whose parameter value η_i is smaller than the threshold value $\hat{\eta}$ defined as

$$\hat{\eta} = \sqrt[3]{\frac{27\,\bar{u}^2}{4}}, \tag{6-2}$$

can be switched from one state to the other by an admissible value of u, that is, $u \in [-\bar{u}, \bar{u}]$. We define cells fulfilling the condition $\eta_i \in (0, \hat{\eta})$ as *controllable* cells. For details on the derivation of $\hat{\eta}$ see Section A.3.6.

All the cells not fulfilling this condition are instead uncontrollable because they are either *monostable* – their parameters differing so much from their nominal values that their bistable nature is lost – or *unswitchable* because their parameters exceed the threshold value (6-2). In the former case η_i will be taking non-positive values in our model, that is, $\eta_i \leq 0$, while in the latter case $\eta_i \geq \hat{\eta}$.

Therefore all cells characterised by parameter values $\eta_i \in (0, \hat{\eta})$, can be switched from one state to the other by means of a common bounded external input u applied into the environment. It is therefore possible to design, for such subset of controllable cells, some feedback control law to automatically regulate their state and keep the balance in the consortium between the two groups $\mathcal{A}_t$ and $\mathcal{B}_t$ to some desired level. As we are going to show next, uncontrollable cells will contribute to a small residual error that can be precisely estimated as a function of the upper bound $\bar{u}$ on the control input and therefore appropriately taken into account in the applications.

Note that even if model (6-1) is not a precise model of any existing bistable memory element, it captures its essential bistable nature and can therefore provide valuable information for the design of reliable and robust controllers able to solve the ratiometric control problem, as we show in the rest of this work.

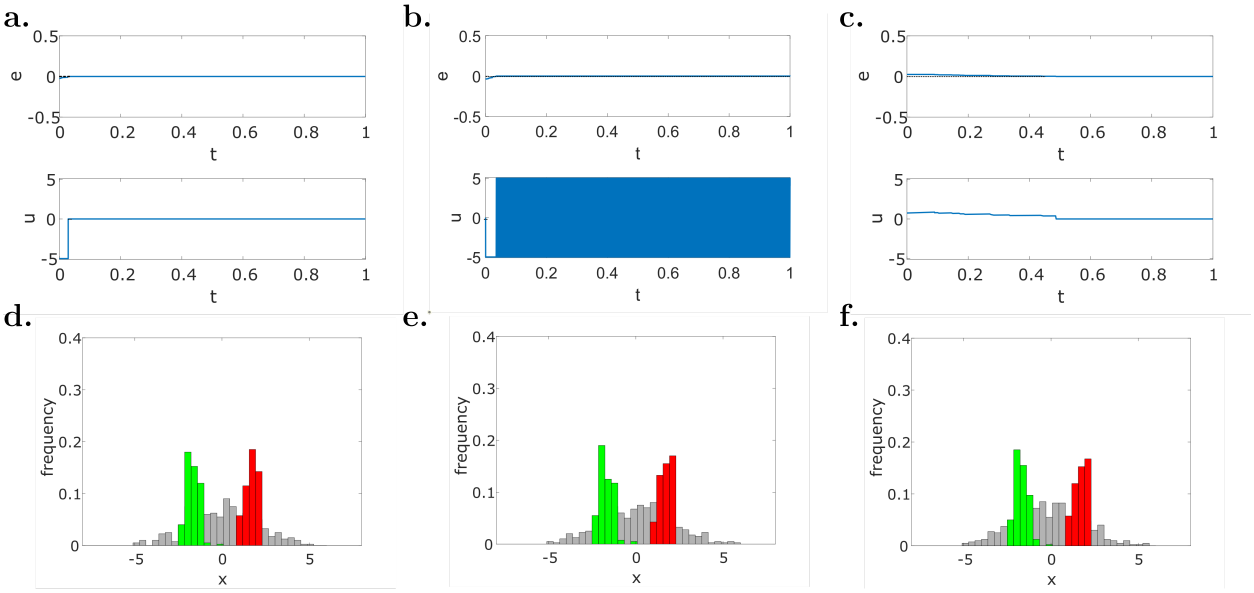


Figure 6-3: **Feedback control strategies are effective to balance two groups of controllable reversible cells to 1:1 ratio ($r_\mathrm{d} = 0.5$).** (**a-c**) Evolution of the error signal $e(t)$ and of the control input $u(t)$ for (**a**) the first implementation of the relay controller (A-6), (**b**) the second implementation of the relay controller (A-8), (**c**) the PI controller (A-10)-(A-11). (**d-f**) Distribution of the cells state at the beginning of the simulation ($t = 0$ a.u., grey histogram) and at steady state ($t = 1$ a.u., green and red histograms), for (**d**) the first implementation of the relay controller, (**e**) the second implementation of the relay controller, (**f**) the PI controller. The green and red bars in panels **(d-f)** correspond to cells being in the region of attraction of A_i and B_i, respectively. The maximum control input is set to $\bar{u} = 5$ and the gains of the PI controller are set to $k_\mathrm{P} = 30$ and $k_\mathrm{I} = 50$. All cells ($N = 400$) have initial conditions $x_i(0)$ drawn from the normal random distribution $\mathcal{N}(0, 4)$, and the parameters η_i are drawn with uniform distribution from the interval $[1, 5]$, therefore all cells are controllable, as no monostable ($\underline{\eta} > 0$) and no unswitchable ($\overline{\eta} < \hat{\eta} \approx 5.53$) cells are present in the population. (See also Figure **A-8** for more simulations with different desired ratios r_d.)

6.1.2 Ratiometric control of cell populations can be achieved by using external feedback strategies

The goal of ratiometric control is to regulate and maintain the *relative ratios* between the number of cells in $\mathcal{A}_t$ and the number of cells in $\mathcal{B}_t$, defined as

$$r_A(t) = \frac{n_A(t)}{N(t)}, \qquad r_B(t) = \frac{n_B(t)}{N(t)}. \tag{6-3}$$

As by definition, $r_A(t) + r_B(t) = 1$ for all time, it suffices to control either r_A or r_B to control the other. Without loss of generality, we assume the ratio $r(t)$ to be controlled is $r_B(t)$. Note that accurate measurement of each cell state is not needed. Indeed, if only noisy measurement of the state of the cell is available, as long as it is possible to identify the region of attraction the cell currently is in, the values of both r_A and r_B can still be accurately quantified.

More formally, the objective of the *ratiometric control problem* can be stated as follows.

Objective 1 *Given a consortium of reversible differentiable cells whose macroscopic dynamics can be described by* (6-1) *and a desired ratio* $r_\mathrm{d} \in [0, 1]$*, design a feedback control law* $u = u(t, x)$*, where* $x = [x_1, \ldots, x_{N(t)}]^\top$*, such that at steady state the consortium is divided into two cell groups whose ratio converges to some desired value,* r_d*, that is,*

$$r(t) \to r_\mathrm{d} \quad \textit{as} \quad t \to \infty. \tag{6-4}$$

The previous statement can be also reformulated in terms of the control error signal $e(t) := r_\mathrm{d} - r(t)$, by requiring that it goes to zero at steady state, that is, $e_\infty = 0$, where $e_\infty := \lim_{t\to\infty} e(t)$. The definition of the control error does not guarantee *per se* that the cells express the desired phenotype, as the cells whose state are near the unstable equilibrium will not be producing the target compound at the desired rate. However, belonging to the correct region of attraction guarantees that, after a transient time, the cells will converge to the correct equilibrium, where the desired phenotype is expressed.

Notice that it is possible to guarantee the solution of the ratiometric control problem as defined in (6-4) for *any* $r_\mathrm{d} \in [0, 1]$ only if *all* cells in the consortium are controllable, as described in Section 6.1.1. Indeed, if part of the population is not controllable, a residual steady-state error might still be present (see Section 6.1.3 for details and an analytical estimate of such residual error).

We present here two different feedback control strategies to solve the ratiometric control problem (see Figure **6-3**), an on-off relay controller and a proportional-integral (PI) controller. Both solutions are easy to implement, robust, and often used in other control applications of cell populations in microfluidic devices [139, 140, 125, 96] (see Section A.3.3).

Relay controllers (also known as bang-bang controllers) are simple yet effective feedback control laws that, by comparing the current output of the process of interest with its desired value, generate a piece-wise constant input signal $u_\mathrm{r}(t)$ whose value belongs to a discrete set $\mathcal{U}_\mathrm{r}$. Here, we propose the use of two alternative implementations of the relay controller, the former where the control input can also be set to zero, i.e., $\mathcal{U}_\mathrm{r}' = \{0, \bar{u}, -\bar{u}\}$, and the latter where u_r is always non-zero, i.e., $\mathcal{U}_\mathrm{r}' = \{\bar{u}, -\bar{u}\}$.

In the ideal case where all cells are controllable, the first implementation of the relay controller guarantees finite time convergence to zero of the error signal (Figure **6-3**.a). The second implementation instead can only guarantee bounded convergence of the error to zero since the input signal cannot be turned off once the error reaches zero. Hence, when such an implementation is adopted, the control input will continue to oscillate between its possible values (Figure **6-3**.b). As is common practice in applications where noise and uncertainties are unavoidable, a dead-zone or a delay can be added in the control loop to avoid high frequency oscillations of the control input that may cause excessive stress to cells and to the actuation system [18]. The details of the proof of convergence for the proposed relay controllers is reported in Section A.3.7.

An alternative strategy is the use of a PI controller that generates a control input $u_\mathrm{PI}(t)$ computed as the sum of one term proportional (P) to the error $e(t)$ and an other proportional to its integral (I) in time. In general, PI controllers guarantee zero regulation error at steady state in the presence of constant output disturbances [141]. In our implementation, this controller is complemented with a (anti-windup) reset condition that sets to zero the internal state of the integrator whenever the error signal $e(t)$ is equal to 0 or changes its sign (see Section A.3.3). When all cells are controllable, this strategy was also proved to solve the ratiometric control problem and guarantee convergence of the error to zero (Section A.3.7).

The evolution of the error signal $e(t)$ under the action of the PI controller is reported in Figure **6-3**.c. The error converges to zero as expected and the control input $u_\mathrm{PI}(t)$ also settles to zero after a short transient, similarly to what observed in the first implementation of the relay controller presented before.

Effective balancing of groups of reversible cells is also achieved by feedback control when the goal is to achieve groups of different sizes, that is, for r_d different from 0.5, e.g., equal to 0.75 or 0.25, corresponding to 1:3 and 3:1 ratios, respectively (Figure **A-8**). Notice that the similarity of the plots for the relay and PI control actions in Figure **A-8** are mainly due to the different set point chosen in these simulations. Indeed, since the initial control error is bigger with respect to the case presented in the main text, the time needed for the integral control action to saturate the control input is relatively low. When this happens, the control actions generated by the relay and the PI controllers are identical.

The performance of the controller are also compared with open loop simulations where no control is applied (Figure **A-14**). While using a closed loop control algorithm it is possible to regulate the relative numbers at some desired value, independently of the initial conditions, in the absence of any control action the error does not change over time and the final configuration of the consortium will strongly depend on the initial conditions chosen for the cellular population.

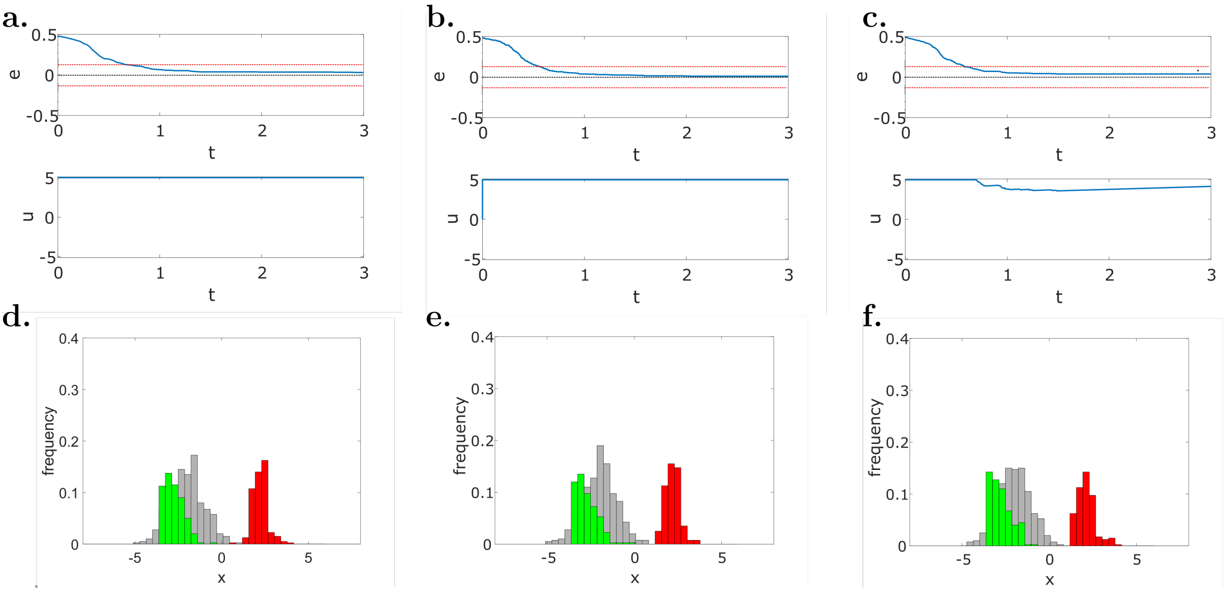


Figure 6-4: **Balance to 1:1 ratio is achieved with a residual steady-state error in the presence of uncontrollable cells ($r_\mathrm{d} = 0.5$).** (**a-c**) Evolution of the error signal $e(t)$ and of the control input $u(t)$ for (**a**) the first implementation of the relay controller (A-6), (**b**) the second implementation of the relay controller (A-8), (**c**) the PI controller (A-10)-(A-11). (**d-f**) Distribution of the cells' state at the beginning of the simulation ($t = 0$ a.u., grey histogram) and at steady state ($t = 3.0$ a.u., green and red histograms), for (**d**) the first implementation of the relay controller, (**e**) the second implementation of the relay controller, (**f**) the PI controller. The green and red bars in panels **d-f** correspond to cells being in the region of attraction of A_i and B_i, respectively. The maximum control input is set to $\bar{u} = 5$ and the gains of the PI controller are set to $k_\mathrm{P} = 30$ and $k_\mathrm{I} = 10$. All cells ($N = 400$) have initial conditions $x_i(0)$ drawn from the normal random distribution $\mathcal{N}(-2, 1)$, and the parameters η_i are drawn with uniform distribution from the interval $[-1, 14]$, therefore both monostable ($\underline{\eta} < 0$) and unswitchable ($\overline{\eta} > \hat{\eta} \approx 5.53$) cells can be present in the population. The steady-state errors observed in the *in silico* experiment are equal to **(a)** 0.05, **(b)** 0.0225, and **(c)** 0.0825. Note that all the observed errors are below the theoretical upper bound on the control error that can be estimated using (6-5) as $e_r = e_r^0 + e_r^u \approx 0.07 + 0.06 = 0.13$ (depicted in the panels **a-c** as red dashed lines). (See also Figure **A-9** for more simulations with different desired ratios r_d.)

6.1.3 Robust bounded regulation of the ratio is still possible in the presence of cell variability and physical constraints

When uncontrollable cells are present in the consortium, that is, cells that cannot be moved from one group to the other in response to any admissible inputs, the ratiometric control problem cannot be solved asymptotically, that is, we cannot guarantee that $e_\infty \to 0$ for an arbitrary initial configuration of the consortium. However, we can still guarantee that the absolute value of the steady-state error $|e_\infty|$ will be upper bounded by some positive quantity e_r, that is, $|e_\infty| \leq e_r$. This effect is well illustrated in Figure **6-4**, where it is shown that, regardless of the control algorithm being used, the error $e(t)$ approaches, but does not converge exactly to, zero. The error bound at steady state will depend upon the interplay between heterogeneity of the cells dynamics, and the constraints on the maximum input value of $\bar{u}$ that can be applied to the cells, as discussed in Section 6.1.1,.

The upper bound e_r can be estimated as being composed by two terms, that is,

$$e_r = e_r^0 + e_r^u, \tag{6-5}$$

each related to the probability of finding one of the two types of uncontrollable cells (i.e., monostable and unswitchable, respectively) in the consortium (Figure **6-2**). The first term, denoted as e_r^0, is related to the fraction of monostable cells, associated to a non-positive value of η_i, and so admitting only one stable equilibrium point for all values of u. For $N \to \infty$, where N is the number of cells in the population, assuming that the probability distribution from where the parameters η_i are drawn is known, we can estimate e_r^0 as

$$e_r^0 = \Pr[\eta_i \leq 0], \tag{6-6}$$

where Pr denotes the probability measure. The second term, denoted as e_r^u, is related to the fraction of unswitchable cells, that is, cells that are bistable but cannot be switched by any admissible value of the control input $u \in \mathcal{U}$. For a given upper bound input value $\bar{u}$, the fraction of unswitchable cells can be estimated, for $N \to \infty$, as the probability that the parameter η_i is greater than $\hat{\eta}$, that is, $\Pr[\eta_i > \hat{\eta}]$. Therefore, the residual error at steady state due to this second type of uncontrollable cells can be quantified as

$$e_r^u = \max\left\{0,\, \Pr[\eta_i > \hat{\eta}] - r_\mathrm{d},\, \Pr[\eta_i > \hat{\eta}] - (1 - r_\mathrm{d})\right\}. \tag{6-7}$$

Equation (6-7) is derived assuming that all the unswitchable cells are in the wrong region of attraction at the beginning of the simulation, which represents the worst case scenario. Specifically, if either $\Pr[\eta_i > \hat{\eta}] - r_\mathrm{d} > 0$ or $\Pr[\eta_i > \hat{\eta}] - (1 - r_\mathrm{d}) > 0$ it means that the amount of cells that need to be switched to achieve the control goal is greater than the number of switchable cells in the consortium, meaning that there may be some steady state error. Note that this value depends on the relationship between the desired ratio for that group (either r_d or $1-r_\mathrm{d}$) and the number of unswitchable cells therein, because they affect the error only when this number exceeds the desired value (see Section A.3.8 for details).

Similar results are observed also when the desired goal is to split the cell population into groups of different sizes, e.g., with 3:1 or 1:3 ratios (Figure **A-9**), confirming that this undesired effect is not due to the particular control strategy adopted or to the chosen desired ratio.

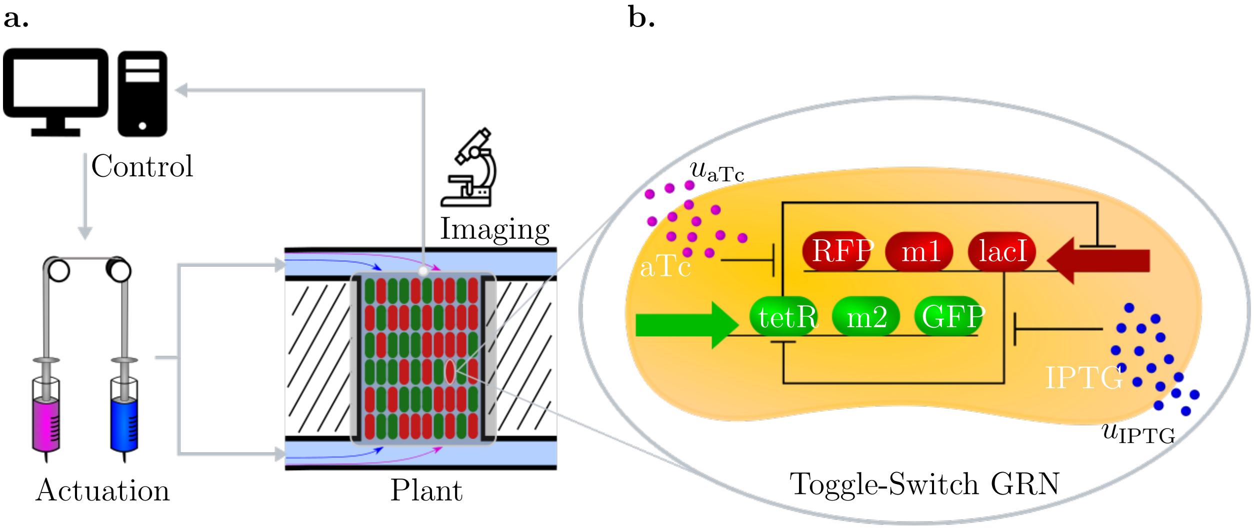


Figure 6-5: **Controlled cooperative bioproduction of a dimer in microfluidic devices.** (**a**) Reversible differentiable cells are hosted in microfluidic chambers where they grow and produce a specific molecule related to the corresponding active state of their internal memory. The current ratio of the two cell groups in the chamber, and hence the production level of the corresponding monomers, is evaluated by measuring with a fluorescence microscope the expression of reporter proteins in each cell. This information is acquired by the feedback control algorithm that compares the current ratio $r(t)$ to the desired ratio r_{d} and computes online the corresponding control inputs. Finally, these signals are sent to the cells by actuating a pair of syringes connected to the microfluidic chambers and containing mixtures of growth medium and inducer molecules. (**b**) The required reversible bistable memory mechanism is implemented by using an inducible toggle-switch. Depending on which of the two repressor proteins, either LacI or TetR, is currently expressed, the cell produces the corresponding monomer and the reporter protein (either M1 and RFP or M2 and GFP, respectively). The state of the toggle-switch can be flipped by changing the concentration of the inducer molecules aTc and IPTG in the microfluidic chamber (denoted as u_{aTc} and u_{IPTG}), which diffuse through the cell membrane and bind to TetR and LacI, respectively.

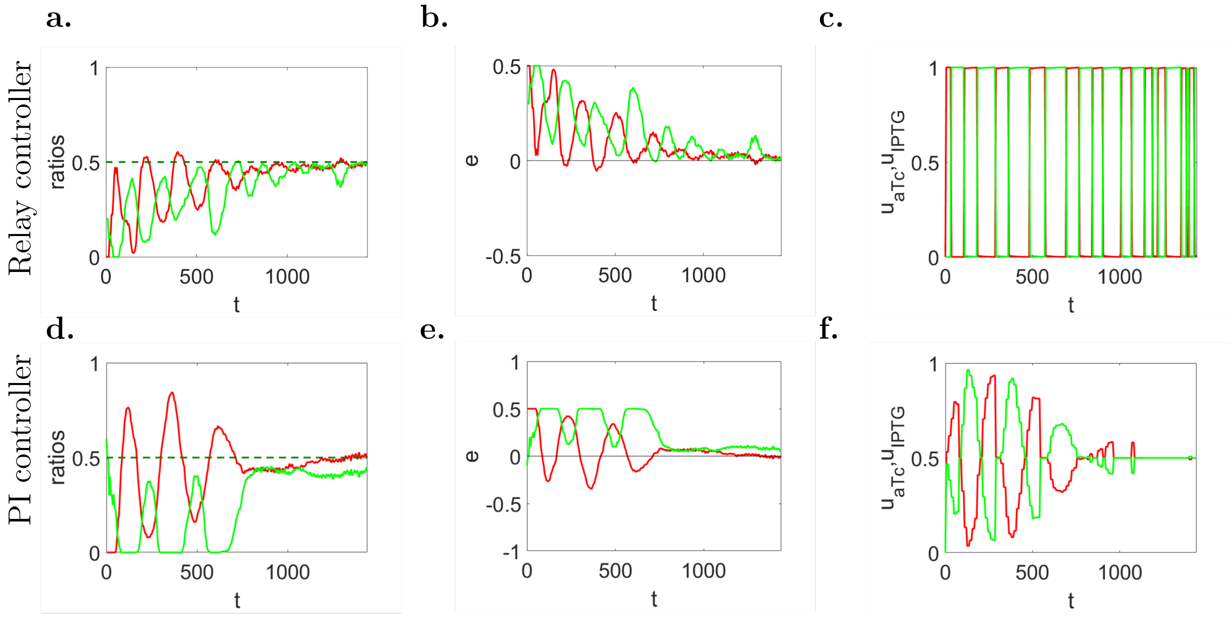


Figure 6-6: **Cooperative production of two monomers to 1:1 population ratio can be achieved by means of feedback ratiometric controllers in microfluidics.** (**a,d**) Evolution in time of populations' ratio r_A (solid green line) and r_B (solid red line) with their respective desired reference values in dashed lines ($r_\mathrm{d} = 0.5$), (**b,e**) of the error signals e_A (solid green line) and e_B (solid red line), and (**c,f**) inducer control signals u_aTc (solid red line) and u_IPTG (solid green line), normalised to their maximum values U_aTc and U_IPTG, respectively. (**a-c**) Parameters of the relay control (A-41): $U_\mathrm{aTc} = 60\,\mathrm{ng/mL}$, $U_\mathrm{IPTG} = 0.5\,\mathrm{mM}$. (**d-f**) Parameters of the PI controller (A-43): $U_\mathrm{aTc} = 100\,\mathrm{ng/mL}$, $U_\mathrm{IPTG} = 1\,\mathrm{mM}$, $k_{\mathrm{P},1} = 100$, $k_{\mathrm{P},2} = 1.5$, $k_{\mathrm{I},1} = 1.5$, $k_{\mathrm{I},2} = 0.05$. Cells (about 200) in the simulated microfluidic chamber (with dimensions $40\,\mu\mathrm{m} \times 50\,\mu\mathrm{m} \times 1\,\mu\mathrm{m}$) have the same parameters' value, and their evolution has been obtained using the agent-based simulator BSim [122, 123] (See Section A.3.11 for further details on the simulator setup). (See also Figure **A-10** for more *in silico* experiments with different desired ratios r_d.)

6.1.4 Ratiometric control enables cooperative bioproduction in microbial consortia

So far the analysis has been conducted by considering the scalar model (6-1) capturing the macroscopic bistable nature of the cells considered in this paper. As we are going to show by means of the representative application that follows, the behaviour captured by the reduced model in (6-1) is qualitatively preserved also in more complex and realistic cell models exhibiting the required memory-like property. Therefore, albeit simple, we demonstrate that the model in (6-1) can be effectively used to design feedback control laws to solve the ratiometric control problem in realistic applications.

As a representative case of study, we consider the agent-based *in silico* implementation of ratiometric control for the bioproduction of protein dimers in microfluidic devices (Figure **6-5**). In this scenario, according to its state (A or B), each cell in the consortium produces either one of two monomers. By acting on the available control inputs, we want to regulate the relative number of cells producing the two monomers so as to balance the overall production of the resulting dimer.

We assume that the mechanism required by the *E. coli* cells to guarantee their correct coordinated behaviour is implemented by means of an inducible genetic toggle-switch [124]. Specifically, we consider the circuits design presented in [125] and further analysed in [126, 127, 142, 143]. This genetic regulatory network consists of two repressor proteins, LacI and TetR, both repressing each other's promoter, so that only one protein is fully expressed at any time. The expression level of the two repressor proteins can be flipped by changing the concentration of two inducer molecules, aTc and IPTG. The former input, aTc, binds to TetR, increasing the rate of production of LacI, and therefore causing the cell to converge to the steady-state corresponding to high expression of LacI. Analogously, IPTG binds to LacI, causing the commitment of the cell to a steady-state corresponding to high expression of TetR.

The sixth order dynamical model of each cell is described in details in Section A.3.10, in which the variables u_{aTc} and u_{IPTG} (as reported in Figure **6-5**) denote the concentrations of the inducer molecules in the growth medium of the microfluidic chambers and they represent the control inputs that can be applied to all cells to change their production role in the consortium

We further assume that the genes m1 and m2 encoding the two monomers of interest are each transcribed together with the repressor genes lacI and tetR of the toggle-switch circuit. So that, at steady state, each cell fully produces only one monomer at the time and at a rate assumed to be proportional to the concentration of the corresponding repressor protein.

Reporter genes of red and green fluorescent proteins (RFP and GFP) are also bound to the repressor genes to monitor the current level of production of the monomers by using fluorescence microscopy (Figure **6-5**). Finally, we assume that the two monomers have equal transcription and translation rates. Therefore, for the dimer to be produced at high rate, the consortium must be split and maintained into two symmetric groups with a 1:1 ratio, that is, we set $r_{\mathrm{d}} = 0.5$. Note that this assumption does not hinder the generality of the framework presented, as different transcription and translation rates would simply require the consortium to operate around a different setpoint, that we can reach and stabilise by just modifying r_{d}.

In the *in silico* experiments, we also take into account realistic physical and technological constraints of a possible implementation in the microfluidic experimental platform described in [139, 12]. The choice of such platform derives from its extensive use in the context of external control [18, 51, 96]. Specifically, we consider constraints on (i) the possible classes of input signals that can be generated by the actuators, (ii) an upper bound on the switching frequency of the inputs to limit osmotic stress to the cells, (iii) a time delay accounting for the time the chemical inducers take to flow from the reservoirs to the cell chambers, and (iv) a safety lower bound on the sampling time of the measurements to avoid excessive photo-toxicity (see Section A.3.11).

In silico control experiments have been conducted by using *ad hoc* implementations in BSim [122, 123] of the two feedback control algorithms presented in Section 6.1.2 (see Section A.3.12). To test the relay control strategy, we assumed that the actuation of the inputs is realised using an ordinary T-junction [133], which allows only one inducer species at a time to be injected into the microfluidic chambers. For the PI controller we assumed that the actuation is realised by a Dial-A-Wave system, as described in [144]. This actuation system is more advanced than the previous one as it allows mixtures of the two inducers to be injected in different proportions into the chambers. Notice that, albeit two inducers are needed to manoeuvre the state of the toggle-switch, they are constrained by our control strategy either to be mutually exclusive (for the relay controller) or in a convex combination (for the PI controller), and therefore they can be viewed as being a single input (see Section A.3.12). Both feedback control algorithms take into account the characteristics of the experimental platform and in particular of the actuators. Full details about the control algorithms and the technological constraints of the platform are reported in Section A.3.12.

Note that the constraints imposed to the control inputs by the platform chosen imply that it would be impossible for our control algorithm to provide no control action at any given time instant. Therefore, the implementation in this context of a relay control algorithm complemented with a null control action was not considered.

The agent-based simulations in BSim accurately capture cells' reproduction, spatial distribution and geometry of the cells and of the microfluidic chambers, diffusion of chemicals into the environment and, more importantly, flush-out of the cells from the chambers. Further details on the stochastic simulation algorithm, geometry and other parameters used for *in silico* experiments in BSim are reported in Section A.3.11.

We observed that both controllers can successfully regulate, after relatively short transients, the populations' ratio to the desired value (Figure **6-6**), which otherwise would have converged to some value that strongly depends on the initial conditions of the cells (Figure **A-15**) The relay controller shows a faster response with more severe oscillations (Figure **6-6**.a-c), while the PI controller presents a smoother but slower response with higher accuracy at steady state (Figure **6-6**.d-f). This is expected as it is well-known that the relay control strategy is in general more robust to uncertainties and noise affecting the controlled process but has poorer accuracy at steady state; the PI control strategy showing better steady-state performance thanks to the presence of an integral action.

The difference in the performance of the two strategies is also seen by the different actuation systems employed in our experiments. Indeed, the Dial-A-Wave system allows for a finer regulation of the concentrations of the inducer molecules than the simpler (and cheaper) T-junction, allowing better accuracy of the control system at steady state. Similar performances are obtained also when the goal is changed to achieve different population ratios, e.g., a 1:3 ratio or a 3:1 ratio (Figure **A-10**). These scenarios may correspond, for example, to situations in which the two monomers have equal transcription rates but different translation rates, requiring the consortium to be split into two asymmetric groups, for efficient production of the dimer.

Besides biochemical noise, the fluctuations at steady state (Figure **6-6**.a,d) are essentially due to cells being flushed out of the microfluidic chamber as they grow and duplicate, and they do not depend on the control parameters used, which might only affect the transient response of the cell populations. These fluctuations are more relevant when cells are host in a small chamber and becomes less significant as the size of the growth chambers increases. Indeed, for the sake of simplicity, assuming the chamber to be a square of side ℓ, the magnitude ε of the fluctuations is proportional to $1/\ell$ (see Section A.3.9), hence the fluctuations increase as the chamber size decreases (see Figure **A-11**), and vice versa.

In addition to this, we assessed the robustness of the control algorithms to cell-to-cell variability, modelled as variability between cells parameters. In detail, the numerical simulations confirmed that the ratiometric controllers presented here were also able to regulate the relative numbers in this case, as shown in Figure **A-12**. The coefficient of variation was selected to be CV $= 0.2$, similarly as done in previous works in the literature [4].

Another important factor determining the evolution of the relative numbers between sub-populations in microbial communities is their generally different growth rates (assumed so far to be identical for both phenotypes) due to different metabolic load. To test the robustness of the designed algorithms to this issue, we assumed that over-expression of the LacI pathway caused a reduction (of 50%) of the growth rate. We found out that even when significant discrepancies in the growth rates are present, no divergence phenomena was observed with the cells still splitting into two sub-populations with a steady-state error that never exceeds 0.15 (see Figure **A-13**).

6.1.5 Discussion

We presented a general framework to guide the design of external feedback controllers for phenotype regulation in microbial consortia. We showed that, by exploiting the memory-like property of reversible differentiable cells, a single-strain cell population can be divided into two groups, expressing different sets of genes, whose relative numbers, i.e., the ratio, can be regulated by means of common exogenous inputs. Notice that, although the proposed controllers can effectively regulate the ratio of the two subpopulations, it is not possible to regulate the rate of convergence at which cells express the desired level of the phenotype of interest, as this is determined by the inherent dynamics of the cells. Also, we wish to highlight that ratiometric control of a population of reversible cells by means of a common input signal is only made possible by the heterogeneity of their response to that input. Indeed, heterogeneous reversible cells characterised by different parameter values switch at different time instants when subject to the same input, and this is a crucial property that allows their state to be controlled by an external feedback action. Stochastic effects, such as biochemical noise or delays, can indeed facilitate the stabilisation of the reversible cells into different groups, by amplifying the cell-to-cell variability, as it was demonstrated in the *in silico* experiments we provided. Future developments might include a more in-depth analytical investigation of their effect on the design and performance of feedback control strategies designed to solve the *ratiometric control* problem.

6.2 Embedded control of cell growth using tunable genetic systems

As we stressed in the introduction to this chapter, channelling cell resources from biomass production (i.e., cell growth and division) to protein production while preventing accumulation of toxic by-products is key for the development of complex engineered genetic circuits. [145, 146, 147, 148, 149] Specifically, the cell growth rate needs to be carefully regulated in order to make the biomass density reach and maintain a desired reference value so as to guarantee higher efficiency of the bioproduction process. This objective is generally achieved by means of external systems controlling the density of the cell populations inside the bioreactors.[150, 151, 37, 152, 153, 100] However, in scenarios where it is not possible to modify the culture conditions either for cost or technological limitations, designing cellular populations capable of self-regulating their growth rate by means of embedded genetic feedback mechanisms could be more suitable, providing intrinsic robustness to the bio production process, e.g., preventing extinction or starvation.

A pioneering solution for the self-regulation of population density in single strain microbial populations was developed in Ref [154] where a quorum sensing based mechanism is used to self-limit the number of the cells. Alternative approaches involve engineering multiple microbial populations to achieve self-regulation of their relative numbers [47, 48, 46, 155, 156, 45] and in some cases also regulation of both the consortium size and composition.[44, 107] In all previous work, the desired population densities are hard-wired into the design of the synthetic gene regulatory networks. Hence, achieving a different working condition requires re-engineering the entire consortium.

In this section, inspired by Ref [154], we present a genetic feedback strategy for population control. Specifically, by means of a quorum sensing mechanism, which allows cells to sense the population density, we render cells capable of self-regulating their own growth rate by producing a growth inhibitor protein. The genetic controller governing this process inside the cells is implemented by means of a *tunable expression system* (TES), recently proposed in Refs [157, 158]. By exploiting this circuit, the production rate of the inhibitor protein can be dynamically changed depending on the concentration of the quorum sensing molecule present into the environment. In so doing, a feedback mechanism can be embedded into the cells so that they become able to regulate their own relative numbers. Thanks to its characteristic features, the TES provides additional flexibility to the control system by allowing the desired set-point for the population density to be changed online according to

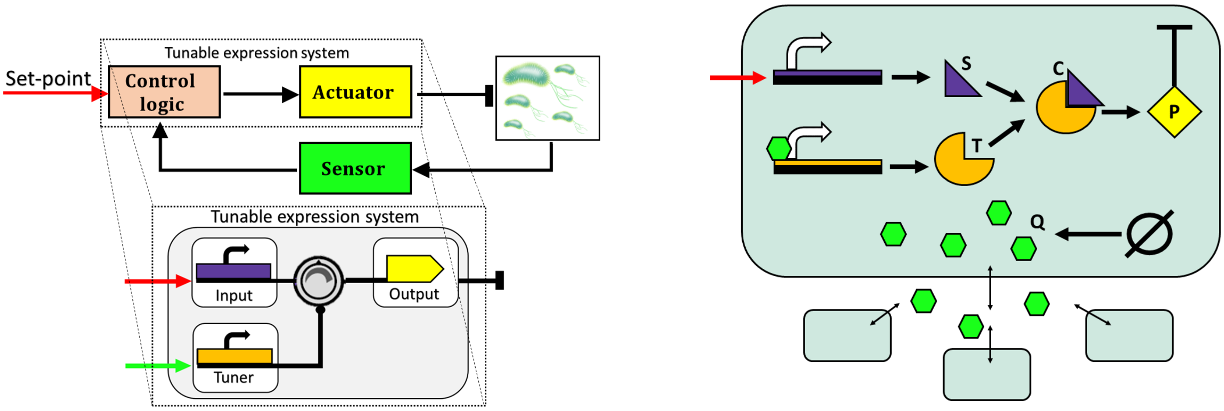


Figure 6-7: Genetic feedback control of the cell population density. *Left panel:* Block diagram representation of the population control strategy. The cell population can self-regulate its growth rate by sensing its current density in the culture environment and comparing it to a desired set-point. This comparison is realized by means of a tunable expression system (TES), which allows its input-output relationship to be dynamically changed as a function of the tuner signal, so that its output can modulate the production of a growth inhibitor protein, closing the feedback loop. *Right panel:* Abstract biological implementation scheme of the population control system. Each cell produces a quorum sensing molecule Q (green hexagons) that diffuses into the environment and in the other cells. The molecule Q, proportional to the population density, regulates the production of the species T, which, reacting with the species S, acts as the tuner signal for the production of the complex C. The production of the growth inhibitor protein P is activated by the complex C, thus closing the loop. Notice that the rate of production of the species S can be either tuned offline or regulated online to specify the desired set-point for the population density N_d.

some exogenous input, and to compensate for unavoidable inaccuracies in the offline tuning of the system parameters, and uncertainties due to cell-to-cell variability.

After presenting the details of the proposed genetic controller, we analyze its steady-state response by considering a simplified model obtained by averaging the dynamics over the entire population assumed to grow in a culture tube. This allows us to analytically derive an approximate relationship between some tunable parameter and the steady-state population density that we then exploit for control design. Finally, the performance and the robustness of the proposed control system are validated *in silico* by means of realistic agent-based simulations performed in BSim, [122, 123] a platform to simulate microbial populations which was originally designed for microfluidic experiments only and is adapted here to simulate cells growing in test tube cultures.

6.2.1 Population Control

The control strategy we present here consists of closing a feedback loop between the cell population and its environment allowing the population to self-regulate its own density by producing a growth inhibitor protein (Figure **6-7**).

This is realized by comparing the concentration of a quorum sensing molecule, produced by the cells and diffusing into the environment, to some desired set-point value of the population density. The concentration of such diffusing molecule into the environment is proportional to the density of the population, therefore, regulating its concentration, directly affects the number of cells into the environment.

The production of the growth inhibitor protein is activated by a genetic controller implemented by means of a tunable expression system (TES) recently proposed in Ref [157] (Figure **6-7**, left panel). Specifically, this genetic circuit allows the input-output relationship between the set-point value and the production rate of the inhibitor protein to be dynamically changed in response to the concentration of the quorum sensing molecule. In this way, the quorum sensing molecule acts as a *tuner* input affecting the production rate of the inhibitor protein, closing the loop. Indeed, when the concentration of the quorum sensing molecule into the environment increases (corresponding to a growth in population density), the production rate of the inhibitor protein also increases and thus the population density decreases. Vice versa, when the population density decreases, the production of the inhibitor protein also decreases and the growth rate of the population increases again (Figure **6-7**, right panel).

Notice that the desired set-point value for the population density can be set either online by means of external inputs, e.g., by changing the concentration of some inducer molecule in the growth medium [159] or by applying light stimuli via optogenetics, [160, 161] or offline by adequately tuning the parameters of the genetic controller.

Modeling We assume that the cells are cultured in an environment of volume V with limited availability of nutrients, such as a test tube. We label with i, $i = 1, \ldots, M(t)$, each cell in the population, where $M(t) : \mathbb{R}_{\geq 0} \mapsto \mathbb{N}$ is the total number of cells in the population at time t. The dynamics of the genetic control system embedded in each cell i and depicted in Figure **6-7**, right panel, is described by a deterministic model we derived from the laws of mass-action and Michaelis-Menten kinetics as the following set of ordinary differential equations:

$$\frac{dS_i}{dt} = \alpha_0^S + \alpha^S(t) - \kappa_{C+} T_i S_i + \kappa_{C-} C_i - \gamma S_i \tag{6-8}$$

$$\frac{dT_i}{dt} = \alpha_0^T + \alpha^T \frac{Q_i^n}{Q_i^n + \theta^n} - \kappa_{C+} T_i S_i + \kappa_{C-} C_i - \gamma T_i \tag{6-9}$$

$$\frac{dC_i}{dt} = \kappa_{C+} T_i S_i - \kappa_{C-} C_i - \gamma C_i \tag{6-10}$$

$$\frac{dP_i}{dt} = \kappa_P C_i - \gamma P_i \tag{6-11}$$

$$\frac{dQ_i}{dt} = \frac{\rho_0}{V_c} + \frac{\eta}{V_c} (Q_e - Q_i) - \gamma Q_i \tag{6-12}$$

where the state variables S_i, T_i, C_i, P_i denote, respectively, the concentrations inside cell i of the toehold switch transcript S, the tuner sRNA T, the switch-sRNA complex C of the TES, [157] and of the inhibitor protein P. Moreover, Q_i and Q_e denote the concentrations of the quorum sensing molecule Q inside each cell i and into the external environment, respectively. That is, $Q_i := q_i/V_c$ and $Q_e := q_e/V_e$, where q_i and q_e denote the amount of molecules into each cell with volume V_c and into the environment with volume $V_e = V - M\,V_c$, respectively. For the sake of simplicity, in the above equations we assumed that all cells are identical and all molecular species degrade, mostly because of dilution, following first-order kinetics with rate γ. This simplifying assumption is commonly made in the literature to derive analytical results that can be used to guide the following implementation of the real biological circuits. [162, 142, 89]

The dynamics of the TES is described by equations (6-8)-(6-10). In (6-8), the toehold switch S is transcribed with rate $\alpha_0^S + \alpha^S(t)$, where the value of the function $\alpha^S(t)$ encodes the desired set-point value for the population density, which will be assumed constant in the rest of the paper, that is, $\alpha^S(t) = \alpha^S$, for all $t > 0$; this corresponds to solving a regulation problem, typical in industrial applications in which the biomass density must be regulated to some value such that the optimal production rate of the molecules of interest is reached. In (6-9), the tuner sRNA T is transcribed with rate $\alpha_0^T + \alpha^T \frac{Q_i^n}{Q_i^n+\theta^n}$, where θ is the activation coefficient and n is the Hill coefficient, and finally in (6-10) the switch-sRNA complex C binds and unbinds with rates κ_{C+} and κ_{C-}, respectively. The inhibitor protein P, with dynamics given in (6-11), is the output of the TES circuit and is produced at rate $\kappa_P C_i$. Finally, the quorum sensing molecule Q, whose dynamics is described in (6-12), is produced by each cell with a constant rate ρ_0, regulating the activation of the transcription of the tuner sRNA T, and diffuses through the cell membrane with diffusion rate η, so that the evolution of the quorum sensing molecules into the environment is described by

$$\frac{dQ_e}{dt} = \sum_{i=1}^{M} \frac{\eta}{V_e} (Q_i - Q_e) - \gamma_e Q_e \tag{6-13}$$

where γ_e is the external degradation rate.

In the following sections, we first present a simpler average model of the closed-loop system allowing us to analyze the effects that a change of the reference parameter α^S has on the steady-state value of the population density. Then, we describe the results of an exhaustive set of realistic *in silico* experiments performed by means of the agent-based bacterial simulator BSim [123, 122] to validate the effectiveness of the proposed approach.

6.2.2 Steady-State Analysis

Model (6-8)-(6-13) can be recast to describe the average behavior of a populations of cells, each embedding the genetic control system described in Section 6.2.1. Specifically, denote with $N(t) : \mathbb{R}_{\geq 0} \mapsto \mathbb{R}_{\geq 0}$ the density of the cell population growing in the volume V at time t, that is, $N(t) := M(t)/V$. The average model of the cell population dynamics can be derived as (see Appendix A.4.1 for further details):

$$\frac{dS}{dt} = \alpha_0^S + \alpha^S - \kappa_{C+} TS + \kappa_{C-} C - \gamma S \tag{6-14}$$

$$\frac{dT}{dt} = \alpha_0^T + \alpha^T \frac{Q^n}{Q^n + \theta^n} - \kappa_{C+} TS + \kappa_{C-} C - \gamma T \tag{6-15}$$

$$\frac{dC}{dt} = \kappa_{C+} TS - \kappa_{C-} C - \gamma C \tag{6-16}$$

$$\frac{dP}{dt} = \kappa_P C - \gamma P \tag{6-17}$$

$$\frac{dQ}{dt} = \rho_0 N - \gamma_e Q \tag{6-18}$$

$$\frac{dN}{dt} = kN \left(1 - \frac{N}{N_m}\right) - dPN \tag{6-19}$$

where the state variables S, T, C, P denote, respectively, the average concentrations of molecules in each cell of the toehold switch transcript S, the tuner sRNA T, the switch-sRNA complex C of the TES, and of the inhibitor protein P, while Q denotes the concentration of quorum sensing molecule Q in the volume V, that is, $Q := q/V$, with q being the total amount of Q in the culture volume. The notation and the meaning of the coefficients in (6-14)-(6-18) are the same as in (6-8)-(6-12). Moreover, in equation (6-19), k denotes the intrinsic growth rate of the population, N_m the carrying capacity, i.e., the maximum density the population can reach given the nutrients available, and d the growth inhibition rate due to the protein P.

To derive a simple, yet meaningful analytical result, in what follows we substitute the Hill function $\frac{Q^n}{Q^n+\theta^n}$ in (6-15) with its piece-wise linear approximation; this approach is widely employed in the literature to study the dynamics of genetic regulatory networks.[163, 164, 165] Specifically, assuming that the Hill function in (6-15) is not saturated, the activation function of T due to Q can be approximated by its first-order approximation evaluated at $Q=\theta$, that is, $\frac{Q^n}{Q^n+\theta^n} \approx \frac{(2-n)}{4} + \frac{n}{4\theta}Q$, which reduces to $\frac{Q}{2\theta}$ for $n=2$. In this way, the average model (6-14)-(6-19) allows us to obtain a very simple relationship at steady state between the value of the population density, $N_{\mathrm{ss}} := \lim_{t\to\infty} N(t)$, and the reference parameter α^S, which is described by the following proposition.

Theorem 1 *Under the assumption that the basal synthesis rates of the species S and T are negligible, that is, $\alpha_0^S = \alpha_0^T = 0$, that $n=2$, and that*

$$\alpha^S \ll \frac{\gamma^2 k}{\kappa_P d} - \gamma \frac{\gamma + \kappa_{C-}}{\kappa_{C+}}, \tag{6-20}$$

$$\frac{\gamma + \kappa_{C-}}{\kappa_{C+}} < \frac{\gamma k}{\kappa_P d}, \tag{6-21}$$

$$\alpha^T \gg \frac{6\gamma_e \theta}{\rho_0 N_m} \frac{\gamma^2 k}{\kappa_P d}, \tag{6-22}$$

then the value of the population density N_{ss} at steady state, normalized by the carrying capacity N_m, can be approximated by

$$\frac{N_{\mathrm{ss}}}{N_m} \approx -\frac{\kappa_P d}{\gamma^2 k}\alpha^S + 1. \tag{6-23}$$

Proof 1 *The steady-state value N_{ss} can be obtained more easily by considering the nondimensional model of* (6-14)-(6-19) *(see Appendix A.4.2 for the complete details), setting the time derivatives to zero and solving the resulting algebraic equations. Such equations have only one admissible solution, the others being either the trivial solution or negative. The unique positive solution can be computed to be:*

$$\frac{N_{\mathrm{ss}}}{N_m} = -\frac{(-a_T a_S + k_C - 1)}{2a_T} + \sqrt{\left(\frac{-a_T a_S + k_C - 1}{2a_T}\right)^2 + \left(\frac{-a_S + k_C}{a_T}\right)}, \tag{6-24}$$

where

$$a_S = 1 - \bar{\alpha}^S = 1 - \frac{\alpha^S}{\gamma S_0} \tag{6-25}$$

$$a_T = 1 + \bar{\gamma}\bar{\alpha}^T = 1 + \frac{Q_0}{2\gamma_e \theta S_0}\alpha^T \tag{6-26}$$

$$k_C = \frac{1 + \bar{\kappa}_{C-}}{\bar{\kappa}_{C+}} = \frac{1}{S_0}\frac{\gamma + \kappa_{C-}}{\kappa_{C+}} \tag{6-27}$$

with $S_0 = (\gamma k)/(\kappa_P d)$ *and* $Q_0 = (N_m \rho_0)/\gamma$ *being nondimensional parameters (see Appendix A.4.2). Now, if the system's parameters satisfy the following conditions*

$$a_S > 0, \tag{6-28a}$$

$$|k_C| \ll |a_S|, \tag{6-28b}$$

$$|k_C - 1| \ll |a_S a_T|. \tag{6-28c}$$

equation (6-24) can be recast as

$$\frac{N_{\text{ss}}}{N_m} \approx \frac{a_S}{2} + \sqrt{\left(-\frac{a_S}{2}\right)^2 - \frac{a_S}{a_T}} = \frac{a_S}{2} + \frac{a_S}{2}\sqrt{1 - \frac{4}{a_T a_S}}. \tag{6-29}$$

In addition to this, if

$$|a_S a_T| \gg 4, \tag{6-30}$$

equation (6-29) can be further simplified, yielding:

$$\frac{N_{\text{ss}}}{N_m} \approx \frac{a_S}{2} + \frac{a_S}{2} = a_S = -\frac{\kappa_P d}{\gamma^2 k}\alpha^S + 1. \tag{6-31}$$

The previous conditions (6-28a)-(6-28c) and (6-30) hold under the assumptions (6-20)-(6-22). To see this, notice that, after some algebraic manipulations, inequality (6-28a) can be rewritten as

$$\alpha^S < \frac{\gamma^2 k}{\kappa_P d}, \tag{6-32}$$

while, noting that both a_S *and* k_C *are positive, (6-28b) can be rewritten as* $k_C \ll a_S$ *and thus we obtain*

$$\alpha^S \ll \frac{\gamma^2 k}{\kappa_P d} - \gamma\frac{\gamma + \kappa_{C-}}{\kappa_{C+}} < \frac{\gamma^2 k}{\kappa_P d}. \tag{6-33}$$

Therefore, under assumption (6-20), both the previous conditions (6-28a) and (6-28b) are verified. Now, notice that assumption (6-21) implies that $k_C < 1$, *and so also that* $0 < |k_C - 1| < 1$. *Therefore, conditions (6-28c) and (6-30) can be rewritten as*

$$0 < |k_C - 1| < 1 < 4 \ll |a_T a_S| \approx |a_T|, \tag{6-34}$$

and thus both conditions (6-28c) and (6-30) are verified if

$$|a_T| \gg 4, \tag{6-35}$$

where the last step on the right hand-side in (6-34) follows from the fact that condition (6-33) implies that $a_S \approx 1$. *Substituting (6-26) into the last condition we obtain*

$$4 \ll \left|1 + \frac{Q_0}{2\gamma_e \theta S_0}\,\alpha^T\right| < 1 + \frac{Q_0}{2\gamma_e \theta S_0}\,\alpha^T, \tag{6-36}$$

which is satisfied by assumption (6-22), thus completing the proof.

A more accurate relationship linking N_{ss} and α^S can be obtained by means of a similar analysis as done in Proposition 1, in the case that the basal production rates α_0^S and α_S^T are non-zero, that is,

$$\frac{N_{\mathrm{ss}}}{N_m} \approx -\frac{\kappa_P d}{\gamma^2 k}\alpha^S + 1 - \frac{\kappa_P d}{\gamma^2 k}\alpha_0^S. \tag{6-37}$$

Even if equation (6-37) has been obtained under several simplifying assumptions, it can be observed from Figure **6-8**, left panel, that it well approximates the input-output relationship at steady state between N_{ss}/N_m and α^S of the original average model (6-14)-(6-19). Specifically, provided that conditions (6-21) and (6-22) hold, the numerical predictions of the average model overlap with those of the analytical approximation from (6-37) for values of α^S satisfying condition (6-20), which requires that $\alpha^S \ll 191\,\mathrm{transcripts/h}$, for the values of the parameters we considered in this work (see Table **A-4**). Moreover, when condition (6-22) is not strictly satisfied, the range in which equation (6-37) gives a good approximation narrows, and the range of values at which N can be regulated also decreases, due to the increasing effect of nonlinear terms that were neglected in the analysis. Therefore, conditions (6-20)-(6-22) represent constraints on the parameters of the biological system that the designer must satisfy when selecting the genetic components to implement the control system.

Note that, although we assumed α^S to be constant, Proposition 1 is still valid under the assumption that the reference parameter is slowly varying w.r.t. the dynamics of the system. Indeed, in this case the time-varying parameter can be considered as being frozen. [138] For example, the reference parameter $\alpha^S(t)$ can be taken as slowly varying when it is desired to gently move the system from one constant working point to another.

By solving (6-37) for α^S, we can obtain a useful expression to set the reference parameter α^S given the desired set-point of the population density at steady state, say N_{d}, that is,

$$\alpha_{\mathrm{ref}}^S = \frac{\gamma^2 k}{\kappa_P d}\left(1 - \frac{N_{\mathrm{d}}}{N_m}\right) - \alpha_0^S. \tag{6-38}$$

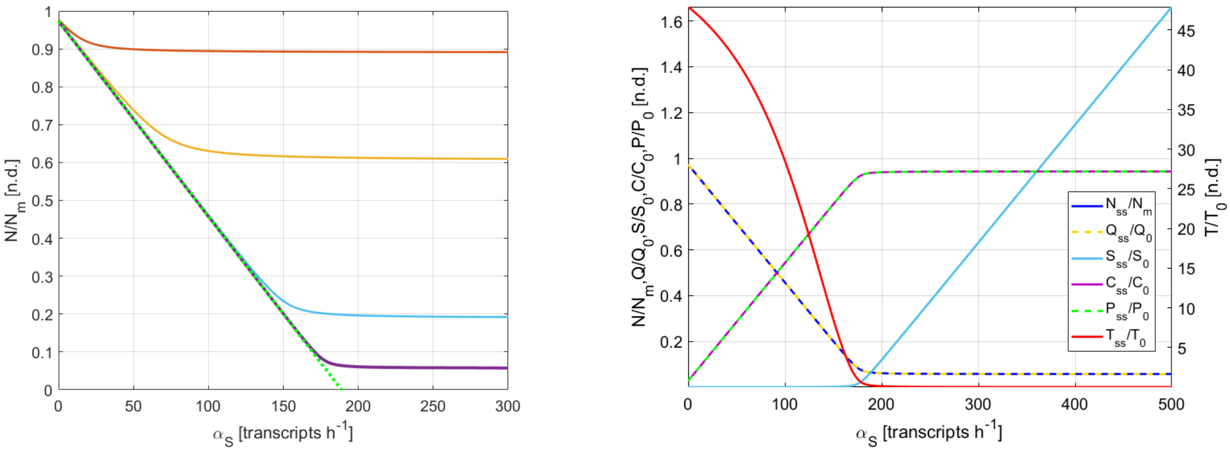


Figure 6-8: Steady-state response of the population control system. *Left panel:* Steady-state input-output relationship between the normalized cellular density N_{ss}/N_m and the reference parameter α^S obtained for the average model (6-14)-(6-19) for different values of α^T, compared to equation (6-37) (dotted green line). Namely, α^T has been selected equal to its nominal value, 11400 (purple line), 1000 (blue line), 100 (yellow line), 10 (red line), respectively. Notice that the linear function in (6-37) gives a good approximation of the steady-state relationship for values of α^S and α^T satisfying (6-20) (i.e., $\alpha^S \ll 191$) and (6-22) (i.e., $\alpha^T \gg 540$), respectively. *Right panel:* Steady-state values of the state variables of the average model as functions of α^S. All the steady-state curves were obtained by integrating the average model, with nominal values of the parameters (see Table **A-4**), in the time interval $[0, 200\,\mathrm{h}]$ starting from $N(0) = 0.1\,N_m$ and the initial value of the other state variables set to zero, by sampling α^S in the interval $[0, 500]$ with step 1. Nondimensional parameters were set to $S_0 = T_0 = C_0 = 97$, $P_0 = 242.5$, $Q_0 = 43.152$.

It can be noticed from the approximated expressions in (6-37) and (6-38) that the coefficient $\frac{\kappa_P d}{\gamma^2 k}$ represents the *static gain* of the self-regulation process and its parameters play a crucial role.

In the next section, after validating the proposed embedded control architecture in different working conditions, we numerically investigate the robustness of our design with respect to parameters' perturbations using model (6-14)-(6-19).

6.2.3 In Silico Experiments

Agent-based simulations To validate the effectiveness of the proposed population control strategy and to verify the accuracy of the analytical results obtained in Section 6.2.2, we implemented a set of *in silico* experiments by means of BSim, [122, 123] a realistic agent-based simulator of bacterial populations. BSim allows to simulate cells' dynamics by also accounting for their spatial distribution and geometry, as well as for the spatio-temporal

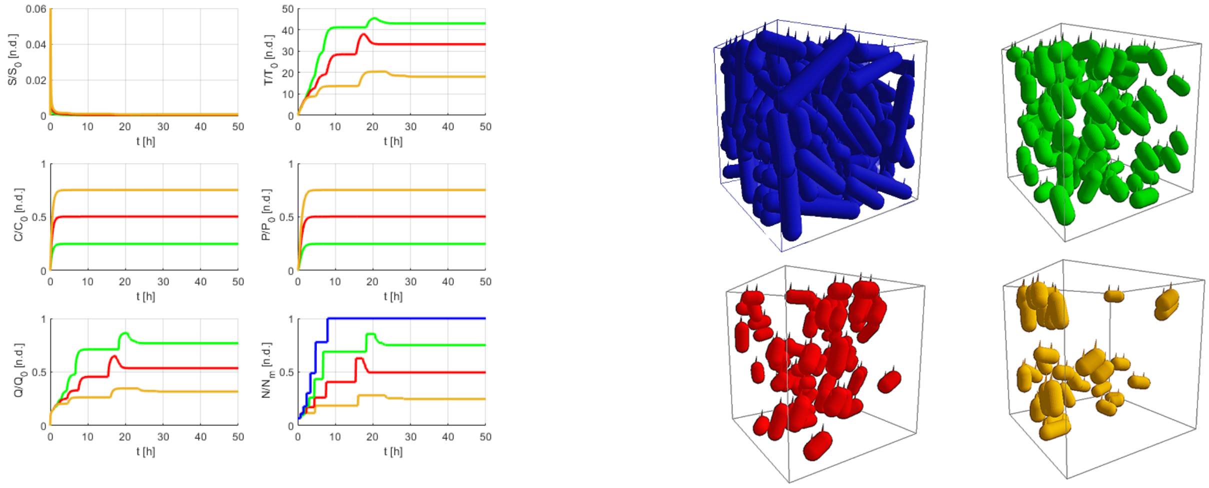


Figure 6-9: Set-point regulation experiments in BSim. *Left panel:* Time evolution of the state variables of model (6-8)-(6-12), averaged over the whole population, and of the population density N/N_m in the cases of no growth control (blue line) and with growth control and desired set-point N_d equal to $0.25\,N_m$ (yellow line), $0.50\,N_m$ (red line), and $0.75\,N_m$ (green line). The values for the desired density of the bacterial population were selected by choosing α^s as in equation (6-38). Nondimensional parameters were set to $S_0 = T_0 = C_0 = 97$, $P_0 = 242.5$, $Q_0 = 43.152$. *Right panel:* Snapshots of the culture chamber at steady state (color code correspond to the operating condition in the right panel). All cells are assumed to be identical, that is, having all the parameters equal to their nominal values, as reported in Table **A-4**. All simulations were initialized with a population of $M = 10$ cells randomly arranged in the chamber, and the initial value of the state variables in (6-8)-(6-12) set to zero.

diffusion of signaling molecules into the environment and into the cells. In addition, it is also possible to simulate cell-to-cell variability, bio-mechanics, growth, division and death of the cells.

We considered a bacterial population, endowed with the proposed genetic regulatory networks encoding our control strategy, growing into an environment with limited availability of resources. Specifically, we assumed the culture environment to be a cube-shaped micro-vial with edges set to $10\,\mu$m, that can at most host a population of 145 cells (Figure **6-9**, right panel). This in order to strike a satisfactory trade-off between the computational cost of the simulations and the number of cells in the population. The numerical routine in BSim accounting for the cell growth dynamics was modified to replicate the typical logistic growth of cell populations under limited resources (further details are reported in Appendix A.4.3). The dynamics of each cell was implemented in BSim by using the agent-based model (6-8)-(6-12), complemented by the cell growth dynamics defined in the Appendix A.4.3 (see equations (A-58) and (A-62)). The nominal values of the parameters used in the simulations are reported in Table **A-4**.

Set-point regulation We started by carrying out open-loop control experiments. As expected, in the absence of an embedded controller, the cell population density reaches the carrying capacity of the environment (Figure **6-9**, left panel, blue line) with the cells filling completely the cubic chamber, that is, $N_{\mathrm{ss}} = N_m$ (Figure **6-9**, right panel, top left chamber).

We then validated the proposed embedded genetic controller for three different set-points of the population density N_{d}, namely 25%, 50% and 75% of the carrying capacity N_m, assuming, to start with all the cells to be *identical*, that is, with all parameters being equal to their nominal values (Table **A-4**). The reference parameter α^S_{ref}, required to reach the desired set-point N_{d}/N_m, was tuned according to the linear approximation in (6-38). In practice, this can be achieved *in vivo* either by engineering it via an offline design or by means of an additional external control loop (see for instance the external control of a tunable promoter proposed in Ref [139]). The *in silico* experiments confirmed that, under nominal conditions, the proposed control strategy allows the cells to self-regulate their density to the desired set-points by dynamically adapting their growth rate (Figure **6-9**). Moreover, as shown in Figure **6-10**, left panel, the control system was found to be robust to both sudden addition or removal of cells from the chamber, being able to recover the desired set-point after a short transient. Next we test its robustness to parameter variations.

Robustness to parameter variations We analyzed the sensitivity of the controlled cells to variations of the parameters from their nominal values by running a series of simulations in MATLAB on the average model (6-14)-(6-19). Specifically, we set the desired set-point to 50% of the carrying capacity, that is, $N_{\mathrm{d}} = 0.5\, N_m$. Then, for any given parameter, say μ, and for each value of the coefficient of variation CV, we ran 200 simulations by drawing each time a parameter value from a normal distribution centered on its nominal value $\bar{\mu}$ and with standard deviation $\sigma = CV \cdot \bar{\mu}$, while keeping all the other parameters fixed to their nominal values. For each value of the coefficient of variation CV we considered, we evaluated the average percentage error at steady state over all the 200 simulations:

$$\bar{e}_{\%} = \frac{1}{200} \sum_{j=1}^{200} \left| \frac{N^j_{\mathrm{ss}} - N_{\mathrm{d}}}{N_{\mathrm{d}}} \right| \cdot 100\%, \tag{6-39}$$

where N^j_{ss} is the density at steady state reached by the cell population in the j-th simulation trial. The result of this analysis for $CV = \{0.05, 0.1, 0.15, 0.2\}$ is reported in Figure **6-10**, right panel, showing that only a subset of parameters significantly affects the steady-state

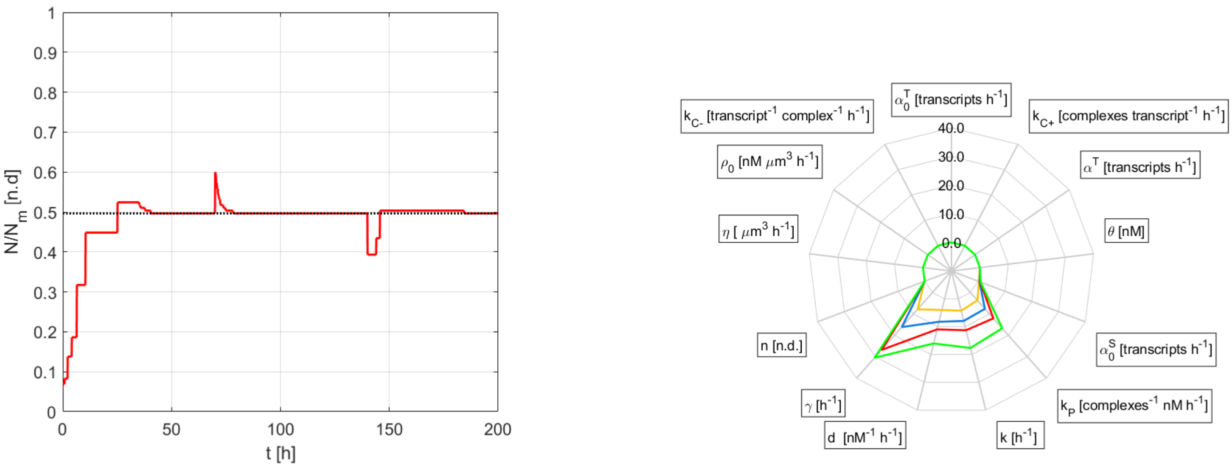


Figure 6-10: Robustness analysis. *Left panel:* Robustness to impulsive disturbances. Time evolution of the population density of the closed-loop agent-based system in BSim in the presence of addition (at $t = 70\,\text{h}$) and removal (at $t = 140\,\text{h}$) of 15 cells (corresponding to about $0.1\,N_m$). *Right panel:* Robustness to parameter variations. Average percentage error at steady state $\bar{e}_{\%}$ of the average model (6-14)-(6-19) as a function of the variability CV of the parameters' value, for CV equal to 0.05 (yellow line), 0.1 (blue line), 0.15 (red line) and 0.2 (green line). For each parameter, say μ, and each value of CV, the results of 200 simulations were averaged, each obtained by drawing the value of μ from a normal distribution centered on its nominal value $\bar{\mu}$ and with standard deviation $\sigma = CV \cdot \bar{\mu}$, while keeping all the other parameters fixed to their nominal values. All simulations were run in MATLAB by setting the set-point $N_\text{d} = 0.5\,N_m$, initial density $0.1\,N_m$, and on the time interval $[0,\ 80\,\text{h}]$.

error with a maximal variation relative to the unperturbed case of up to about 16% when $CV = 0.2$ for all parameters but γ which causes the largest variation (up to 30.44 % for $CV = 0.2$). These are the same parameters appearing in the linear approximations (6-23) and (6-37), that is, the growth rate k, the death rate d, the rate of production of the growth inhibitor protein κ_P, and the dilution rate γ.

Cell-to-cell variability Finally, cell-to-cell variability was modeled in BSim by assigning a different set of values of the parameters to daughter cells when they split from their mothers. We assumed that the heterogeneity in the response of the cells is essentially due to (i) the different copy numbers of the artificial plasmids into the cells and (ii) the different effects that the growth inhibitor protein P can have on the growth rate of each cell. Therefore, the parameters we considered to be varied in the agent-based model (6-8)-(6-12), and growth dynamics defined by (A-58) and (A-62), were α_0^S, α_0^T, α^T, κ_P, ρ_0, and the death rate d due to P. Specifically, each of these parameters, say μ, was drawn independently from a normal

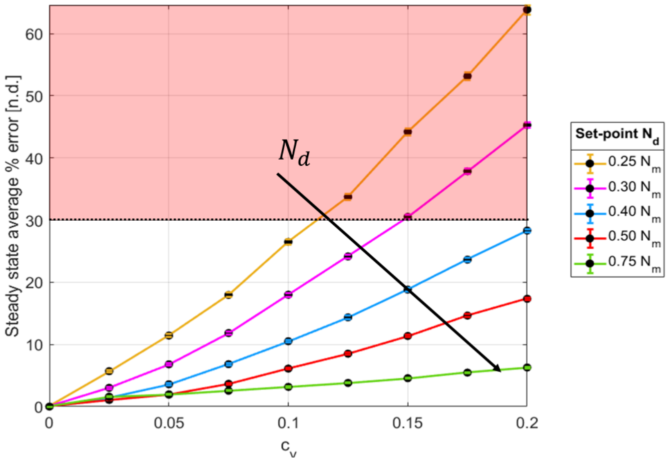


Figure 6-11: Cell-to-cell variability. Average value and standard deviation of the percentage error at steady state $\bar{e}_{\%}$ of a heterogeneous population of cells (6-8)-(6-12), for the desired set-point N_{d} equal to $0.25\,N_m$ (yellow), $0.30\,N_m$ (pink), $0.40\,N_m$ (blue), $0.50\,N_m$ (red), and $0.75\,N_m$ (green). Each point and its whiskers represent the average value and the standard deviation of $\bar{e}_{\%}$ evaluated over 100 simulation trials in BSim. For each simulation, the parameters of mother and daughter cells were drawn independently from normal distributions centered on their nominal values, $\bar{\mu}$, and with standard deviation $\sigma = CV \cdot \bar{\mu}$. All simulations were initialized with a population of 10 cells, and the initial value of the state variables in (6-8)-(6-12) set to zero.

distribution centered on its nominal value $\bar{\mu}$ and with standard deviation $\sigma = CV \cdot \bar{\mu}$. From the numerical experiments we observed that the average percentage error $\bar{e}_{\%}$, obtained over 100 simulation trials for each pair of values of the set-point N_{d} and coefficient of variation CV, increases with the heterogeneity among the cells (i.e., larger CV), with lower values of the set-point N_{d} corresponding to higher sensitivity to cell-to-cell variability (Figure **6-11**). Assuming, for example, that a steady-state percentage error below the threshold of 30% is acceptable in practice, the proposed controller guarantees good performance in all perturbed conditions when the reference set-point is greater than $0.4\,N_m$, while exhibiting larger deviations from the unperturbed case when the desired density goes below that value. However, acceptable steady-state errors can be achieved when the desired set-point values is smaller than $0.40\,N_M$ if parameters remain closer to their nominal values. In all cases, at the population level the cells exhibited very low variance over 100 trials, represented in Figure **6-11** by very small whiskers at each point.

6.2.4 Discussion

By engineering a tunable synthetic gene pathway, we were able to endow a cell population with the ability of self-regulating its own density to some desired value. We used a combination of analytical derivations and *in silico* experiments to test the performance and robustness of the proposed control architecture showing its reliability and robustness to parameter perturbations and cell-to-cell variability.

The growth control approach we presented can be used either to substitute external feedback control loops to guarantee growth regulations of cell populations in chemostats or to robustify existing external feedback control strategies, providing some regulation capability in the presence of critical failures of the sensing or actuation systems. In addition, the circuit proposed can put the basis for the realisation of synthetic systems enabling long-term coexistence of different cell populations in microbial consortia for multicellular applications. As such, the embedded control strategy we proposed can be used as an alternative solution providing a good trade-off between strategies allowing an online tuning of the consortium composition (by modifying the external environment), and strategies granting self-regulation of the consortium composition without allowing for the consortium composition to be changed dynamically.

Chapter 7

Conclusions and perspectives

The design of microbial consortia is a promising solution to expand the functional complexity of synthetically engineered cells, as each cellular population implements a specific functionality and cooperate with the others to achieve a common goal. To realize communities carrying out complex tasks it is necessary for each member of the consortium to robustly express a specific phenotype and that all population coexist within the consortium.

In this Thesis, we addressed two crucial problems towards the design of reliable, robust and modular microbial communities. One objective was to implement and validate *in vivo* a distributed feedback control loop split across populations in a two strain consortium. The other was to develop control strategies to shape the composition of a consortium, allowing stable, and long term coexistence between different populations.

In Chapter 2, we emphasized the importance of synthetic microbial consortia to expand the range of functionalities that is possible to embed in living cells. Also, we discussed the importance of mathematical modeling to speed up circuit design and of control theory to improve reliability and robustness of the realized communities. Then, we reviewed in Chapter 3 the recent efforts made to interface control theory and synthetic biology. Here, we introduced the idea of multicellular control, where the elements of a feedback control loop are distributed among different populations comprising a microbial consortium. We highlighted how this approach can alleviate the current limitations of embedded biomolecular controllers, such as limited modularity and high metabolic load on the hosts. Although different *in silico* proof-of-concept designs had been proposed in the literature (see e.g. [4, 67]), no biological realisation of a consortium implementing such paradigm has been proposed. Hence, in Chapter 4, we proposed and optimized the first *in vivo* biological implementation of a consortium comprising a controller population that was able to regulate the expression of a desired gene in a second population, the targets. We validated each component of the architecture, showing that it is effectively possible to make the two populations communicate with each other. In addition, we provided evidence that the architecture enabled robust regulation of gene expression in the targets, with the possibility of modulating the desired output by modifying the concentration of IPTG in the culture environment. Building on the biological design proposed in Chapter 4, in Chapter 5, we developed *in silico* the design of a multicellular PI controller, devising design guidelines based on the analytical study of a mathematical model of the consortium. We also developed multicellular controllers to toggle the state of a bistable memory element using two controller populations that induce activation or deactivation of an inducible genetic switch hosted in the third population.

The development of robust distributed multicellular controllers was complemented by the investigation of solutions for the problem of regulating the relative numbers between different populations within a microbial consortium. This was expounded in Chapter 6 where we developed external feedback controllers able to robustly stabilise a cell population endowed with a genetic toggle-switch functioning as a bistable memory element, even in the presence of realistic physical and technological constraints. In addition to this, in the same chapter we presented a genetic embedded controller able to steer the density of a bacterial population onto an arbitrary desired value, granting adaptation to loss or addition of cells while keeping the density close to the desired value even in the presence of cell-to-cell variability.

7.1 Open problems and perspectives

Although the work presented in this Thesis addressed crucial aspects for exploiting the full potential of synthetic microbial consortia, some key problems remain open. The main one is to realize systems enabling both robust control of gene expression and the reliable regulation of the relative sizes of populations within the consortium. That could be achieved by combining an external feedback loop regulating the size of each population within the community, such as the one presented in Chapter 6, with an embedded multicellular strategy such as the one described in Chapter 4. However, this solution requires a tightly controlled environment to work, which may not be suitable in some applications. An alternative route could be to design a single genetic controller that can also orchestrate the composition of the community. Nonetheless, this solution requires embedding additional pathways inside the controllers, adding extra metabolic load on the hosts.

Solving these problems would lead to the development of distributed feedback controllers across microbial communities with a controlled composition and spatial distribution, which is key to enable the construction of complex interacting communities, where each population carries out a specific functionality [166]. Such microbes could be used in applications ranging from the integration of bacteria in the human gut microbiome to treat a class of diseases [167] to engineering the soil microbiome to increase plant growth and health [168], and bacteria-based production of fuels, which presents a promising source for sustainable energy [21].

Bibliography

[1] Domitilla Del Vecchio, Yili Qian, Richard M Murray, and Eduardo D Sontag. Future systems and control research in synthetic biology. *Annual Reviews in Control*, 45:5–17, 2018.

[2] Domitilla Del Vecchio, Alexander J Ninfa, and Eduardo D Sontag. Modular cell biology: retroactivity and insulation. *Molecular systems biology*, 4(1):161, 2008.

[3] Philip Bittihn, M Omar Din, Lev S Tsimring, and Jeff Hasty. Rational engineering of synthetic microbial systems: from single cells to consortia. *Current Opinion in Microbiology*, 45:92–99, 2018.

[4] Gianfranco Fiore, Antoni Matyjaszkiewicz, Fabio Annunziata, et al. In-silico analysis and implementation of a multicellular feedback control strategy in a synthetic bacterial consortium. *ACS Synthetic Biology*, 6(3):507–517, 2017.

[5] Davide Fiore, Davide Salzano, Enric Cristobal-Coppulo, Josep M Olm, and Mario di Bernardo. Multicellular feedback control of a genetic toggle-switch in microbial consortia. *bioRxiv*, 2020.

[6] Elise Weill-Duflos, Virgile Andreani, Chetan Aditya, Pierre Martinon, Gregory Batt, Frederic Bonnans, and Jakob Ruess. Optimal control of an artificial microbial differentiation system for protein bioproduction. In *Proc. of the 2019 European Control Conference*, pages 2663–2668, 2019.

[7] M Khammash, Mario Di Bernardo, and Diego Di Bernardo. Cybergenetics: Theory and methods for genetic control system. In *2019 IEEE 58th Conference on Decision and Control (CDC)*, pages 916–926. IEEE, 2019.

[8] Stephanie K Aoki, Gabriele Lillacci, Ankit Gupta, Armin Baumschlager, David Schweingruber, and Mustafa Khammash. A universal biomolecular integral feedback controller for robust perfect adaptation. *Nature*, 570(7762):533–537, 2019. ** First implementation of synthetic feedback controller in living cells, particularty in *E. coli*, demonstrates tunability of the set point as well as robust perfect adaptation to a constant disturbance.

[9] Michael Chevalier, Mariana Gómez-Schiavon, Andrew H Ng, and Hana El-Samad. Design and analysis of a proportional-integral-derivative controller with biological molecules. *Cell systems*, 9(4):338–353, 2019.

[10] Mehdi Sadeghpour, Alan Veliz-Cuba, Gábor Orosz, Krešimir Josić, and Matthew R Bennett. Bistability and oscillations in co-repressive synthetic microbial consortia. *Quantitative Biology*, 5(1):55–66, 2017.

[11] James J Winkle, Bhargav R Karamched, Matthew R Bennett, William Ott, and Kresimir Josić. Emergent spatiotemporal population dynamics with cell-length control of synthetic microbial consortia. *PLoS Comput Biol*, 17(6):e1009381, 2021.

[12] Davide Salzano, Davide Fiore, and Mario di Bernardo. Ratiometric control for differentiation of cell populations endowed with synthetic toggle switches. *Conference on Decision and Control (CDC)*, 2019.

[13] Iacopo Ruolo, Sara Napolitano, Davide Salzano, Mario di Bernardo, and Diego di Bernardo. Control engineering meets synthetic biology: Foundations and applications. *Current Opinion in Systems Biology*, 28:100397, 2021.

[14] Vittoria Martinelli, Davide Salzano, Davide Fiore, and Mario di Bernardo. Multicellular pi control for gene regulation in microbial consortia. *IEEE Control Systems Letters*, 6:3373–3378, 2022.

[15] Davide Salzano, Davide Fiore, and Mario di Bernardo. Ratiometric control of cell phenotypes in monostrain microbial consortia. *Journal of the Royal Society Interface*, 19(192):20220335, 2022.

[16] Virginia Fusco, Davide Salzano, Davide Fiore, and Mario di Bernardo. Embedded control of cell growth using tunable genetic systems. *International Journal of Robust and Nonlinear Control*, 2021.

[17] Agostino Guarino, Davide Fiore, Davide Salzano, and Mario di Bernardo. Balancing cell populations endowed with a synthetic toggle switch via adaptive pulsatile feedback control. *ACS synthetic biology*, 9(4):793–803, 2020.

[18] Barbara Shannon, Criseida G Zamora-Chimal, Lorena Postiglione, Davide Salzano, Claire S Grierson, Lucia Marucci, Nigel J Savery, and Mario di Bernardo. In vivo feedback control of an antithetic molecular-titration motif in escherichia coli using microfluidics. *ACS Synthetic Biology*, 9(10):2617–2624, 2020. * An in-vivo study in microfluidics showing that it is possible to finely regulate gene expression in a population embedding an integral feedback control.

[19] Irene de Cesare, Davide Salzano, Mario di Bernardo, Ludovic Renson, and Lucia Marucci. Control-based continuation: a new approach to prototype synthetic gene networks. *ACS Synthetic Biology*, 11(7):2300–2313, 2022.

[20] Fabio Della Rossa, Davide Salzano, Anna Di Meglio, Francesco De Lellis, Marco Coraggio, Carmela Calabrese, Agostino Guarino, Ricardo Cardona-Rivera, Pietro De Lellis, Davide Liuzza, et al. A network model of italy shows that intermittent regional strategies can alleviate the covid-19 epidemic. *Nature communications*, 11(1):1–9, 2020.

[21] Jasmine Shong, Manuel Rafael Jimenez Diaz, and Cynthia H. Collins. Towards synthetic microbial consortia for bioprocessing. *Current Opinion in Biotechnology*, 23(5):798–802, 2012.

[22] Filippo Menolascina. *Synthetic Gene Circuits*. Springer, 2021.

[23] Timothy S Gardner, Charles R Cantor, and James J Collins. Construction of a genetic toggle switch in escherichia coli. *Nature*, 403(6767):339–342, 2000.

[24] Michael B Elowitz and Stanislas Leibler. A synthetic oscillatory network of transcriptional regulators. *Nature*, 403(6767):335–338, 2000.

[25] Domitilla Del Vecchio, Aaron J. Dy, and Yili Qian. Control theory meets synthetic biology. *Journal of the Royal Society Interface*, 13(120), 2016.

[26] Stephanie G. Hays, William G. Patrick, Marika Ziesack, Neri Oxman, and Pamela A. Silver. Better together: Engineering and application of microbial symbioses. *Current Opinion in Biotechnology*, 36:40–49, 2015.

[27] W. Shou, Sri Ram, and Jose M. G. Vilar. Synthetic cooperation in engineered yeast populations. *European Cells and Materials*, 7(SUPPL.1):25, 2006.

[28] Xinying Ren and Richard M. Murray. Role of interaction network topology in controlling microbial population in consortia. *Proceedings of the IEEE Conference on Decision and Control*, 2018-Decem(Cdc):2691–2697, 2019.

[29] Goni-Moreno Angel, Martyn Amos, and Fernando de la Cruz. Multicellular Computing Using Conjugation for Wiring. *PLoS*, 8(6), 2013.

[30] Ye Chen, Jae Kyoung Kim, Andrew J. Hirning, Krešimir Josić, and Matthew R. Bennett. Emergent genetic oscillations in a synthetic microbial consortium. *Science*, 349(6251):986–989, 2015.

[31] Nicolas Kylilis, Zoltan A. Tuza, Guy Bart Stan, and Karen M. Polizzi. Tools for engineering coordinated system behaviour in synthetic microbial consortia. *Nature Communications*, 9(1), 2018.

[32] Razan N. Alnahhas, James J. Winkle, Andrew J. Hirning, Bhargav Karamched, William Ott, Krešimir Josić, and Matthew R. Bennett. Spatiotemporal dynamics of synthetic microbial consortia in microfluidic devices. *ACS Synthetic Biology*, 8(9):2051–2058, 2019. PMID: 31361464.

[33] Patrick Tabeling. *Introduction to microfluidics*. OUP Oxford, 2005.

[34] Agostino Guarino, Barbara Shannon, Lucia Marucci, Claire Grierson, Nigel Savery, and Mario di Bernardo. A low-cost, open-source turbidostat design for in-vivo control experiments in synthetic biology. *IFAC-PapersOnLine*, 52(26):244–248, 2019.

[35] Jae Kyoung Kim, Ye Chen, Andrew J. Hirning, Razan N. Alnahhas, Krešimir Josić, and Matthew R. Bennett. Long-range temporal coordination of gene expression in synthetic microbial consortia. *Nature Chemical Biology*, 15(11):1102–1109, 2019.

[36] Fatima-Zahra Tani, Alain Rapaport, and Térence Bayen. A hybrid control against species invasion in the chemostat. In *2019 IEEE 58th Conference on Decision and Control (CDC)*, pages 2814–2819. IEEE, 2019.

[37] Davide Fiore, Fabio Della Rossa, Agostino Guarino, and Mario di Bernardo. Feedback ratiometric control of two microbial populations in a single chemostat. *IEEE Control Systems Letters*, 6:800–805, 2021.

[38] Souichiro Kato, Shin Haruta, Zong Jun Cui, Masaharu Ishii, and Yasuo Igarashi. Network relationships of bacteria in a stable mixed culture. *Microbial Ecology*, 56(3):403–411, 2008.

[39] Shridhar Jayanthi, Kayzad Soli Nilgiriwala, and Domitilla Del Vecchio. Retroactivity controls the temporal dynamics of gene transcription. *ACS Synthetic Biology*, 2(8):431–441, 2013.

[40] Katie Brenner, David K Karig, Ron Weiss, and Frances H Arnold. Engineered bidirectional communication mediates a consensus in a microbial biofilm consortium. *PNAS*, 104(44), 2007.

[41] Alvin Tamsir, Jeffrey J. Tabor, and Christopher A. Voigt. Robust multicellular computing using genetically encoded NOR gates and chemical 'wiresg'. *Nature*, 469(7329):212–215, 2011.

[42] Arturo Urrios, Javier Macia, Romilde Manzoni, Núria Conde, Adriano Bonforti, Eulàlia De Nadal, Francesc Posas, and Ricard Solé. A Synthetic Multicellular Memory Device. *ACS Synthetic Biology*, 5(8):862–873, 2016.

[43] Kang Zhou, Kangjian Qiao, Steven Edgar, and Gregory Stephanopoulos. Distributing a metabolic pathway among a microbial consortium enhances production of natural products. *Nature Biotechnologies*, 33(4):377–383, 2016.

[44] Xinying Ren, Ania-Ariadna Baetica, Anandh Swaminathan, and Richard M Murray. Population regulation in microbial consortia using dual feedback control. In *2017 IEEE 56th Annual Conference on Decision and Control (CDC)*, pages 5341–5347. IEEE, 2017.

[45] Spencer R Scott, M Omar Din, Philip Bittihn, Liyang Xiong, Lev S Tsimring, and Jeff Hasty. A stabilized microbial ecosystem of self-limiting bacteria using synthetic quorum-regulated lysis. *Nature Microbiology*, 2(17083), 2017.

[46] Kristina Stephens, Maria Pozo, Chen-Yu Tsao, Pricila Hauk, and William E Bentley. Bacterial co-culture with cell signaling translator and growth controller modules for autonomously regulated culture composition. *Nature Communications*, 10(4129), 2019.

[47] Alex JH Fedorec, Behzad D Karkaria, Michael Sulu, and Chris P Barnes. Single strain control of microbial consortia. *Nature Communications*, 12(1977), 2021.

[48] Razan N Alnahhas, Mehdi Sadeghpour, Ye Chen, Alexis A Frey, William Ott, Krešimir Josić, and Matthew R Bennett. Majority sensing in synthetic microbial consortia. *Nature communications*, 11(1):1–10, 2020. * Engineering a corepressive two strain microbial consortium allows the community to sense each population density and trigger a response depending on the more abundent population.

[49] Hao Song, Jun Ozaki, Cynthia H Collins, Matthew Barnet, Frances H Arnold, and Stephen R Quake. A synthetic Escherichia coli predator – prey ecosystem. *Molecular Systems Biology*, 4(187):1–8, 2008.

[50] Xinying Ren and Richard M Murray. Role of interaction network topology in controlling microbial population in consortia. In *2018 IEEE Conference on Decision and Control (CDC)*, pages 2691–2697. IEEE, 2018.

[51] Remy Chait, Jakob Ruess, Tobias Bergmiller, Gašper Tkačik, and Călin C Guet. Shaping bacterial population behavior through computer-interfaced control of individual cells. *Nature communications*, 8(1):1–11, 2017.

[52] Subhayu Basu, Yoram Gerchman, Cynthia H. Collins, Frances H. Arnold, and Ron Weiss. A synthetic multicellular system for programmed pattern formation. *Nature*, 434(7037):1130–1134, 2005.

[53] Hao Song, Ming-zhu Ding, Xiao-qiang Jia, Qian Ma, and Ying-jin Yuan. Synthetic microbial consortia : from systematic analysis to construction and applications. *Chemical Society Review*, pages 6954–6981, 2014.

[54] Spencer R. Scott, M Omar Din, Philip Bittihn, Liyang Xiong, Lev S. Tsimring, and Jeff Hasty. A stabilized microbial ecosystem of self-limiting bacteria using synthetic quorum-regulated lysis. *Physiology & behavior*, 176(3):139–148, 2017.

[55] Christian Cuba Samaniego, Nicholas A Delateur, Giulia Giordano, and Elisa Franco. Biomolecular stabilisation near the unstable equilibrium of a biological system. In *2019 IEEE 58th Conference on Decision and Control (CDC)*, pages 958–964. IEEE, 2019.

[56] Jean-Jacques E Slotine, Weiping Li, et al. *Applied nonlinear control*, volume 199. Prentice hall Englewood Cliffs, NJ, 1991.

[57] Harrison Steel, Gabriele Lillacci, Mustafa Khammash, and Antonis Papachristodoulou. Challenges at the interface of control engineering and synthetic biology. *2017 IEEE*

56th Annual Conference on Decision and Control, CDC 2017, 2018-Janua:1014–1023, 2018.

[58] Philip Bittihn, M Omar Din, Lev S Tsimring, Jeff Hasty, San Diego, La Jolla, San Diego, La Jolla, Biology Section, and San Diego. Rational engineering of synthetic microbial systems: From single cells to consortia. *Current Opinion in Microbiology*, 45:92–99, 2018.

[59] Michael J Liao, M Omar Din, Lev Tsimring, and Jeff Hasty. Rock-paper-scissors: Engineered population dynamics increase genetic stability. *Science*, 1049(September):1045–1049, 2019.

[60] Nicholas Marchand and Cynthia H Collins. Synthetic Quorum Sensing and Cell-Cell Communication in Gram- Positive Bacillus megaterium. *ACS Synthetic Biology*, pages 597–606, 2016.

[61] Travis S Bayer, Daniel M Widmaier, Karsten Temme, Ethan A Mirsky, Daniel V Santi, and Christopher A Voigt. Synthesis of Methyl Halides from Biomass Using Engineered Microbes. *Jurnal of American Chemical Society*, pages 6508–6515, 2009.

[62] Jeremy J. Minty, Marc E. Singer, Scott A. Scholz, Chang Hoon Bae, Jung Ho Ahn, Clifton E. Foster, James C. Liao, and Xiaoxia Nina Lin. Design and characterization of synthetic fungal-bacterial consortia for direct production of isobutanol from cellulosic biomass. *Proceedings of the National Academy of Sciences of the United States of America*, 110(36):14592–14597, 2013.

[63] Xinying Ren, Ania Ariadna Baetica, Anandh Swaminathan, and Richard M. Murray. Population regulation in microbial consortia using dual feedback control. *2017 IEEE 56th Annual Conference on Decision and Control, CDC 2017*, 2018-Janua(Cdc):5341–5347, 2018.

[64] Gianfranco Fiore, Antoni Matyjaszkiewicz, Fabio Annunziata, Claire Grierson, Nigel J. Savery, Lucia Marucci, and Mario Di Bernardo. In-Silico Analysis and Implementation of a Multicellular Feedback Control Strategy in a Synthetic Bacterial Consortium. *ACS Synthetic Biology*, 6(3):507–517, 2017.

[65] Elise Weill, Virgile Andréani, Chetan Aditya, Pierre Martinon, Jakob Ruess, Gregory Batt, and Frederic Bonnans. Optimal control of an artificial microbial differentiation system for protein bioproduction. *European Control Conference*, pages 2663–2668, 2019.

[66] Timothy J. Hanly and Michael A. Henson. *Unstructured modeling of a synthetic microbial consortium for consolidated production of ethanol*, volume 12. IFAC, 2013.

[67] L Postiglione, J Wan, Mario di Bernardo, and Lucia Marucci. A strategy for multicellular feedback control in mammalian cells. In *2019 18th European Control Conference (ECC)*, pages 2669–2674. IEEE, 2019.

[68] Rory L Williams and Richard M Murray. Tunable integrase-mediated differentiation facilitates improved output of burdensome functions in e. coli. *arXiv preprint arXiv:1303.3183*, 2019.

[69] Liyang Xiong, Robert Cooper, and Lev S. Tsimring. Coexistence and Pattern Formation in Bacterial Mixtures with Contact-Dependent Killing. *Biophysical Journal*, 114(7):1741–1750, 2018.

[70] Arthur Prindle, Jangir Selimkhanov, Howard Li, Ivan Razinkov, Lev S. Tsimring, and Jeff Hasty. Rapid and tunable post-translational coupling of genetic circuits. *Nature*, 508(7496):387–391, 2014.

[71] Bo Hu, Jin Du, Rui yang Zou, and Ying Jin Yuan. An environment-sensitive synthetic microbial ecosystem. *PLoS ONE*, 5(5):1–9, 2010.

[72] Corentin Briat, Christoph Zechner, and Mustafa Khammash. Design of a synthetic integral feedback circuit: dynamic analysis and dna implementation. *ACS synthetic biology*, 5(10):1108–1116, 2016.

[73] Timothy Frei, Federica Cella, Fabiana Tedeschi, Joaquín Gutiérrez, Guy-Bart Stan, Mustafa Khammash, and Velia Siciliano. Characterization and mitigation of gene expression burden in mammalian cells. *Nature communications*, 11(1):1–14, 2020. * Natural and synthetic miRNA-based incoherent feedforward loop (iFFL) circuits were identified and engineered to mitigate gene expression burden.

[74] Ross D Jones, Yili Qian, Velia Siciliano, Breanna DiAndreth, Jin Huh, Ron Weiss, and Domitilla Del Vecchio. An endoribonuclease-based feedforward controller for decoupling resource-limited genetic modules in mammalian cells. *Nature communications*, 11(1):1–16, 2020. * The effects of transcriptional and translational resource loading were quantified in engineered mammalian genetic systems and an endoribonuclease-based feedforward controller was developed in mammalian cells to adapt the expression level of a gene of interest to significant resource loading.

[75] Domitilla Del Vecchio, Aaron J Dy, and Yili Qian. Control theory meets synthetic biology. *Journal of The Royal Society Interface*, 13(120):20160380, 2016.

[76] Victoria Hsiao, Anandh Swaminathan, and Richard M Murray. Control theory for synthetic biology: recent advances in system characterization, control design, and controller implementation for synthetic biology. *IEEE Control Systems Magazine*, 38(3):32–62, 2018.

[77] Gabriele Lillacci, Stephanie Aoki, David Schweingruber, and Mustafa Khammash. A synthetic integral feedback controller for robust tunable regulation in bacteria. *BioRxiv*, page 170951, 2017.

[78] T Frei, C-H Chang, M Filo, and M Khammash. Genetically engineered integral feedback controllers for robust perfect adaptation in mammalian cells. *bioRxiv*, 2020. * First synthetic implementation of integral feedback in mammalian cells maintains constant levels of a transcription factor, even when its degradation is significantly increased, and the controlled system is structural robust.

[79] Deepak K Agrawal, Ryan Marshall, M Ali Al-Radhawi, Vincent Noireaux, and Eduardo D Sontag. Some remarks on robust gene regulation in a biomolecular integral controller. In *2019 IEEE 58th Conference on Decision and Control (CDC)*, pages 2820–2825. IEEE, 2019.

[80] Theodore W Grunberg and Domitilla Del Vecchio. Time-scale separation based design of biomolecular feedback controllers. In *2019 IEEE 58th Conference on Decision and Control (CDC)*, pages 6616–6621. IEEE, 2019.

[81] Hsin-Ho Huang, Yili Qian, and Domitilla Del Vecchio. A quasi-integral controller for adaptation of genetic modules to variable ribosome demand. *Nature communications*, 9(1):1–12, 2018.

[82] Yili Qian and Domitilla Del Vecchio. Realizing ‘integral control’ in living cells: how to overcome leaky integration due to dilution? *Journal of The Royal Society Interface*, 15(139):20170902, 2018.

[83] Gabriele Lillacci, Yaakov Benenson, and Mustafa Khammash. Synthetic control systems for high performance gene expression in mammalian cells. *Nucleic acids research*, 46(18):9855–9863, 2018.

[84] Francesca Ceroni, Alice Boo, Simone Furini, Thomas E Gorochowski, Olivier Borkowski, Yaseen N Ladak, Ali R Awan, Charlie Gilbert, Guy-Bart Stan, and Tom Ellis. Burden-driven feedback control of gene expression. *Nature methods*, 15(5):387–393, 2018.

[85] Hsin-Ho Huang, Massimo Bellato, Yili Qian, Pablo Cárdenas, Lorenzo Pasotti, Paolo Magni, and Domitilla Del Vecchio. dcas9 regulator to neutralize competition in crispri circuits. *bioRxiv*, 2020.

[86] M Omar Din, Aida Martin, Ivan Razinkov, Nicholas Csicsery, and Jeff Hasty. Interfacing gene circuits with microelectronics through engineered population dynamics. *Science advances*, 6(21):eaaz8344, 2020. ** For the first time, system output is not an engineered fluorescent protein but a canonical cellular quantity (cellular impedance, that is correlated to cell colony growth) measured by means of microelectrodes.

[87] Gianfranco Fiore, Giansimone Perrino, Mario Di Bernardo, and Diego Di Bernardo. In vivo real-time control of gene expression: a comparative analysis of feedback control strategies in yeast. *ACS synthetic biology*, 5(2):154–162, 2016.

[88] Andreas Milias-Argeitis, Marc Rullan, Stephanie K Aoki, Peter Buchmann, and Mustafa Khammash. Automated optogenetic feedback control for precise and robust regulation of gene expression and cell growth. *Nature communications*, 7(1):1–11, 2016.

[89] Jean-Baptiste Lugagne, Sebastián Sosa Carrillo, Melanie Kirch, Agnes Köhler, Gregory Batt, and Pascal Hersen. Balancing a genetic toggle switch by real-time feedback control and periodic forcing. *Nature communications*, 8(1):1–8, 2017.

[90] Marc Rullan, Dirk Benzinger, Gregor W Schmidt, Andreas Milias-Argeitis, and Mustafa Khammash. An optogenetic platform for real-time, single-cell interrogation of stochastic transcriptional regulation. *Molecular cell*, 70(4):745–756, 2018.

[91] Lorena Postiglione, Sara Napolitano, Elisa Pedone, Daniel L Rocca, Francesco Aulicino, Marco Santorelli, Barbara Tumaini, Lucia Marucci, and Diego di Bernardo. Regulation of gene expression and signaling pathway activity in mammalian cells by automated microfluidics feedback control. *ACS synthetic biology*, 7(11):2558–2565, 2018.

[92] Alexander G Goglia and Jared E Toettcher. A bright future: optogenetics to dissect the spatiotemporal control of cell behavior. *Current opinion in chemical biology*, 48:106–113, 2019.

[93] Alvaro Banderas, Matthias Le Bec, Céline Cordier, and Pascal Hersen. Autonomous and assisted control for synthetic microbiology. *International Journal of Molecular Sciences*, 21(23):9223, 2020.

[94] Jonathan P Newman, Ming-fai Fong, Daniel C Millard, Clarissa J Whitmire, Garrett B Stanley, and Steve M Potter. Optogenetic feedback control of neural activity. *Elife*, 4:e07192, 2015.

[95] Elisa Pedone, Lorena Postiglione, Francesco Aulicino, Dan L Rocca, Sandra Montes-Olivas, Mahmoud Khazim, Diego di Bernardo, Maria Pia Cosma, and Lucia Marucci. A tunable dual-input system for on-demand dynamic gene expression regulation. *Nature communications*, 10(1):1–13, 2019.

[96] Giansimone Perrino, Sara Napolitano, Francesca Galdi, Antonella La Regina, Davide Fiore, Teresa Giuliano, Mario di Bernardo, and Diego di Bernardo. Automatic synchronisation of the cell cycle in budding yeast through closed-loop feedback control. *Nature communications*, 12(1):1–12, 2021. * First example of control engineering applied to a very complex cellular mechanism. Control task is, indeed, the synchronization of cell cycle across a population of budding yeast.

[97] Patrick Harrigan, Hiten D Madhani, and Hana El-Samad. Real-time genetic compensation defines the dynamic demands of feedback control. *Cell*, 175(3):877–886, 2018.

[98] Giansimone Perrino, Cathal Wilson, Marco Santorelli, and Diego di Bernardo. Quantitative characterization of α-synuclein aggregation in living cells through automated microfluidics feedback control. *Cell reports*, 27(3):916–927, 2019.

[99] Aivar Sootla, Natalja Strelkowa, Damien Ernst, Mauricio Barahona, and Guy-Bart Stan. Toggling a genetic switch using reinforcement learning. *arXiv preprint arXiv:1303.3183*, 2013.

[100] Neythen J Treloar, Alex JH Fedorec, Brian Ingalls, and Chris P Barnes. Deep reinforcement learning for the control of microbial co-cultures in bioreactors. *PLoS computational biology*, 16(4):e1007783, 2020.

[101] Mahmoud Khazim, Lorena Postiglione, Elisa Pedone, Dan L Rocca, Carine Zahra, and Lucia Marucci. Towards automated control of embryonic stem cell pluripotency. *IFAC-PapersOnLine*, 52(26):82–87, 2019.

[102] Melinda Liu Perkins, Dirk Benzinger, Murat Arcak, and Mustafa Khammash. Cell-in-the-loop pattern formation with optogenetically emulated cell-to-cell signaling. *Nature communications*, 11(1):1–10, 2020. ** External control was used to simulate cell-to-cell signaling demonstrating the feasibility of multicellular control through quorum sensing mechanism.

[103] Gianfranco Fiore, Antoni Matyjaszkiewicz, Fabio Annunziata, Claire Grierson, Nigel J Savery, Lucia Marucci, and Mario di Bernardo. In-silico analysis and implementation of a multicellular feedback control strategy in a synthetic bacterial consortium. *ACS Synthetic Biology*, 6(3):507–517, 2017.

[104] Marvin Whiteley, Stephen P Diggle, E Peter Greenberg, and E O Wilson. Bacterial quorum sensing: the progress and promise of an emerging research area. *Nature*, 551(7680):313–320, 2017.

[105] Fabio Annunziata, Antoni Matyjaszkiewicz, Gianfranco Fiore, Claire S Grierson, Lucia Marucci, Mario di Bernardo, and Nigel J Savery. An orthogonal multi-input integration system to control gene expression in escherichia coli. *ACS synthetic biology*, 6(10):1816–1824, 2017.

[106] Michael J Liao, M Omar Din, Lev Tsimring, and Jeff Hasty. Rock-paper-scissors: Engineered population dynamics increase genetic stability. *Science*, 365(6457):1045–1049, 2019.

[107] Reed D McCardell, Ayush Pandey, and Richard M Murray. Control of density and composition in an engineered two-member bacterial community. *bioRxiv*, page 632174, 2019.

[108] Giansimone Perrino and Diego di Bernardo. Synchronisation of yeast cell cycle through quorum sensing coupling. *IFAC-PapersOnLine*, 53(2):16779–16784, 2020. 21th IFAC World Congress.

[109] Tal Danino, Octavio Mondragón-Palomino, Lev Tsimring, and Jeff Hasty. A synchronized quorum of genetic clocks. *Nature*, 463(7279):326–330, 2010.

[110] Ye Chen, Jae Kyoung Kim, Andrew J Hirning, Krešimir Josić, and Matthew R Bennett. Emergent genetic oscillations in a synthetic microbial consortium. *Science*, 349(6251):986–989, 2015.

[111] Murat Arcak. Pattern formation by lateral inhibition in large-scale networks of cells. *IEEE Transactions on Automatic Control*, 58(5):1250–1262, 2012.

[112] Virgil A Rhodius, Thomas H Segall-Shapiro, Brian D Sharon, Amar Ghodasara, Ekaterina Orlova, Hannah Tabakh, David H Burkhardt, Kevin Clancy, Todd C Peterson, Carol A Gross, et al. Design of orthogonal genetic switches based on a crosstalk map of σs, anti-σs, and promoters. *Molecular systems biology*, 9(1):702, 2013.

[113] Kathleen E McGinness, Tania A Baker, and Robert T Sauer. Engineering controllable protein degradation. *Molecular cell*, 22(5):701–707, 2006.

[114] Katherine M McKinnon. Flow cytometry: an overview. *Current protocols in immunology*, 120(1):5–1, 2018.

[115] Yves Mattenberger, Shawn Mattson, Johann Métrailler, Filo Silva, and Dominique Belin. 55.1, a gene of unknown function of phage t4, impacts on escherichia coli folate metabolism and blocks dna repair by the ner. *Molecular microbiology*, 82(6):1406–1421, 2011.

[116] DF Ackerley, Y Barak, SV Lynch, J Curtin, and A Matin. Effect of chromate stress on escherichia coli k-12. *Journal of Bacteriology*, 188(9):3371–3381, 2006.

[117] Harrison Steel, Robert Habgood, Ciarán L Kelly, and Antonis Papachristodoulou. In situ characterisation and manipulation of biological systems with chi. bio. *PLoS biology*, 18(7):e3000794, 2020.

[118] Stefan Oehler, Elisabeth R Eismann, Helmut Krämer, and Benno Müller-Hill. The three operators of the lac operon cooperate in repression. *The EMBO journal*, 9(4):973–979, 1990.

[119] Udo Wegmann, Ana Lucia Carvalho, Martin Stocks, and Simon R Carding. Use of genetically modified bacteria for drug delivery in humans: Revisiting the safety aspect. *Scientific Reports*, 7(2294), 2017.

[120] M Omar Din, Tal Danino, Arthur Prindle, et al. Synchronized cycles of bacterial lysis for in vivo delivery. *Nature*, 536:81–85, 2016.

[121] Patrick Hillenbrand, Georg Fritz, and Ulrich Gerland. Biological signal processing with a genetic toggle switch. *PLoS One*, 8(7), 2013.

[122] Thomas E Gorochowski, Antoni Matyjaszkiewicz, Thomas Todd, et al. BSim: An agent-based tool for modeling bacterial populations in systems and synthetic biology. *PLoS One*, 7(8):e42790, 2012.

[123] Antoni Matyjaszkiewicz, Gianfranco Fiore, Fabio Annunziata, et al. BSim 2.0: an advanced agent-based cell simulator. *ACS Synthetic Biology*, 6(10):1969–1972, 2017.

[124] Timothy S Gardner, Charles R Cantor, and James J Collins. Construction of a genetic toggle switch in Escherichia coli. *Nature*, 403:339–342, 2000.

[125] Jean-Baptiste Lugagne, Sebastián Sosa Carrillo, Melanie Kirch, et al. Balancing a genetic toggle switch by real-time feedback control and periodic forcing. *Nature Communications*, 8(1671), 2017.

[126] Davide Fiore, Agostino Guarino, and Mario di Bernardo. Analysis and control of genetic toggle switches subject to periodic multi-input stimulation. *IEEE Control Systems Letters*, 3(2):278–283, 2019.

[127] A. Guarino, D. Fiore, D. Salzano, and M. di Bernardo. Balancing cell populations endowed with a synthetic toggle switch via adaptive pulsatile feedback control. *ACS Synthetic Biology*, 9(4):793–803, 2020.

[128] Christian Cuba Samaniego, Nicholas A Delateur, Giulia Giordano, and Elisa Franco. Biomolecular stabilisation near the unstable equilibrium of a biological system. In *Proc. of the 58th IEEE Conference on Decision and Control*, pages 958–964, 2019.

[129] Christian Cuba Samaniego and Elisa Franco. A robust molecular network motif for period-doubling devices. *ACS Synthetic Biology*, 7(1):75–85, 2018.

[130] Joshua L Cherry and Frederick R Adler. How to make a biological switch. *Journal of Theoretical Biology*, 203(2):117–133, 2000.

[131] Aivar Sootla, Diego Oyarzún, David Angeli, and Guy-Bart Stan. Shaping pulses to control bistable systems: Analysis, computation and counterexamples. *Automatica*, 63:254–264, 2016.

[132] Corentin Briat, Ankit Gupta, and Mustafa Khammash. Antithetic integral feedback ensures robust perfect adaptation in noisy biomolecular networks. *Cell Systems*, 2(1):15–26, 2016.

[133] Octavio Mondragón-Palomino et al. Entrainment of a population of synthetic genetic oscillators. *Science*, 333(6047):1315–1319, 2011.

[134] Kevin D Litcofsky, Raffi B Afeyan, Russell J Krom, Ahmad S Khalil, and James J Collins. Iterative plug-and-play methodology for constructing and modifying synthetic gene networks. *Nature Methods*, 9:1077–1080, 2012.

[135] Chetan Aditya, François Bertaux, Gregory Batt, and Jakob Ruess. A light tunable differentiation system for the creation and control of consortia in yeast. *Nature communications*, 12(1):1–10, 2021.

[136] Joaquin Gutierrez, Sant Kumar, and Mustafa Khammash. Dynamic cybergenetic control of bacterial co-culture composition via optogenetic feedback. *bioRxiv*, 2022.

[137] Danielle B. Pedrolli, Nathan V. Ribeiro, Patrick N. Squizato, et al. Engineering microbial living therapeutics: The synthetic biology toolbox. *Trends in Biotechnology*, 37(1):100 – 115, 2019.

[138] Hassan K Khalil and Jessy W Grizzle. *Nonlinear systems*, volume 3. Prentice hall Upper Saddle River, NJ, 2002.

[139] Filippo Menolascina, Gianfranco Fiore, Emanuele Orabona, Luca De Stefano, Mike Ferry, Jeff Hasty, Mario di Bernardo, and Diego di Bernardo. In-vivo real-time control of protein expression from endogenous and synthetic gene networks. *PLoS Computational Biology*, 10(5):e1003625, 2014.

[140] Gianfranco Fiore, Giansimone Perrino, Mario di Bernardo, and Diego di Bernardo. In vivo real-time control of gene expression: a comparative analysis of feedback control strategies in yeast. *ACS Synthetic Biology*, 5(2):154–162, 2015.

[141] Karl Johan Aström and Richard M Murray. *Feedback systems: an introduction for scientists and engineers.* Princeton University Press, 2010.

[142] Agostino Guarino, Davide Fiore, and Mario di Bernardo. In-silico feedback control of a MIMO synthetic toggle switch via pulse-width modulation. In *Proc. of the 18th European Control Conference*, pages 680–685, 2019.

[143] Alejandro F Villaverde, Nikolaos Tsiantis, and Julio R Banga. Full observability and estimation of unknown inputs, states and parameters of nonlinear biological models. *Journal of the Royal Society Interface*, 16(156):20190043, 2019.

[144] Mike S Ferry, Ivan A Razinkov, and Jeff Hasty. Microfluidics for synthetic biology: from design to execution. In *Methods in Enzymology*, volume 497, pages 295–372. Elsevier, 2011.

[145] Marco Mauri, Jean-Luc Gouzé, Hidde De Jong, and Eugenio Cinquemani. Enhanced production of heterologous proteins by a synthetic microbial community: Conditions and trade-offs. *PLoS Comput Biol*, 16(4):e1007795, 2020.

[146] Tat-Ming Lo, Si Hui Chng, Wei Suong Teo, Han-Saem Cho, and Matthew Wook Chang. A two-layer gene circuit for decoupling cell growth from metabolite production. *Cell Systems*, 3(2):133–143, 2016.

[147] Rongzhen Tian, Yanfeng Liu, Yanting Cao, Zhongjie Zhang, Jianghua Li, Long Liu, Guocheng Du, and Jian Chen. Titrating bacterial growth and chemical biosynthesis for efficient n-acetylglucosamine and n-acetylneuraminic acid bioproduction. *Nature Communications*, 11(5078), 2020.

[148] Peng Xu. Production of chemicals using dynamic control of metabolic fluxes. *Current Opinion in Biotechnology*, 53:12–19, 2018.

[149] Yongkun Lv, Shuai Qian, Guocheng Du, Jian Chen, Jingwen Zhou, and Peng Xu. Coupling feedback genetic circuits with growth phenotype for dynamic population control and intelligent bioproduction. *Metabolic Engineering*, 54:109–116, 2019.

[150] Jacques Monod. The growth of bacterial cultures. *Annual Review of Microbiology*, 3(1):371–394, 1949.

[151] Patrick De Leenheer and Hal Smith. Feedback control for chemostat models. *Journal of Mathematical Biology*, 46(1):48–70, 2003.

[152] Guang-Yan Zhu, Abdelqader Zamamiri, Michael A Henson, and Martin A Hjortsø. Model predictive control of continuous yeast bioreactors using cell population balance models. *Chemical Engineering Science*, 55(24):6155–6167, 2000.

[153] Amos Richmond, Zhang Cheng-Wu, and Yair Zarmi. Efficient use of strong light for high photosynthetic productivity: interrelationships between the optical path, the optimal population density and cell-growth inhibition. *Biomolecular Engineering*, 20(4-6):229–236, 2003.

[154] Lingchong You, Robert Sidney Cox, Ron Weiss, and Frances H Arnold. Programmed population control by cell–cell communication and regulated killing. *Nature*, 428(6985):868–871, 2004.

[155] Behzad D Karkaria, Alex JH Fedorec, and Chris P Barnes. Automated design of synthetic microbial communities. *Nature Communications*, 12(672), 2021.

[156] Christina V Dinh, Xingyu Chen, and Kristala LJ Prather. Development of a quorum-sensing based circuit for control of coculture population composition in a naringenin production system. *ACS Synthetic Biology*, 9(3):590–597, 2020.

[157] Vittorio Bartoli, Grace A Meaker, Mario di Bernardo, and Thomas E Gorochowski. Tunable genetic devices through simultaneous control of transcription and translation. *Nature Communications*, 11(2095), 2020.

[158] Vittorio Bartoli, Mario di Bernardo, and Thomas E. Gorochowski. Self-adaptive biosystems through tunable genetic parts and circuits. *Current Opinion in Systems Biology*, 24:78–85, 2020.

[159] Ye Chen, Joanne ML Ho, David L Shis, Chinmaya Gupta, James Long, Daniel S Wagner, William Ott, Krešmir Josić, and Matthew R Bennett. Tuning the dynamic range of bacterial promoters regulated by ligand-inducible transcription factors. *Nature Communications*, 9(64), 2018.

[160] Robert M Hughes. A compendium of chemical and genetic approaches to light-regulated gene transcription. *Critical Reviews in Biochemistry and Molecular Biology*, 53(5):453–474, 2018.

[161] Armin Baumschlager, Marc Rullan, and Mustafa Khammash. Exploiting natural chemical photosensitivity of anhydrotetracycline and tetracycline for dynamic and setpoint chemo-optogenetic control. *Nature Communications*, 11(3834), 2020.

[162] Evgeni V Nikolaev and Eduardo D Sontag. Quorum-sensing synchronization of synthetic toggle switches: A design based on monotone dynamical systems theory. *PLoS computational biology*, 12(4):e1004881, 2016.

[163] Richard Casey, Hidde De Jong, and Jean-Luc Gouzé. Piecewise-linear models of genetic regulatory networks: equilibria and their stability. *Journal of Mathematical Biology*, 52(1):27–56, 2006.

[164] Erik Plahte and Sissel Kjøglum. Analysis and generic properties of gene regulatory networks with graded response functions. *Physica D: Nonlinear Phenomena*, 201(1):150–176, 2005.

[165] Grégory Batt, Calin Belta, and Ron Weiss. Temporal logic analysis of gene networks under parameter uncertainty. *IEEE Transactions on Automatic Control*, 53(Special Issue):215–229, 2008.

[166] Nicolas E Grandel, Kiara Reyes Gamas, and Matthew R Bennett. Control of synthetic microbial consortia in time, space, and composition. *Trends in Microbiology*, 29(12):1095–1105, 2021.

[167] In Young Hwang and Matthew Wook Chang. Engineering commensal bacteria to rewire host–microbiome interactions. *Current opinion in biotechnology*, 62:116–122, 2020.

[168] Inessa Arif, Maria Batool, and Peer M Schenk. Plant microbiome engineering: expected benefits for improved crop growth and resilience. *Trends in Biotechnology*, 38(12):1385–1396, 2020.

[169] MS Guyer, RR Reed, JA Steitz, and KB Low. Identification of a sex-factor-affinity site in e. coli as $\gamma\delta$. In *Cold Spring Harbor symposia on quantitative biology*, volume 45, pages 135–140. Cold Spring Harbor Laboratory Press, 1981.

[170] Abigail J Smith, Franziska Thomas, Deborah Shoemark, Derek N Woolfson, and Nigel J Savery. Guiding biomolecular interactions in cells using de novo protein–protein interfaces. *ACS Synthetic Biology*, 8(6):1284–1293, 2019.

[171] Harrison Steel, Robert Habgood, Ciarán Kelly, and Antonis Papachristodoulou. Chi. bio: An open-source automated experimental platform for biological science research. *bioRxiv*, page 796516, 2019.

[172] Yuri A Kuznetsov. *Elements of applied bifurcation theory*. Springer-Verlag, 1998.

[173] Agostino Guarino, Davide Fiore, and Mario di Bernardo. In-silico feedback control of a MIMO synthetic Toggle Switch via Pulse-Width Modulation. In *Proc. of the 2019 European Control Conference*, pages 680–685, 2019.

[174] Eszter Lakatos. *Stochastic analysis and control methods for molecular cell biology*. PhD thesis, Imperial College London, 2017.

[175] Victoria Hsiao, Anandh Swaminathan, and Richard M. Murray. Control Theory for Synthetic Biology: Recent Advances in System Characterization, Control Design, and Controller Implementation for Synthetic Biology. *IEEE Control Systems*, 38(3):32–62, 2018.

[176] M Gen and R Cheng. *Genetic algorithms and engineering optimization*. John Wiley & Sons, 2000.

[177] Ping Wang, Lydia Robert, James Pelletier, Wei Lien Dang, Francois Taddei, Andrew Wright, and Suckjoon Jun. Robust growth of escherichia coli. *Current Biology*, 20(12):1099–1103, 2010.

[178] Lucia Marucci. Nanog dynamics in mouse embryonic stem cells: results from systems biology approaches. *Stem cells international*, 2017, 2017.

[179] Ann C Babtie, Paul Kirk, and Michael PH Stumpf. Topological sensitivity analysis for systems biology. *Proceedings of the National Academy of Sciences*, 111(52):18507–18512, 2014.

[180] Margrith E Mattmann and Helen E Blackwell. Small molecules that modulate quorum sensing and control virulence in pseudomonas aeruginosa. *The Journal of organic chemistry*, 75(20):6737–6746, 2010.

[181] Jean-Denis Pédelacq, Stéphanie Cabantous, Timothy Tran, Thomas C Terwilliger, and Geoffrey S Waldo. Engineering and characterization of a superfolder green fluorescent protein. *Nature biotechnology*, 24(1):79–88, 2006.

[182] Susan Gottesman, Eric Roche, YanNing Zhou, and Robert T Sauer. The clpxp and clpap proteases degrade proteins with carboxy-terminal peptide tails added by the ssra-tagging system. *Genes & development*, 12(9):1338–1347, 1998.

[183] Karl Johan Åström and Richard M Murray. *Feedback systems: an introduction for scientists and engineers*. Princeton university press, 2021.

[184] Giuseppe C Calafiore, Carlo Novara, and Corrado Possieri. A modified sir model for the covid-19 contagion in italy. In *2020 59th IEEE Conference on Decision and Control (CDC)*, pages 3889–3894. IEEE, 2020.

[185] Giulia Giordano, Franco Blanchini, Raffaele Bruno, Patrizio Colaneri, Alessandro Di Filippo, Angela Di Matteo, and Marta Colaneri. Modelling the covid-19 epidemic and implementation of population-wide interventions in italy. *Nature medicine*, 26(6):855–860, 2020.

[186] Liangrong Peng, Wuyue Yang, Dongyan Zhang, Changjing Zhuge, and Liu Hong. Epidemic analysis of covid-19 in china by dynamical modeling. *arXiv preprint arXiv:2002.06563*, 2020.

[187] Francesco Casella. Can the covid-19 epidemic be controlled on the basis of daily test reports? *IEEE Control Systems Letters*, 5(3):1079–1084, 2020.

[188] Cameron Nowzari, Victor M Preciado, and George J Pappas. Optimal resource allocation for control of networked epidemic models. *IEEE Transactions on Control of Network Systems*, 4(2):159–169, 2015.

[189] Eduardo Ramírez-Llanos and Sonia Martínez. A distributed dynamics for virus-spread control. *Automatica*, 76:41–48, 2017.

[190] Julien Arino and P Van den Driessche. A multi-city epidemic model. *Mathematical Population Studies*, 10(3):175–193, 2003.

[191] Ayalvadi Ganesh, Laurent Massoulié, and Don Towsley. The effect of network topology on the spread of epidemics. In *Proceedings IEEE 24th Annual Joint Conference of the IEEE Computer and Communications Societies.*, volume 2, pages 1455–1466. IEEE, 2005.

[192] Marino Gatto, Enrico Bertuzzo, Lorenzo Mari, Stefano Miccoli, Luca Carraro, Renato Casagrandi, and Andrea Rinaldo. Spread and dynamics of the covid-19 epidemic in italy: Effects of emergency containment measures. *Proceedings of the National Academy of Sciences*, 117(19):10484–10491, 2020.

[193] Herbert W Hethcote. An immunization model for a heterogeneous population. *Theoretical population biology*, 14(3):338–349, 1978.

[194] Ana Lajmanovich and James A Yorke. A deterministic model for gonorrhea in a nonhomogeneous population. *Mathematical Biosciences*, 28(3-4):221–236, 1976.

[195] Wenjun Mei, Shadi Mohagheghi, Sandro Zampieri, and Francesco Bullo. On the dynamics of deterministic epidemic propagation over networks. *Annual Reviews in Control*, 44:116–128, 2017.

[196] Lisa Sattenspiel and Klaus Dietz. A structured epidemic model incorporating geographic mobility among regions. *Mathematical biosciences*, 128(1-2):71–91, 1995.

[197] Lucas M Stolerman, Daniel Coombs, and Stefanella Boatto. Sir-network model and its application to dengue fever. *SIAM Journal on Applied Mathematics*, 75(6):2581–2609, 2015.

[198] Anna Mummert and Olusegun M Otunuga. Parameter identification for a stochastic seirs epidemic model: case study influenza. *Journal of mathematical biology*, 79(2):705–729, 2019.

[199] Michelangelo Bin, Peter Cheung, Emanuele Crisostomi, Pietro Ferraro, Connor Myant, Thomas Parisini, and Robert Shorten. On fast multi-shot epidemic interventions for post lock-down mitigation: Implications for simple covid-19 models. *arXiv preprint arXiv:2003.09930*, 1125, 2020.

[200] Ruiyun Li, Sen Pei, Bin Chen, Yimeng Song, Tao Zhang, Wan Yang, and Jeffrey Shaman. Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (sars-cov-2). *Science*, 368(6490):489–493, 2020.

Appendix A

Supplementary material

A.1 Methodologies for *in-vivo* implementation of the multicellular control

.

A.1.1 Strains and constructs

Escherichia coli strain TOP10 was used for all the cloning manipulations of this work. *Escherichia coli* strain MG1655 ($\lambda-$, rph $-$ 1)[169] was used for all the assays in this work, and transformed with the plasmids constituting either the controllers or the targets. The controllers embed three plasmids: pVRa20_992_BS2 (medium copy number), pVRc20_992_BS1 (medium copy number) and pVRb_LasI (low copy number). The genes constituting the essential core of the computation module (σ, anti-σ and lasI) are ssrA tagged (AANDENYALAA) to ensure fast dynamics of expression and degradation.

Vector backbones were derived from plasmids pVRa20_992 and pVRc20_992 [112].

The pVRb_lasI plasmid encodes the lasI gene, under the control of the *p_20992* promoter, and was was built in [105], together with the pVRb_ssrA where the lasI was swapped with a GFP and that was trasformed in the controllers for their input output characterisation.

The pVRa20_992_BS2 produces the σ factor proporionally to the IPTG concentration present in the culture media. To construct the plasmid first the pVRa20_992_BS1 plasmid was obtained by standard restriction digestion and ligation procedure, by amplifying the Sigma gene from the pLusB plasmid [105] with primers Sigma_plus_tag_F and Sigma_plus_tag_F and cloning the PCR product in the pVRa_20_992 plasmid [112] after NcoI/BamHI digestion. Then, the *pLac* promoter was substituted with the *pLac-UV5* promoter to obtain the pVRa20_992_BS2 plasmid. The construct was built by standard restriction digestion and ligation procedure, by amplifying the *pLac-UV5* from the pVRa_placASb_Flag plasmid [105] with primers PlacUV5_F and PlacUV5_R and cloning the PCR product in the pVRa_20_992_BS1 plasmid after XbaI digestion.

The pVRc20_992_BS1 produces the anti-σ factor proportionally to the concentration of 3-O-C6-HSL present in the culture media. The plasmid was constructed by standard restriction digestion and ligation procedure, by amplifying the anti-σ from the pVRa_placASb_Flag plasmid [105] with primers Anti_Sigma_plus_tag_F and Anti_Sigma_plus_tag_R and cloning the PCR product in the pVRc_20_992 plasmid after BamHI/PstI digestion.

Also, the improved version of the controllers, where a second lac operator was introduced upstream the *plac_UV5* promoter was transformed substituting the pVRa20_992_BS2 plasmid with the pVRa20_992_DS plasmid. The pVRa20_992_DS was constructed by Gibson assembly after PCR amplification of the pVRa20_992_BS2 using primers Back_extralac_for and Back_extralac_rev, and of the pVRbLacO1O1Del from [170] using the Ins_extralac_for and Ins_extralac_rev primers.

Finally, the controllers expressing also a Red Fluorescent Protein were implemented by substituting the pVRb_lasI plasmid with the pVRb_lasI_RFP plasmid. The pVRb_lasI_RFP was assembled using standard restriction digestion and ligation procedure, by amplifying a synthetically generated RFP with primers RFP_Afe_for and RFP_Nde_rev and cloning the PCR product in the pVRb_lasI plasmid after AfeI/NdeI digestion.

The targets embedded the 3-O-C12-HSL inducible promoter *plas* driving the expression of a green fluorescent protein and of the luxI gene. They were transformed with the p_Las_Lux_GFP_3.0 plasmid. This plasmid was assembled by Gibson asseembly after PCR amplification of the las_composite_device plasmid from [31] with p_Las_Lux_GFP_3_0_fwd_20 and p_Las_Lux_GFP_3_0_rev_20 primers, and of the luxI from pTeLu_GFP constructed in [105] with the luxI_fwd and luxI_rev primers. Finally, the targets with luxI fused with a GFP were transformed with the p_Las_Lux_GFP_fusion plasmid, constructed via standard restriction digestion and ligation procedure, by ligating a synthetically generated luxI GFP fused protein with p_Las_Lux_GFP_3.0 after digestion with AseI/BsrGI. All the primers sequences and the sequences of the synthetic genes are reported in Tables **A-1** and **A-2**, respectively.

A.1.2 Chemicals

For all experiments, kanamycin (50 μg/mL), ampicillin (100 μg/mL) and chloramphenicol (25 μg/mL) were added to the growth media. These antibiotics were supplied by Sigma-Aldrich. 3-O-C6-HSL (N-(B-Ketocaproyl)-L-Homoserine Lactone from Sigma-Aldrich, cat #

Primer Name	Sequence (5'→3')
Sigma_plus_tag_F	GAGGAGAAATACCATAATCC
Sigma_plus_tag_R	GTTTTATCAGACCGCTTC
placUV5_F	TGGTTTCACATTCACCAC
placUV5_R	GCTAGATCTAGAGAAATTGTTATCCGCTCACA
Anti_Sigma_plus_tag_F	ATGTCATATGCCGCGATAGAGGAGGTAA
Anti_Sigma_plus_tag_R	CTCTCATCCGCCAAAACA
Back_extralac_for	CACCCCAGGCTTTA
Back_extralac_rev	ATTCGGGATCGAGAT
Ins_extralac_for	GAATTCCGATTCATTAATG
Ins_extralac_rev	CGGAAGCATAAAGTGT
RFP_Afe_for	TGATCGAGCGCTGATAAGTCCCTAACTTTTACAGC
RFP_Nde_rev	TGCTCACATATGGGACCAAAACGAAAAAAGGC
p_Las_Lux_GFP_3_0_fwd_20	ACGCCCTTGCAGCGTAATAATACTAGAGAAAGAGGAGAAATACTAGATGCG
p_Las_Lux_GFP_3_0_rev_20	CTCTATCGCGGAAATTGACACTCGGCGTTATGTCATGAAG
luxI_for	TGTCAATTTCCGCGATAGAGGAG
luxI_rev	TTATTACGCTGCAAGGGCGTA

Table A-1: Names and sequences of the primers used for the cloning of the constructs used in this study.

K3007) and 3-O-C12-HSL (N-(3-Oxododecanoyl)-L-homoserine lactone from Sigma-Aldrich, cat # O9139) were dissolved in water, filter-sterilized and added to the LB growth medium at the indicated concentrations. IPTG (Isopropyl β-D-1-thiogalactopyranoside, supplied by Sigma-Aldrich, cat # I5502) was dissolved in water, filter-sterilized and added to LB medium at the indicated concentrations. M9 salts were prepared by suspending in 1L of water 6g Na_2HPO4, 3g KH_2PO_4, 1g NH_4Cl and 0.5g NaCl, then autoclaving the solution. M9 media was obtained by supplementing 50 mL of M9 salts with 500 μl 20% casamino acids, 100 μl 1M $MgSO_4$, 250 μl 0.1M $CaCl_2$, 250 μl 50% glycerol and 5 μl 20 mg/ml thiamine.

A.1.3 Samples preparation and analysis using flow cytometry

At the end of each experiment, each sample (0.5 mL volume) was resuspended in 1xPBS and diluted so as to achieve a final OD for the culture of 0.05. Each sample was then analysed using a flow cytometer, and the results analysed using FlowJo.

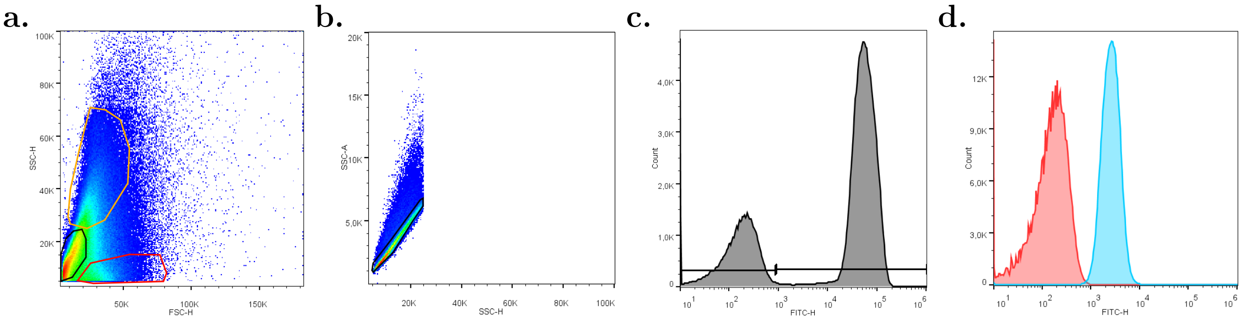


Figure A-1: **a.** Density plot of events in the FSC-H,SSC-H plane. Red areas represent zones of the plane with more events. The Polygons drawn represent events classified as healthy cells (black), stressed cells (orange) and cellular debris (red). **b.** Density plot healthy cells in the SSC-H,SSC-A plane. Red areas represent zones of the plane with more events. The black polygon drawn is the gate used to select single cells against aggregates. **c.** Histogram of single cells in the FITC-H (Green Fluorescence) channel. The two horizontal gates highlights controllers and targets in the consortium. **d.** Comparison of the FITC-h fluorescent profiles of controllers (red) and uninduced targets (blue).

A.1.4 Gating Strategy

The events recorded by a flow cytometer (BD LSRFortessa X-20 or Acea Novocyte) were first gated in the forward scatter (FSC-H, proxy of cellular size), side scatter (SSC-H, proxy of celllular complexity) plane. Here, the events were divided in three subpopulations: healthy cells where size and complexity were in physiologiacal ranges [116], stressed cells where the bacteria had increased size and complexity due to filamentation, and cellular debris corresponding to events with high size with respect to their complexity (Figure **A-1**a). Subsequently, healthy cells were gated in the SSC-H, SSC-A plane to select for single cells. In detail, cells with comparable signals in the two channels were selected as single cells, whereas events with much larger SSC-A were classified as cellular conglomerates and excluded from the analysis (Figure **A-1**b). Finally, single cells were gated using the FITC-H (490 nm excitation, 530/30 nm emission) channel to discriminate targets and controllers based on their fluorescence levels. Specifically, cells with low FITC-H signal were classified as controllers and cells having high FITC-H fluorescence were selected to be targets (Figure **A-1**c). Alternatively, when controllers constitutively expressed RFP, events with a high signal in the PE-CF594 (560 nm excitation, 610/20 nm emission filter, Red) channel were distinguished as controllers while cells with low signal were designated as targets.

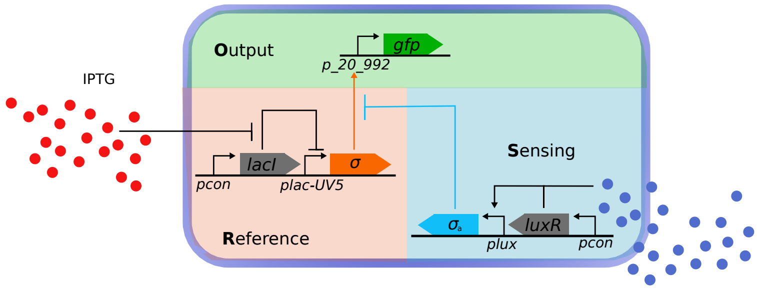


Figure A-2: Schematic representation of the controllers used for the characterisation of their input-output response. In this population the plasmid expressing lasI under the *p_20_992* promoter was modified substituting lasI with a green fluorescent protein (GFP).

A.1.5 Induction protocol

Cells were cultured in 10 mL LB broth in 50 mL Falcons. Overnight cultures were diluted 1:500 in fresh sterile media supplemented with the inducer(s) at the specified concentration. The culture was then incubated at 37°C, shaking at 250rpm for 3h. Then, the samples were prepared and analysed by flow cytometry. This protocol was used to obtain Figures **4-2**a,b and **4-4**b.

A.1.6 Cell-cell communication assay

targets/controllers were diluted in 10 mL 1:500 from Overnight cultures in 10 mL LB broth supplemented with the specified concentration of inducer(s). The culture was then incubated at 37°C, shaking at 250rpm for 3h. Past this time, the culture was spun at 3500rpm for 8' at room temperature to pellet the cells. The supernatant was filtered using a 0.2μm filter in a sterile tube and an overnight culture of the other population was diluted 1:500 in the filtered media. The culture was finally incubated again for 3h at 37°C, shaking at 250rpm. Finally, culture 0.5 mL were sampled and analysed by flow cytometry. This protocol was used to obtain the results shown in Figure **4-2**c,d.

A.1.7 Time-courses in LB

Overnight cultures of targets and controllers were diluted in 35 mL LB supplemented with the specified concentration of IPTG. The initial dilutions used for the populations were 1:600 and 1:3000, respectively. Both cultures were incubated at 37°C, shaking at 250rpm for 6h. Each hour, 5 mL of both cultures were mixed in a sterile 50 mL tube, which was then incubated alongside the single strain cultures. After 6h, 0.5 mL of all mixed cultures were sampled and analysed via flow cytometry. Note that the hours in Figures **4-3**, **4-6** and **4-7** refer to the how long targets and controllers have been mixed for, not to their incubation time. This protocol was followed to retrieve the results portrayed in Figures **4-3**a,c, **4-4**c, **4-5**a, **4-6**b,c and **4-7**f.

A.1.8 Consortium composition assay

Overnight cultures of targets and controllers were diluted in 10 mL LB using different targets:controllers ratios. Specifically, 15:1, 1:10, 1:8, 1:6, 1:5, 1:4, 1:3, 1:2, 1:1.5, 1:1 ratios were used. The volume added of each population was computed so as to have a cell:media dilution of 1:500. All cultures were incubated at $37°C$, shaking at 250rpm for 6h. After the incubation 0.5 mL of each culture were sampled and analysed by flow cytometry. This protocol was used in Figures **4-3**b and **4-7**a-d.

A.1.9 Chi.Bio timecourse

10 mL of overnight cultures of controllers and targets were mixed in a 30 mL sterile glass tube. The protocol described in [171] was followed to set up the chemostat and calibrate the optical density (OD) measurements. The reference OD was set to 1.0. Each hour 0.5 mL from the continuous culture were sampled and stored at $4°C$ until the end of the timecouse, when all the samples were prepared and analysed by flow cytometry. This protocol was followed to obtain Figures **4-5**b and **A-5**b.

A.1.10 M9 Timecourse experiments

Overnight cultures of controllers and targets were mixed in a 1:5 ratio in 10 mL M9 media supplemented with the specified IPTG concentration, so as to have an overall dilution of 1:1000. The culture was then incubated at $37°C$, shaking at 250rpm for 6h. After each hour, 0.5 mL were sampled from the culture and stored at $4°C$. After 6h cells were spun at 3500rpm for 8' at room temperature. The supernatant was filtered, the cells resuspended in fresh media and diluted 1:500 in their supernatant. The diluted culture was then incubated for additional 6h. At the end of the timecourse all cells were prepared and analysed by flow cytometry. This protocol was used to retrieve Figure **4-5**c.

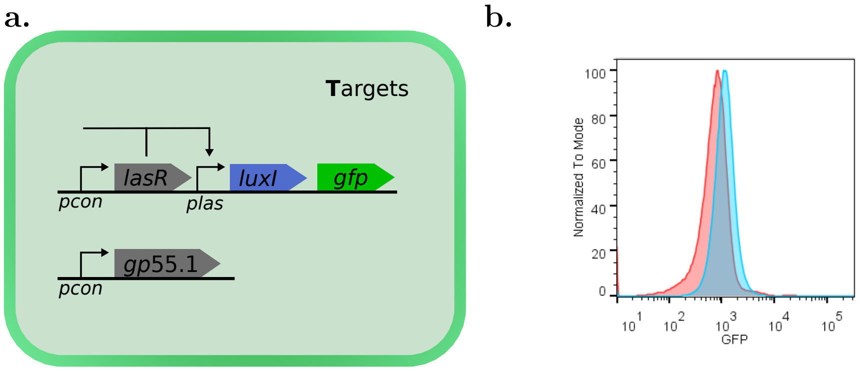


Figure A-3: **a.** Targets bearing extra metabolic load. A burdensome protein (GP55.1) was constitutively expressed in the target population. **b.** Comparison of the FITC-H(green channel) fluorescent profiles of targets with (red) and without (blue) the additional metabolic load.

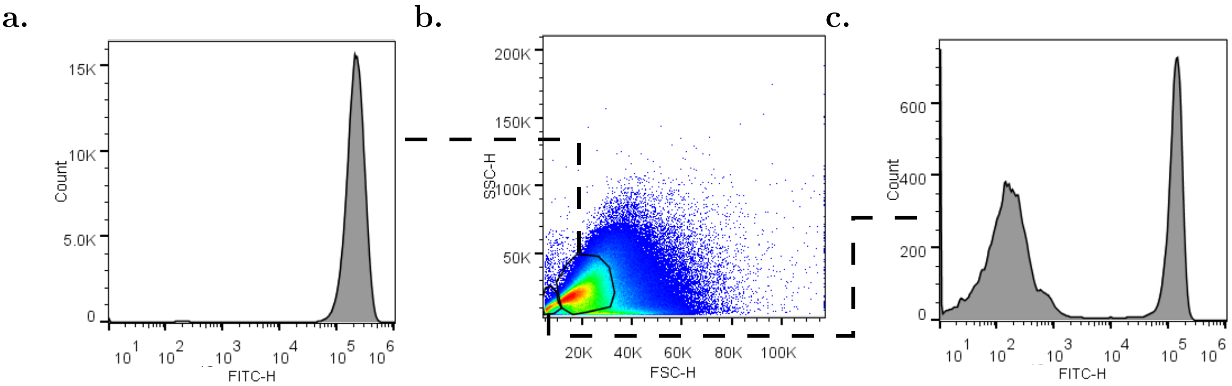


Figure A-4: **a.** FITC-H (greeen) fluorescent profile of controllers with stressed morphology when induced with 100 μM IPTG and 50 nM 3-O-C6 HSL. **b.** Density plot of events in the FSC-H,SSC-H plane. Red areas represent zones of the plane with more events. The black polygons drawn are the gate used to discriminate healthy and stressed cells. **c.** FITC-H (greeen) fluorescent profile of controllers with healthy morphology when induced with 100 μM IPTG and 50 nM 3-O-C6 HSL.

Gene Name	Sequence (5’→3’)
RFP	GATAAGTCCCTAACTTTTACAGCTAGCTCAGTCC AGGTATTATGCTAGCCTGAAGCTGTCACCGGATG TGCTTTCCGGTCTGATGAGTCCGTGAGGACGAAA CAGCCTCTACAAATAATTTTGTTTAATACTAGAG AAAGAGGGGAAATACTAGatggtttccaagggcg aggaggataacatggctatcattaaagagttcat gcgcttcaaagttcacatggagggttctgttaac ggtcacgagttcgagatcgaaggcgaaggcgagg gccgtccgtatgaaggcacccagaccgccaaact gaaagtgactaaaggcggcccgctgccttttgcg tgggacatcctgagcccgcaatttatgtacggtt ctaaagcgtatgttaaacacccagcggatatccc ggactatctgaagctgtcttttccggaaggtttc aagtgggaacgcgtaatgaattttgaagatggtg gtgtcgtgaccgtcactcaggactcctccctgca ggatggcgagttcatctataaagttaaactgcgt ggtactaatttcccatctgatggcccggtgatgc agaaaaagacgatgggttgggaggcgtctagcga acgcatgtatccggaagatggtgcgctgaaaggc gaaattaaacagcgcctgaaactgaaagatggcg gccattatgacgctgaagtgaaaaccacgtacaa agccaagaaacctgtgcagctgcctggcgcgtac aatgtgaatattaaactggacatcacctctcata atgaagattatacgatcgtagagcaatatgagcg cgcggagggtcgtcattctaccggtggcatggat gagctgtacaaataaCTCGGTACCAAATTCCAGA AAAGAGGCCTCCCGAAAGGGGGGCCTTTTTTCGT TTTGGTCC
luxI-GFP fusion	gtattgtctatgcctattaatgaacagtttaaaaaa gcagtcttaaatgcagcgaacgacgaaaattACGCC CTTGCAGCGggcggtggctccggcggtggtagtggt ggcatgcgtaaaggagaagaacttttcactggagtt gtcccaattcttgttgaattagatggtgatgttaat gggcacaaattttctgtcagtggagagggtgaaggt gatgcaacatacggaaaacttacccttaaatttatt tgcactactggaaaactacctgttccatggccaaca cttgtcactactttcggttatggtgttcaatgcttt gcgagatacccagatcatatgaaacagcatgacttt ttcaagagtgccatgcccgaaggttatgtacaggaa agaactatatttttc

Table A-2: Names and sequences of the gene commercially synthesiszed for this study.

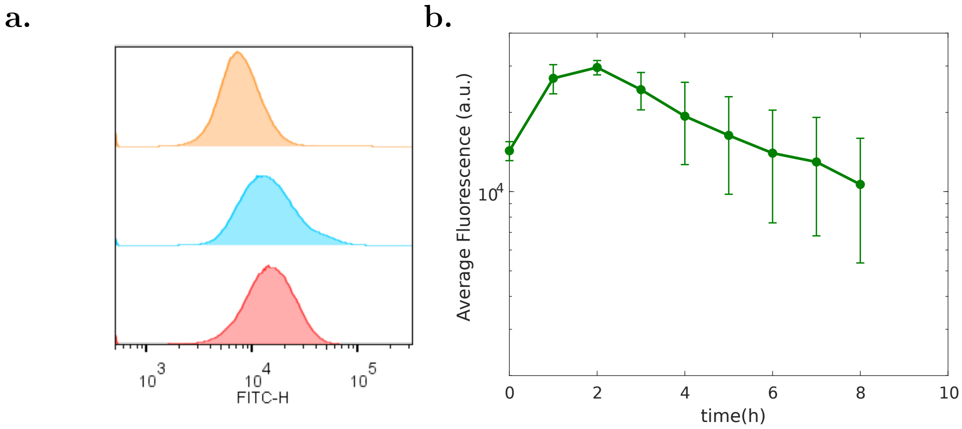


Figure A-5: **a.** Comparison of the FITC-H fluorescent profiles between uninduced targets cultured overnight (red) and for 6h (orange) with targets induced with 1 μM 3-O-C12 for 3h. **b.** Time evolution of an overnight target culture in Chi.Bio. The vertical lines represent the standard deviation between samples over $n = 2$ biological replicas.

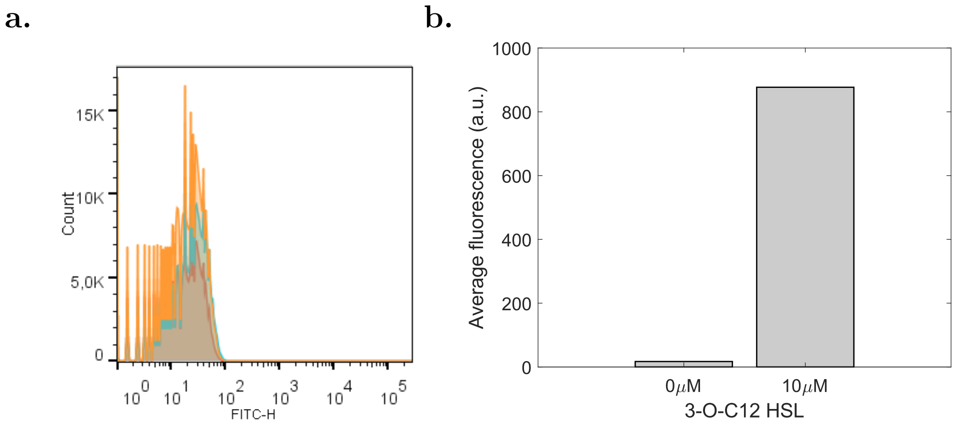


Figure A-6: **a.** Comparison of the FITC-H fluorescent profiles between uninduced targets cultured for 3h (red) and 6h (orange) with the fluorescent profile of cells not expressing GFP. **b.** Steady state average fluorescence of targets uninduced and induced with 10 μM 3-O-C12 HSL. The fluorescence was collected after 6h incubation in LB.

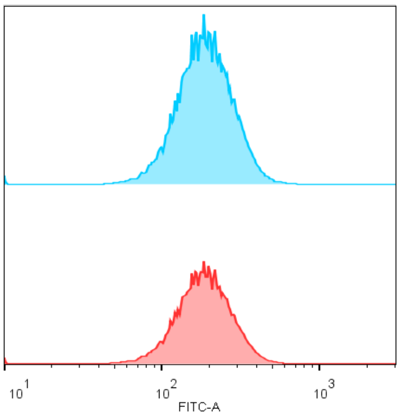


Figure A-7: Comparison of the FITC-H fluorescent profiles between uninduced targets and the steady state profile of targets in Closed Loop when no IPTG was present in the growth environment.

A.2 Supplementary material for the development of a Multicellular PI control architecture

A.2.1 Derivation of the reduced order models

Under the conditions presented in Section 5.2.2, the dynamics of the microbial consortium comprising three populations and described by equations (5-19)-(5-25) can be further simplified making a quasi-steady state assumption on the dynamics of the quorum sensing molecules, that is, by imposing $\dot{Q}_k^j = 0$, with $k \in \{u, x\}$ and $j \in \{t, p, i, e\}$. Under Assumption 1, substituting their steady-state values in $\dot{Q}_x^e = 0$ and $\dot{Q}_u^e = 0$ we obtain that:

$$\begin{aligned} Q_u^e &= \tfrac{1}{\Gamma_{PI}} \cdot \left(\beta_P Y_\mathrm{d} \tfrac{\mu_P Y_\mathrm{d}}{\mu_P Y_\mathrm{d} + \theta_P Q_x^p} + \beta_I Z_1 \right), \\ Q_x^e &= \tfrac{\beta_x X_c}{\Gamma_{PI}}. \end{aligned} \tag{A-1}$$

Then, by substituting (A-1) in $\dot{Q}_k^j = 0$, where $k \in \{x, u\}$ and $j \in \{p, i, t\}$, and leveraging again that $\eta \gg \Gamma_{PI}$, we obtain $Q_u^t \approx Q_u^p \approx Q_u^i \approx Q_u^e$, and $Q_x^t \approx Q_x^p \approx Q_x^i \approx Q_x^e$. We then define those two quantities as Q_u and Q_x and substitute their steady state values in equations (5-19) and (5-22).

A.2.2 Nominal biochemical parameters

The nominal biochemical parameters used in the BSim simulations are chosen as: $\beta_u = 0.06\,\mathrm{min}^{-1}$, $\beta_x = 0.03\,\mathrm{min}^{-1}$, $\gamma = 0.023\,\mathrm{min}^{-1}$, $\eta = 2\,\mathrm{min}^{-1}$ (taken from [5]); $\beta_c = 0.1\,\mathrm{min}^{-1}$, $\mu = 1\,\mathrm{min}^{-1}$, $\theta = 0.3\,\mathrm{min}^{-1}$, $\gamma_z = 0.01\,\mathrm{nM}^{-1}\mathrm{min}^{-1}$ (taken from [9]).

A.3 Supplementary material for the development ratiometric control

A.3.1 Memory-like property

We assume that the macroscopic behaviour of any cell in the consortium we wish to control can be modelled in the domain of interest by a dynamical system of the form

$$\dot{z} = g(z, w), \tag{A-2}$$

with $g : \mathcal{Z} \times \mathcal{W} \mapsto \mathcal{Z}$ being a smooth vector field, $z \in \mathcal{Z} \subset \mathbb{R}^d$ the state variables and $w \in \mathcal{W} \subset \mathbb{R}^m$ the exogenous input variables, representing, for instance, the concentrations of chemical species inside the cell and those of control inducer molecules into the environment, respectively. We assume that stochastic effects, such as fluctuations due to biochemical reactions, do not significantly alter the behaviour of the system at steady state; that is, the region of attraction of any stable equilibrium point of the dynamical system $\dot{z} = g(z, w)$ is large enough so that stochastic noise does not cause undesired switch from one equilibrium point to another.

As discussed in the Introduction and in Section 6.1.1, we are interested to robustly regulate the behaviour of reversible cells, in particular we focus our attention to a specific class of cells whose dynamics satisfy the following fundamental property.

Definition 1 *Consider a dynamical system of the form*

$$\dot{z} = f(z, \alpha), \tag{A-3}$$

where $z \in \mathcal{Z}$ *and the parameter* $\alpha \in \mathcal{I} \subset \mathbb{R}$ *depends on some exogenous input signal* $w \in \mathcal{W}$*, that is,* $\alpha = \alpha(w)$*. We say that system* (A-3) *is* memory-like *if:*

- *There exists some* $\bar{\alpha}$ *such that the system* $\dot{z} = f(z, \bar{\alpha})$ *has two stable equilibria and an unstable equilibrium. Furthermore, the regions of attraction of the two stable equilibria form a partition of* $\mathcal{Z}$*.*
- *There exist two values* $\hat{\alpha}_1$ *and* $\hat{\alpha}_2$ *such that for* $\alpha < \hat{\alpha}_1 \cup \alpha > \hat{\alpha}_2$ *the system* $\dot{z} = f(z, \bar{\alpha})$ *has a single equilibrium point whose region of attraction is the whole* $\mathcal{Z}$*.*

A.3.2 Event-driven modelling of the control error evolution

Here we derive an event-driven model for the evolution of the control error $e(t) = r_\mathrm{d} - r(t)$, presented in Section 6.1.2. Recall that the finite set of all cells in the consortium and its cardinality at time t are denoted by $\mathcal{N}_t$ and $N(t)$, respectively. Now, we denote with $E^i_{A\to B}(t')$ the event at time instant $t = t'$, corresponding to when the state x_i of cell $i \in \mathcal{N}_t$ enters the region of attraction of the equilibrium point B_i, that is, $x_i(t) \in \mathcal{R}_{B_i}$ for $t \in [t', t' + \varepsilon]$, and $x_i(t) \notin \mathcal{R}_{B_i}$ for $t \in [t' - \varepsilon, t')$, where ε is a small positive real number. Likewise, we denote by $E^i_{B\to A}(t'')$ the event at time instant $t = t''$, corresponding to when the state of cell $i \in \mathcal{N}_t$ enters the region of attraction of A_i. Specifically, for solutions to dynamical system (6-1) with $u = 0$ and $\eta_i > 0$, an event $E^i_{A\to B}$ occurs when the state of cell i, x_i, crosses zero and becomes positive, while an event $E^i_{B\to A}$ occurs when x_i becomes negative. In this case the threshold at zero is defined by the unstable equilibrium point at the origin dividing the regions of attraction of A_i and B_i. Moreover, we denote by $\mathcal{E}_{A\to B}$ ($\mathcal{E}_{B\to A}$) the sets of all events $E^i_{A\to B}$ ($E^i_{B\to A}$) occurring for all i at any time t, and we denote by $\mathcal{E}$ the set of all events occurring in the population, that is, $\mathcal{E} := \mathcal{E}_{A\to B} \cup \mathcal{E}_{B\to A}$.

To derive the discrete (event-driven) dynamics of the control error $e(t)$, we make the following standing assumptions:

Assumption 3 *At any time t only one event in $\mathcal{E}$ can occur, that is, there exists a unique $i \in \mathcal{N}_t$ such that either $E^i_{A\to B}$ or $E^i_{B\to A}$, but not both, occurs at time t.*

Assumption 4 *The number of cells in the host chamber is assumed to be constant, that is, $N(t) = N$, for all t.*

Assumption 3 implies that any two events in $\mathcal{E}$ cannot occur simultaneously. Assumption 4 follows from the fact that, after a short transient from the beginning of the experiment, cells grow occupying the entire host chamber. From this time on, cells exceeding the maximum capacity of the chamber are flushed out. Therefore, the number of cells in the chamber can be assumed with a good approximation to be constant (except for a small, negligible oscillation due to flush-out, further discussed in Section A.3.9). Note that Assumption 4 is true not only in case of a finite dimensional microfluidic device. Indeed, in industrial applications where chemostats are employed to maintain the density of the microbial culture to constant levels, it is reasonable to assume that also the number of cells is constant. However, in the case of large-scale cell co-cultures such as when the experiments are run in bioreactors, since measuring the fluorescence levels of each individual in the consortium can become not feasible as the number of cells increases the method we propose will need to be adapted, for example by measuring the average fluorescence in the whole population.

Then, there exists a sequence of discrete time instants $\{t_k\}_{k\in\mathbb{N}}$, each one corresponding to the occurrence of an event in $\mathcal{E}$ and such that $t_{k+1} = t_k + \Delta t_k$, where $\Delta t_k > 0$ is the time interval between two consecutive events occurring at time t_k and t_{k+1}. respectively. Moreover, from Assumption 4 it follows that the functions $n_A(t)$ and $n_B(t)$, defining the number of cells converging to either A_i or B_i, respectively, are piece-wise constant functions, that is, $n_A(t) = n_A(t_k)$, and $n_B(t) = n_B(t_k)$, $\forall t \in [t_k, t_{k+1})$. Since $n_A(t)$ and $n_B(t)$ are constrained by the relation $n_A(t) + n_B(t) = N$, for all t, for the sake of brevity, we will refer to $n(t) := n_B(t)$ only; $n_A(t)$ being given by $N - n_B(t)$.

That said, being $n(t)$ the number of cells in the region of attraction of B_i at time t, we can write the following discrete-time update law:

$$n(t_{k+1}) = \begin{cases} n(t_k) + 1, & \text{if an event in } \mathcal{E}_{A\to B} \text{ occurs} \\ n(t_k) - 1, & \text{if an event in } \mathcal{E}_{B\to A} \text{ occurs} \end{cases} \tag{A-4}$$

As a consequence, since $e(t) = r_\mathrm{d} - \frac{n(t)}{N}$, we have that when an event in $\mathcal{E}_{A\to B}$ occurs, then

$$e(t_{k+1}) = r_\mathrm{d} - \frac{n(t_{k+1})}{N} = \left(r_\mathrm{d} - \frac{n(t_k)}{N}\right) - \frac{1}{N} = e(t_k) - \frac{1}{N}.$$

A similar reasoning holds when an event in $\mathcal{E}_{B\to A}$ occurs. Therefore, the discrete-time dynamics of the control error can be expressed as:

$$e(t_{k+1}) = \begin{cases} e(t_k) - \frac{1}{N}, & \text{if an event in } \mathcal{E}_{A\to B} \text{ occurs} \\ e(t_k) + \frac{1}{N}, & \text{if an event in } \mathcal{E}_{B\to A} \text{ occurs} \end{cases} \tag{A-5}$$

A.3.3 Implementation details of the proposed feedback control algorithms and error dynamics

Here we discuss the design of the control strategies proposed to solve the ratiometric control problem, namely, the relay controller and the PI controller.

A.3.4 Relay control algorithm.

A relay controller is a feedback control law that generates a piece-wise constant input $u_{\mathrm{r}}(t)$ by comparing an output measured from the plant to some desired reference value r_{d}. The input $u_{\mathrm{r}}(t)$ takes value from a finite set of real values $\mathcal{U}_{\mathrm{r}}$, generally composed by only two values, one chosen such that the control error $e(t) = r_{\mathrm{d}} - r(t)$ decreases when $e(t) > 0$, and the other such that it increases when $e(t) < 0$.

We have considered two implementations of the relay controller. Specifically, the first implementation includes a control shutdown condition when $e(t) = 0$, and the second one does not.

Formally, the *first* implementation of the relay control input is defined as:

$$u'_{\mathrm{r}}(t) = \begin{cases} \bar{u}, & e(t) > 0 \\ 0, & e(t) = 0 \\ -\bar{u}, & e(t) < 0 \end{cases} \tag{A-6}$$

The value $\bar{u} > 0$ is chosen such that (ideally) all cells are controllable, that is, for $u'_{\mathrm{r}} = \bar{u}$ ($u'_{\mathrm{r}} = -\bar{u}$), the equation $\eta_i x_i - x_i^3 + u'_{\mathrm{r}} = 0$ has a unique stable solution, namely $\bar{x}_i = B_i(\bar{u})$ ($\bar{x}_i = A_i(\bar{u})$), for all i such that $\eta_i > 0$. This guarantees that when $e(t) > 0$ (i.e., $r_B(t) < r_{\mathrm{d}}$), the next event occurring must belong to $\mathcal{E}_{A\to B}$, forcing the error to decrease, that is, $e(t_{k+1}) = e(t_k) - \frac{1}{N}$. Likewise, when $e(t) < 0$ (i.e., $r_B(t) > r_{\mathrm{d}}$), $u'_{\mathrm{r}} = -\bar{u}$ implies that the next event occurring belongs to $\mathcal{E}_{B\to A}$, and so $e(t_{k+1}) = e(t_k) + \frac{1}{N}$. Moreover, notice that, when no control is applied (i.e., $u'_{\mathrm{r}} = 0$), each cell will converge to either A_i or B_i, depending on its current state, without any other event in $\mathcal{E}$ having to occur. Therefore, the shutdown condition ensures that, if there exists a t^* such that $e(t^*) = 0$, then $e(t) = 0$ for all $t \geq t^*$.

Combining (A-6) and (A-5), the discrete-time update law for the control error becomes:

$$e(t_{k+1}) = \begin{cases} e(t_k) - \frac{1}{N}, & \text{if } e(t_k) > 0 \\ e(t_k), & \text{if } e(t_k) = 0 \\ e(t_k) + \frac{1}{N}, & \text{if } e(t_k) < 0 \end{cases} \tag{A-7}$$

The previous discrete map can also be rewritten as $e(t_{k+1}) = e(t_k) - \frac{1}{N}\mathrm{sgn}(e(t_k))$.

The *second* implementation of the relay control input, without the shutdown condition, is defined as:

$$u''_{\mathrm{r}}(t) = \begin{cases} \bar{u}, & e(t) \geq 0 \\ -\bar{u}, & e(t) < 0 \end{cases} \tag{A-8}$$

In this case it is not possible to ensure that the error remains equal to zero indefinitely. By using a similar reasoning as before, (A-5) can be recast as:

$$e(t_{k+1}) = \begin{cases} e(t_k) - \frac{1}{N}, & \text{if } e(t_k) > 0 \\ -\frac{1}{N}, & \text{if } e(t_k) = 0 \\ e(t_k) + \frac{1}{N}, & \text{if } e(t_k) < 0 \end{cases} \tag{A-9}$$

A.3.5 PI control algorithm.

The control input of the PI controller is defined as:

$$u_{\mathrm{PI}}(t) = k_{\mathrm{P}} e(t) + k_{\mathrm{I}} z(t), \tag{A-10}$$

$$\dot{z}(t) = e(t), \quad z(0) = 0, \tag{A-11}$$

with k_{P} and k_{I} being positive constants. This control action is complemented with a (anti-windup) reset condition that sets the internal state z of the integrator to zero whenever the error becomes 0 or changes its sign. Furthermore, to take into account constraints on the actuation system of the experimental platform, the control input signal u_{PI} is assumed to be saturated at $\bar{u}$ and $-\bar{u}$.

The control algorithm guarantees that when $e(t) > 0$ the control input $u_{\mathrm{PI}}(t)$ is positive and $\frac{d}{dt}u_{\mathrm{PI}}(t) > 0$. So, $u_{\mathrm{PI}}(t)$ will increase and reach some positive value $\hat{u}$ such that, for at least one cell, the equation $\eta_i x_i - x_i^3 + u = 0$ with $u = \hat{u}$ and $\eta_i > 0$ has a unique solution, namely $B_i(\hat{u})$. The cell will be attracted by this stable equilibrium point and, therefore, there will exist a time instant t' such that, for all $t \geq t'$, $u_{\mathrm{PI}}(t) \geq \hat{u}$ and an event in $\mathcal{E}_{A\to B}$ will occur. A similar reasoning holds in the case $e(t) < 0$. Hence, it directly follows that the discrete-time update law for the control error $e(t)$ under the PI control law is the same as in (A-7).

A.3.6 Derivation of $\hat{\eta}$

In this section we show how the expression of $\hat{\eta}$ in (6-2) was derived. The quantity $\hat{\eta}$ represents the limit value of η_i such that the equation $\dot{x}_i = 0$ with $\eta_i > 0$ and $u = \bar{u}$, where $\bar{u} = \max_{u \in \mathcal{U}} |u|$, has a unique solution. That is, any cell whose behaviour can be described by (6-1) with $\eta_i > \hat{\eta}$, cannot be switched from A_i to B_i, or vice versa, by any input value $u \in \mathcal{U}$, because it always admits at least two solutions and it is therefore uncontrollable.

To this aim, we have to characterise the bifurcation points at which two equilibrium points of (6-1) coalesce and disappear [172], under variations of the control input u and of the parameters η_i. The position of the limit points can be easily obtained as:

$$(x_{LP}^{+}, u_{LP}^{+}) = \left(\sqrt{\frac{\eta_i}{3}}, -\sqrt{\frac{4\eta_i^3}{27}} \right) \tag{A-12}$$

$$(x_{LP}^{-}, u_{LP}^{-}) = -(x_{LP}^{+}, u_{LP}^{+}) \tag{A-13}$$

Hence, equation $\dot{x}_i = 0$ admits at least two solutions if its parameters and the control input satisfy the relationship

$$-\sqrt{\frac{4\eta_i^3}{27}} < u < \sqrt{\frac{4\eta_i^3}{27}} \tag{A-14}$$

Therefore, it is easy to verify from (A-14) that, given the maximum value of the control input $\bar{u}$, the threshold value $\hat{\eta}$ over which a cell becomes uncontrollable is

$$\hat{\eta} := \left(\frac{27\bar{u}^2}{4} \right)^{\frac{1}{3}} \tag{A-15}$$

A.3.7 Proof of convergence of the proposed controllers

Here we prove that the feedback control algorithms we proposed ensure finite time convergence to the desired cell ratio and hence solve the ratiometric control problem stated in Section 6.1.2. To this aim we conduct a convergence analysis on the error map $e_{k+1} = f(e_k)$, where $f : D \to D$, with f being defined in (A-7) or (A-9), and D is the domain of definition of the error signal e_k. Specifically, we consider some positive definite scalar function $V : D \to \mathbb{R}$, and we prove convergence of e_k to some invariant set by assessing the negative definiteness of the increment function $\Delta V := V(f(e_k)) - V(e_k)$ for every point outside this invariant set.

Notice that the domain of definition D is *finite*, indeed, from the definition of the control error signal, that is, $e(t) := r_{\mathrm{d}} - r(t) = \frac{1}{N}(n_{\mathrm{d}} - n(t))$, and from the fact that $n_{\mathrm{d}} = N \cdot r_{\mathrm{d}}$

is an integer constant and $n(t)$ is integer-valued function taking values in the finite set $\{0, 1, 2, \dots, N\}$, we have that

$$D := \left\{ -1, \dots, -\frac{2}{N}, -\frac{1}{N}, 0, \frac{1}{N}, \frac{2}{N}, \dots, 1 \right\}. \tag{A-16}$$

Relay control algorithm (A-6). To prove convergence of the error signal e to 0, let us consider the function $V(e) = e^2$. Using the discrete-time map in (A-7), we have that

$$\Delta V = V_{k+1} - V_k = e_{k+1}^2 - e_k^2 = e_k^2 - \frac{2}{N}|e_k| + \frac{1}{N^2} - e_k^2 = -\frac{1}{N^2}(2N|e_k| - 1). \tag{A-17}$$

The increment ΔV is negative definite everywhere except in the points such that

$$|e_k| \leq 1/(2N), \tag{A-18}$$

therefore any solution whose initial condition is outside this set converges towards it. Since the only possible value of e_k in D satisfying (A-18) is 0, we can therefore conclude that the error signal e_k converges to 0.

Relay control algorithm (A-8). From (A-9) it directly follows that the error dynamics does not admit any fixed point. However, a period-2 cycle exists whose stability we are going to prove by considering the second iterate of the error map in (A-9):

$$e_{k+2} = \begin{cases} e_k - \frac{2}{N}, & \text{if } e_k \geq \frac{1}{N} \\ e_k, & \text{if } e_k \in \{0, -\frac{1}{N}\} \\ e_k + \frac{2}{N}, & \text{if } e_k \leq -\frac{2}{N} \end{cases} \tag{A-19}$$

The periodic orbit of the first iterate map (A-9) corresponds to the pair of fixed points located in 0 and $-1/N$ of the second iterate map (A-19).

Let us consider the function $V(e) = |e|$. By using the second iterated map in (A-19) and taking into account (A-16) we have:

$$V(e_{k+2}) = \begin{cases} e_k - \frac{2}{N}, & \text{if } e_k \geq \frac{2}{N} \\ \frac{1}{N}, & \text{if } e_k = \frac{1}{N} \\ -e_k, & \text{if } e_k \in \{0, -\frac{1}{N}\} \\ -e_k - \frac{2}{N}, & \text{if } e_k \leq -\frac{2}{N} \end{cases} \tag{A-20}$$

Hence, after some algebra we can write:

$$\Delta V = V(e_{k+2}) - V(e_k) = \begin{cases} -\frac{2}{N}, & \text{if } |e_k| \geq \frac{2}{N} \\ 0, & \text{if } e_k = \{0, -\frac{1}{N}, \frac{1}{N}\} \end{cases} \tag{A-21}$$

Therefore, we can conclude that all solutions converge to the set where $\Delta V = 0$, that is, such that $|e_k| \leq 1/N$, or, equivalently, to the set $\{-\frac{1}{N},\ 0,\ \frac{1}{N}\}$. Now, by using the second iterate map (A-19), we can easily verify that in the set $\{-\frac{1}{N},\ 0,\ \frac{1}{N}\}$ there are two fixed points, namely $-\frac{1}{N}$ and 0. Therefore, these two fixed points correspond to period-2 cycles for the first iterated map (A-9). More precisely, both points correspond to the same cycle $M = \{-\frac{1}{N},\ 0\}$, indeed, by applying (A-9) we have that $-\frac{1}{N} \mapsto 0$ and $0 \mapsto -\frac{1}{N}$. This proves that the error signal globally converges to the period-2 cycle M.

PI control algorithm (A-10)-(A-11). In this case the discrete-time map describing the dynamics of the control error is also defined by (A-7). Therefore, the proof of convergence for the PI controller is the same as that of the first implementation of the relay controller (A-6) presented earlier on.

Finite time convergence. We complete our proof that the regulation happens in finite time by showing that when $e(t) \neq 0$ the time interval Δt_k between consecutive events is finite. Assuming without loss of generality that the control error right after an event in $\mathcal{E}$ occurred to be positive, say $e(t_k^+) > 0$, there will exist a time instant $\bar{t} > t_k$ after that the control input u is such that, for at least one cell, the equation $\eta_i x_i - x_i^3 + u = 0$ with $\eta_i > 0$ has a unique positive solution, namely $B_i(u) > 0$. This solution is globally asymptotically stable for the solution $x_i(t)$, that is, $\lim_{t\to\infty} x_i(t) = B_i(u)$. Hence, since $x_i(t)$ is a continuous function in time, there will exist a finite time instant t_{k+1} such that $x_i(t_{k+1}^-) < 0$ and $x_i(t_{k+1}^+) > 0$, meaning that an event in $\mathcal{E}_{A\to B}$ occurred. A similar reasoning holds when $e(t_k) < 0$. Combining this result with the observation that the number of events needed to converge to the desired ratio is finite due to the finite cardinality of D, we can conclude that the time needed to reach steady state is finite.

A.3.8 Derivation of e_r^u

Here we illustrate how the expression of e_r^u in (6-7) was obtained. The quantity e_r^u is an upper bound on the residual steady-state error due to the presence of *unswitchable* cells, i.e., bistable cells that cannot be toggled by any admissible value of the control input u. The maximum value of this quantity can be estimated by considering the worst case scenarios in which all these uncontrollable cells are initially in one region of attraction and therefore cannot be moved to the other one by any admissible values of u.

Without loss of generality, assume that all uncontrollable cells are initially in the region of attraction of their equilibrium point B_i and recall that r_d denotes the desired ratio of cells converging to B_i at steady state. Denoting with r_u the ratio between the number of uncontrollable cells and the whole cell population, we have that if $r_u \leq r_d$, that is, the number of uncontrollable cells does not exceed the desired number in B_i, then uncontrollable cells do not cause any problem because we do not need to move any of them from where they are. Therefore, in this case, $e_\infty = 0$. On the other hand, if $r_u > r_\mathrm{d}$, the fraction of uncontrollable cells exceeding the desired ratio, that is, $r_u - r_\mathrm{d}$, will determine an error at steady state, that is, $|e_\infty| \leq r_u - \mathrm{r_d}$. A similar argument can be stated for the dual scenario in which all uncontrollable cells are initially in the region of attraction of A_i. Recalling that in this case we want the ratio of cells converging to A_i to be equal to $(1 - r_\mathrm{d})$, the residual error at steady state in this scenario will satisfy $|e_\infty| \leq r_u - (1 - \mathrm{r_d})$.

From the previous analysis and exploiting that, for a high number of cells, r_u can be estimated as the probability of η_i begin greater that $\hat{\eta}$, we finally obtain that

$$e_r^u = \max\left\{0, \Pr[\eta_i > \hat{\eta}] - \mathrm{r_d}, \Pr[\eta_i > \hat{\eta}] - (1 - \mathrm{r_d})\right\}. \tag{A-22}$$

where Pr denotes the probability measure.

A.3.9 Noise magnitude due to flush-out effects

The fluctuations in the steady-state response of the closed-loop system (see for example Figure **6-6**) are mostly due to cells flowing out of the microfluidic chamber as they grow. The magnitude ε of this noise is indeed proportional to the ratio between the flow rate Q of cells exiting the chamber and the total number N of cells therein. That is, $\varepsilon \propto Q/N$, where $Q = v \cdot A$; being A the cross-section area of the chamber apertures and v the flow velocity, that depends on the cell growth rate. Without loss of generality, assume that, as is usual in microfluidic applications, cells in the chamber are organised in a single-layer and that, on average, the area occupied by a rod-shaped *E. Coli* cell can be approximated by a rectangle of base w and height h. Moreover, let us assume that the chamber has a square geometry with side ℓ with two open sides, as in Figure **6-5**, and that all cells leave the chamber longitudinally. Then, the flow rate Q can be approximated as $Q \approx v \cdot \frac{2\ell}{w}$. Moreover, the total number of cells in the chamber is $N = \ell^2/(w \cdot h)$. Therefore, in conclusion, we have that

$$\varepsilon \propto \frac{Q}{N} = \frac{v \cdot \frac{2\ell}{w}}{\frac{\ell^2}{w \cdot h}} = 2\, v \cdot \frac{h}{\ell},$$

that is, the magnitude of the noise is inversely proportional to the size of the culture chamber. We validated the formula numerically by comparing the noise levels between chambers of different sizes (compare Figure **6-6** and Supplemental Figure **A-11**). Specifically, we observed that, as expected, the average variation of the ratios at steady state in the chamber $40\,\mu m \times 50\,\mu m$ wide (0.009) was three times lower than the one in a chamber where both the dimensions were scaled down to a third (0.03) (see also Supplemental Video 2).

Note that when the cells under observation flow out of the field of view from all the four sides of the chamber (a common situation when higher magnification factors are used on the microscope), the same result still holds but multiplied by a factor of 2.

A.3.10 Mathematical model of the inducible toggle-switch

In this section we present the details of the mathematical model used to describe the time evolution of proteins' concentration in the inducible toggle-switch. The model of the toggle-switch we consider here was originally developed and parameterised from experimental data in [125], and it was further analysed in [126, 127, 173, 143].

The model captures the pseudo-reactions describing transcription

$$\emptyset \xrightarrow{h_{\mathrm{L}}(TetR,aTc)} mRNA_{\mathrm{LacI}}, \tag{A-23}$$

$$\emptyset \xrightarrow{h_{\mathrm{T}}(LacI,IPTG)} mRNA_{\mathrm{TetR}}, \tag{A-24}$$

where

$$h_{\mathrm{L}}(TetR, aTc) := \kappa_{\mathrm{L}}^{\mathrm{m0}} + \frac{\kappa_{\mathrm{L}}^{\mathrm{m}}}{1 + \left(\frac{[TetR]}{\theta_{\mathrm{TetR}}} \cdot \frac{1}{1+\left(\frac{[aTc]}{\theta_{\mathrm{aTc}}}\right)^{\eta_{\mathrm{aTc}}}} \right)^{\eta_{\mathrm{TetR}}}}, \tag{A-25}$$

$$h_{\mathrm{T}}(LacI, IPTG) := \kappa_{\mathrm{T}}^{\mathrm{m0}} + \frac{\kappa_{\mathrm{T}}^{\mathrm{m}}}{1 + \left(\frac{[LacI]}{\theta_{\mathrm{LacI}}} \cdot \frac{1}{1+\left(\frac{[IPTG]}{\theta_{\mathrm{IPTG}}}\right)^{\eta_{\mathrm{IPTG}}}} \right)^{\eta_{\mathrm{LacI}}}}, \tag{A-26}$$

those describing translation

$$mRNA_{\mathrm{LacI}} \xrightarrow{\kappa_L^P} mRNA_{\mathrm{LacI}} + LacI, \tag{A-27}$$

$$mRNA_{\mathrm{TetR}} \xrightarrow{\kappa_T^P} mRNA_{\mathrm{TetR}} + TetR, \tag{A-28}$$

and those related to dilution/degradation

$$mRNA_{\mathrm{LacI}} \xrightarrow{g_L^m} \emptyset, \tag{A-29}$$

$$mRNA_{\mathrm{TetR}} \xrightarrow{g_T^m} \emptyset, \tag{A-30}$$

$$LacI \xrightarrow{g_L^p} \emptyset, \tag{A-31}$$

$$TetR \xrightarrow{g_L^p} \emptyset. \tag{A-32}$$

$\kappa_{\text{L}}^{\text{m0}}$	0.3045 counts min^{-1}	g_{L}^{p}	0.0165 min^{-1}
$\kappa_{\text{T}}^{\text{m0}}$	0.3313 counts min^{-1}	g_{T}^{p}	0.0165 min^{-1}
$\kappa_{\text{L}}^{\text{m}}$	13.01 counts min^{-1}	θ_{LacI}	124.9
$\kappa_{\text{T}}^{\text{m}}$	5.055 counts min^{-1}	η_{LacI}	2.00
$\kappa_{\text{L}}^{\text{p}}$	0.6606 a.u. $^{-1}$ min^{-1}	θ_{TetR}	76.40
$\kappa_{\text{T}}^{\text{p}}$	0.5098 a.u. $^{-1}$ min^{-1}	η_{TetR}	2.152
k_{aTc}	0.04 min^{-1}	θ_{aTc}	35.98
k_{IPTG}	0.04 min^{-1}	η_{aTc}	2.00
g_{L}^{m}	0.1386 min^{-1}	θ_{IPTG}	0.2926
g_{T}^{m}	0.1386 min^{-1}	η_{IPTG}	2.00
k_{ℓ}	0.12 min^{-1}		

Table A-3: Values of the parameters of the cell population models (taken from [125]). k_{ℓ} defines the rate at which the length of a bacterium grows in BSim, chosen as in [123].

The repressive action of LacI and TetR on each other is captured in (A-25)-(A-26) by means of Hill functions (with parameters θs and ηs). Likewise, the repressions of aTc on TetR and of IPTG on LacI are both described by another Hill function nested in the previous two. In the above pseudo-reactions, the parameters $\kappa_{\text{L/T}}^{\text{m0}}$, $\kappa_{\text{L/T}}^{\text{m}}$, $\kappa_{\text{L/T}}^{\text{p}}$, $g_{\text{L/T}}^{\text{m}}$, $g_{\text{L/T}}^{\text{p}}$ are basal transcription, transcription, translation, mRNA degradation and protein degradation rates, respectively. Note that all dilution rates were set based on estimates of cell cycle times. The value of these parameters are reported in Table **A-3**. The pseudo-reactions listed above can be used to obtain the following *deterministic model* of the inducible toggle-switch:

$$\frac{d}{dt}\,mRNA_{\text{LacI}} = \kappa_{\text{L}}^{\text{m0}} + \frac{\kappa_{\text{L}}^{\text{m}}}{1+\left(\frac{[TetR]}{\theta_{\text{TetR}}}\cdot\frac{1}{1+\left(\frac{[aTc]}{\theta_{\text{aTc}}}\right)^{\eta_{\text{aTc}}}}\right)^{\eta_{\text{TetR}}}} - g_{\text{L}}^{\text{m}}\,mRNA_{\text{LacI}} \tag{A-33}$$

$$\frac{d}{dt}\,mRNA_{\text{TetR}} = \kappa_{\text{T}}^{\text{m0}} + \frac{\kappa_{\text{T}}^{\text{m}}}{1+\left(\frac{[LacI]}{\theta_{\text{LacI}}}\cdot\frac{1}{1+\left(\frac{[IPTG]}{\theta_{\text{IPTG}}}\right)^{\eta_{\text{IPTG}}}}\right)^{\eta_{\text{LacI}}}} - g_{\text{T}}^{\text{m}}\,mRNA_{\text{TetR}} \tag{A-34}$$

$$\frac{d}{dt}\,[LacI] = \kappa_{\text{L}}^{\text{p}}\,mRNA_{\text{LacI}} - g_{\text{L}}^{\text{p}}\,[LacI] \tag{A-35}$$

$$\frac{d}{dt}\,[TetR] = \kappa_{\text{T}}^{\text{p}}\,mRNA_{\text{TetR}} - g_{\text{T}}^{\text{p}}\,[TetR] \tag{A-36}$$

where the quantities $mRNA_{\text{LacI}}, mRNA_{\text{TetR}}$ denote abundance of the chemical species in a cell while variables in the form $[\,\cdot\,]$ denote concentrations of molecules inside each cell (see [125] for further details about the mathematical model).

The model is complemented by also taking into account the diffusion of aTc and IPTG across the cell membrane:

$$\frac{d}{dt}\,[aTc] = k_{\text{aTc}}\,\left(u_{\text{aTc}} - [aTc]\right), \tag{A-37}$$

$$\frac{d}{dt}\,[IPTG] = k_{\text{IPTG}}\,\left(u_{\text{IPTG}} - [IPTG]\right), \tag{A-38}$$

where the parameters $k_{\text{aTc/IPTG}}$ are diffusion rates, whose values are reported in Table **A-3**.

A.3.11 Agent-based simulations in BSim

BSim is an agent-based simulator explicitly designed for the simulation of bacterial populations [122, 123]. In this environment the chemicals' spatial distribution and the bio-mechanics of the cells are also considered. Specifically, it is possible to mimic a microfluidic platform, where cells grow and move, and where chemicals diffuse into the environment.

The BSim implementation presented in [123] has been extended here to include also stochastic simulations of the biochemical processes taking place inside the cells. Specifically, we implemented the SDE-based algorithm presented in [174]. Formally, for each cell we solved a system of equations of the form

$$d\mathbf{x} = \mathbf{S} \cdot \mathbf{F}(\mathbf{x}, t)dt + \mathbf{S} \cdot \mathrm{diag}(\mathbf{F}(\mathbf{x}, t))d\mathbf{w} + \Delta(\mathbf{x})dt, \tag{A-39}$$

where $\mathbf{x} = \left[mRNA_{\mathrm{LacI}}, mRNA_{\mathrm{TetR}}, [LacI], [TetR], [aTc], [IPTG]\right]^\top$ is the state vector, $\mathbf{w}$ is a vector of independent standard Wiener processes, $\mathbf{S} = [\mathbf{s}_1, \ldots, \mathbf{s}_6]$ is the stoichiometry matrix and $\mathbf{F}(\mathbf{x}) = [a_1(\mathbf{x}), \ldots, a_6(\mathbf{x})]^\top$, where $a_j(\mathbf{x})$ is the propensity function of the j-th pseudo-reaction defined as in Section A.3.10. Notice that the term $\Delta(\mathbf{x})dt$ models the diffusion dynamics of chemicals across the cell membrane, assumed to be deterministic.

The microfluidic platform we considered consists of a microfluidic device, a microscopy, a computer and an actuation system. The microfluidic device contains a chamber where cells are trapped and grow. Specifically, we simulated the microfluidic chip described in [109], where the growth chambers' dimensions are $40\,\mu\mathrm{m} \times 30\,\mu\mathrm{m} \times 1\,\mu\mathrm{m}$. This allows bacteria to grow only in a mono-layer structure. The chamber is connected to two perfusion channels, which bring nutrients and inducers to the cellular population. The flow of the fluids sent to the chambers is governed by two syringes, each one containing a mixture of growth media and one inducer. By adjusting the relative height of the syringes, it is possible to change the amount of inducers delivered to the cells. The architecture is completed by a computer, controlling the syringes depending on the output of a control algorithm, and a microscopy, necessary to take images whose analysis results in the computation of the error.

The bio-mechanical parameters of the cells (i.e., cells' growth rate) have been set as in [123]. In addition, as also done in [5, 123], at each cell division we assumed the cell contents to be equally divided in the mother and daughter cell, resulting in both having the same state that the mother had the moment right before the division occurred.

Moreover, we have taken into account the following realistic constraints on the experimental platform [18] :

1. the state of the cells cannot be measured more often than 5 min, to avoid excessive photo-toxicity;
2. there is a time delay of 40 s on the actuation of the control inputs due to the time that the flow of the chemical inducers takes to reach the chambers in the microfluidic chip where cells are hosted;
3. the minimum time interval between two consecutive control inputs cannot be smaller than 15 min, to limit excessive osmotic stress on the cells;

4. the maximum duration of any experiment cannot exceed 24 hours (1440 min), to avoid substantial cell mutations during the experiments.

A.3.12 Realistic *in silico* implementation of the controllers

The specific implementation of the microfluidic device we adopted also introduces constraints on the possible classes of input signals $u(t) = [u_{\mathrm{aTc}}(t),\, u_{\mathrm{IPTG}}(t)]^\top$ that can be generated by the actuators. We consider two possible implementations:

1. a T-junction, which limits u_{aTc} and u_{IPTG} to be mutually exclusive and with fixed amplitudes, that is, u is either equals to $[U_{\mathrm{aTc}}, 0]^\top$, or to $[0, U_{\mathrm{IPTG}}]^\top$;

2. a Dial-A-Wave (DAW) system [144], which constraints u_{aTc} and u_{IPTG} to be in a convex combination. Namely, given $u_{\mathrm{aTc}} \in [0, U_{\mathrm{aTc}}]$ we have

$$u_{\mathrm{IPTG}} = \left(1 - \frac{u_{\mathrm{aTc}}}{U_{\mathrm{aTc}}}\right) U_{\mathrm{IPTG}} \tag{A-40}$$

In the above equations, U_{aTc} and U_{IPTG} denote the concentrations of the inducers in the reservoirs, and we assume that they can be varied at the beginning of the experiments in the intervals $[0, 100]$ and $[0, 1]$, respectively. The upper bounds for U_{aTc} and U_{IPTG} are selected to avoid excessive stress on cells and are the same as those used *in vivo* in [125].

Furthermore, to meet the specific dynamical features of the genetic toggle-switch implementing the required bistable memory mechanism in reversible differentiable cells (Section 6.1.4), we propose modified versions of the ideal relay controller and the PI controller proposed in Section 6.1.2.

When considering an inducible toggle-switch the boundaries of the regions of attraction for each stable equilibrium strongly depend on the parameters chosen for the model and, even knowing them with absolute precision for each cell in a deterministic setting, the presence of noise could cause cells to change region of attraction. Hence, it is not possible to define precise boundaries between the regions of attraction of the equilibrium points A_i and B_i for each cell. For this reason, we define conservative estimation of the regions of attraction common to all cells. Specifically, we defined $\mathcal{A}_t := \{i \in \mathcal{N}_t \,:\, 2\,LacI_i(t) < TetR_i(t)\}$ and $\mathcal{B}_t := \{i \in \mathcal{N}_t \,:\, LacI_i(t) > 2\,TetR_i(t)\}$, where $LacI_i$ and $TetR_i$ denote the concentrations of the molecules inside the i-th cell. Moreover, since in reality the union of these two sets does not cover the entire set $\mathcal{N}_t$ of cells in the consortium, cells not belonging to $\mathcal{A}_t$ and $\mathcal{B}_t$ are in an *uncertain* state. We define this uncertain set of cells as $\mathcal{C}_t := \{i \in \mathcal{N}_t : i \notin \mathcal{A}_t, i \notin \mathcal{B}_t\}$ and denote with $n_C(t) = |\mathcal{C}_t|$ its cardinality at time t. For this reason, the error signals e_A and e_B are not complementary (i.e., $e_A(t) \neq -e_B(t)$) and we need to consider both in the design of the control actions.

Relay controller. The relay controller consists of two mutually exclusive inputs with fixed amplitudes, which are applied to the system depending on the current value of the error signals $e_A(t)$ and $e_B(t)$. Specifically, at any time t, it is applied the input that causes

$\max\{|e_A(t)|, |e_B(t)|\}$ to decrease. More formally, the control input $u(t) = [u_{\mathrm{aTc}}(t),\, u_{\mathrm{IPTG}}(t)]^\top$ is chosen as

$$u(t) = \begin{cases} u_1, & |e_B(t)| \geq |e_A(t)| \\ u_2, & |e_B(t)| < |e_A(t)| \end{cases}, \tag{A-41}$$

where

$$u_1 = \begin{cases} [0, U_{\mathrm{IPTG}}]^\top, & e_B(t) \leq 0 \\ [U_{\mathrm{aTc}}, 0]^\top, & e_B(t) > 0 \end{cases}, \quad u_2 = \begin{cases} [U_{\mathrm{aTc}}, 0]^\top, & e_A(t) \leq 0 \\ [0, U_{\mathrm{IPTG}}]^\top, & e_A(t) > 0 \end{cases}. \tag{A-42}$$

In this way, when $e_B > 0$ ($e_B \leq 0$) a high value of aTc (IPTG) is applied, that, by promoting LacI (TetR), causes e_B to decrease (increase). Similar arguments hold for e_A.

PI controller. The PI controller is implemented using a Dial-A-Wave system [144]. It consists of two control actions, each acting to decrease either of the control errors. Formally, we have:

$$u_{\mathrm{aTc}}(t) = k_{\mathrm{P},1} e_B(t) + k_{\mathrm{I},1} \int_0^t e_B(\tau) d\tau - \left(k_{\mathrm{P},2} e_A(t) + k_{\mathrm{I},2} \int_0^t e_A(\tau) d\tau \right) \tag{A-43}$$

where $k_{\mathrm{P},j}$ and $k_{\mathrm{I},j}$, $j = 1, 2$, are the proportional and integral control gains. Note that, due to the constraints on the actuation system, u_{IPTG} is defined by (A-40).

Moreover, to improve performance, the control algorithm is complemented with an anti-windup scheme (that sets to zero the internal state of the integrator whenever the error signals are equal to 0 or change sign) and a dynamic saturation, empirically tuned, defined as:

$$\begin{cases} u_{\mathrm{aTc}} \in [0, 50], & \text{if } |e_B| < |e_A| \\ u_{\mathrm{aTc}} \in [0, 100], & \text{otherwise} \end{cases} \tag{A-44}$$

Note that the choice of u_{IPTG} follows from the above and Equation (A-40).

Tuning of the control parameters. The control parameters U_{aTc}, U_{IPTG}, and gains $k_{\mathrm{P},j}$ and $k_{\mathrm{I},j}$, $j = 1, 2$, in Equations (A-42) and (A-43) used have been empirically selected by means of preliminary *in silico* experiments in MATLAB on the closed-loop system consisting of 30 cells whose dynamics is described by (A-39). The same value of the control parameters have been then used in all the experiments conducted in BSim. From an experimental viewpoint, a better tuning of the control gains might be possible once the model is parameterised from real data so as to provide an *in silico* benchmark better representative of the *in vivo* implementation of the controller under investigation. The numerical values used in each simulation are reported in the caption of the relative figure.

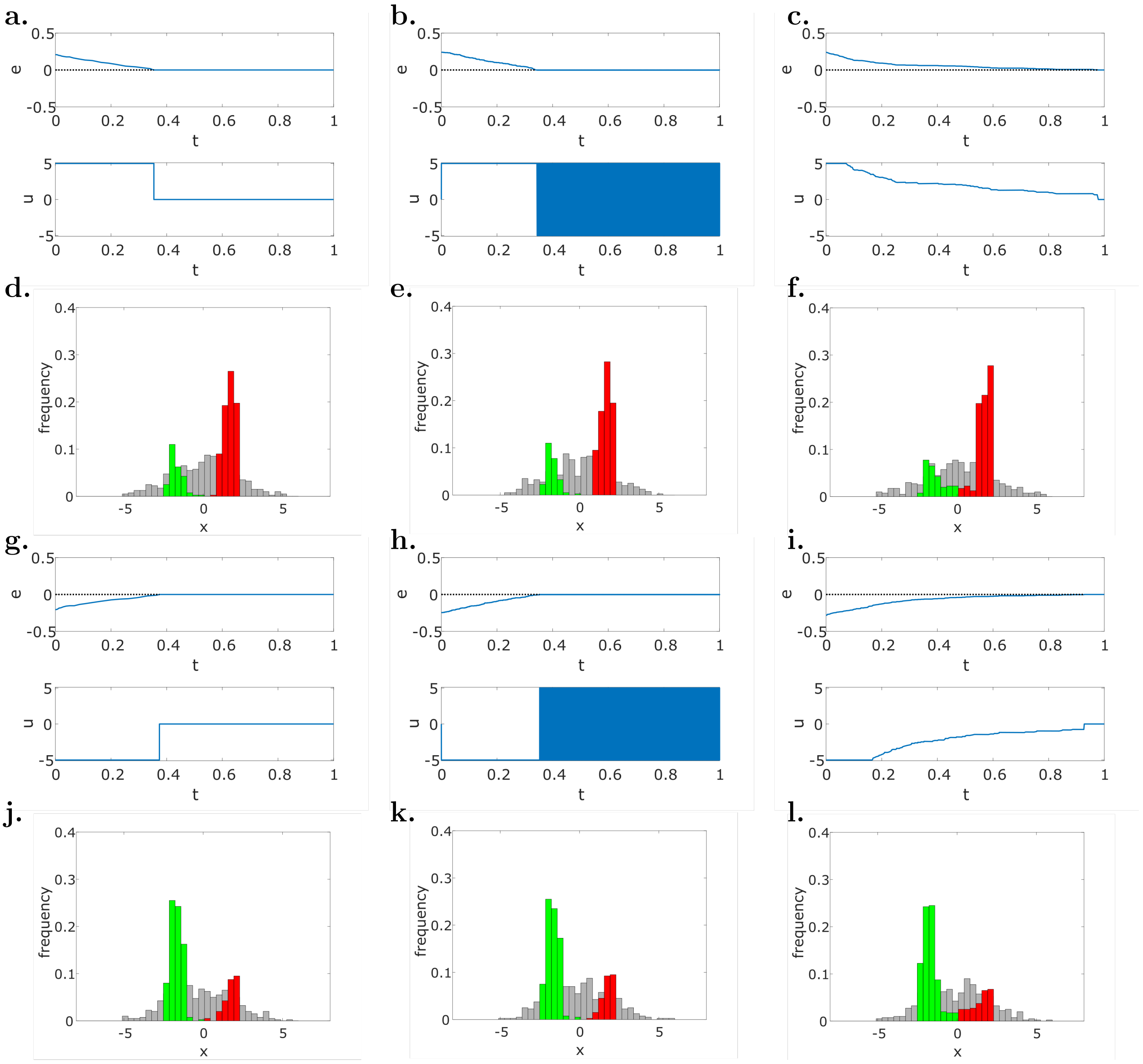


Figure A-8: **Feedback control strategies are effective to balance two *asymmetric* groups of controllable reversible cells to 1:3 ratio ($r_\mathrm{d} = 0.75$) and 3:1 ratio ($r_\mathrm{d} = 0.25$).** (**a-c**, **g-i**) Evolution of the error signal $e(t)$ and of the control input $u(t)$ for (**a**, **g**) the first implementation of the relay controller (A-6), (**b**, **h**) the second implementation of the relay controller (A-8), (**c**, **i**) the PI controller (A-10)-(A-11). (**d-f**, **j-l**) Distribution of the cells state at the beginning of the simulation ($t = 0\,$a.u., grey histogram) and at steady state ($t = 1.0\,$a.u., green and red histograms), for (**d**, **j**) the first implementation of the relay controller, (**e**, **k**) the second implementation of the relay controller, (**f**, **l**) the PI controller. Desired ratios are set to (**a-f**) $r_\mathrm{d} = 0.75$ (1:3 ratio) and (**g-l**) $r_\mathrm{d} = 0.25$ (3:1 ratio). The green and red bars in panels **d-f** and **j-l** correspond to cells in the region of attraction of A_i and B_i, respectively. The maximum control input is set to $\bar{u} = 5$ and the gains of the PI controller has been set to $k_\mathrm{P} = 30$ and $k_\mathrm{I} = 10$. All cells ($N = 400$) have initial conditions $x_i(0)$ drawn from the normal random distribution $\mathcal{N}(0, 4)$, and the parameters η_i are drawn with uniform distribution from the interval $[1, 5]$, therefore all cells are controllable, as no monostable ($\underline{\eta} > 0$) and no unswitchable ($\overline{\eta} < \hat{\eta} \approx 5.53$) cells are present in the population.

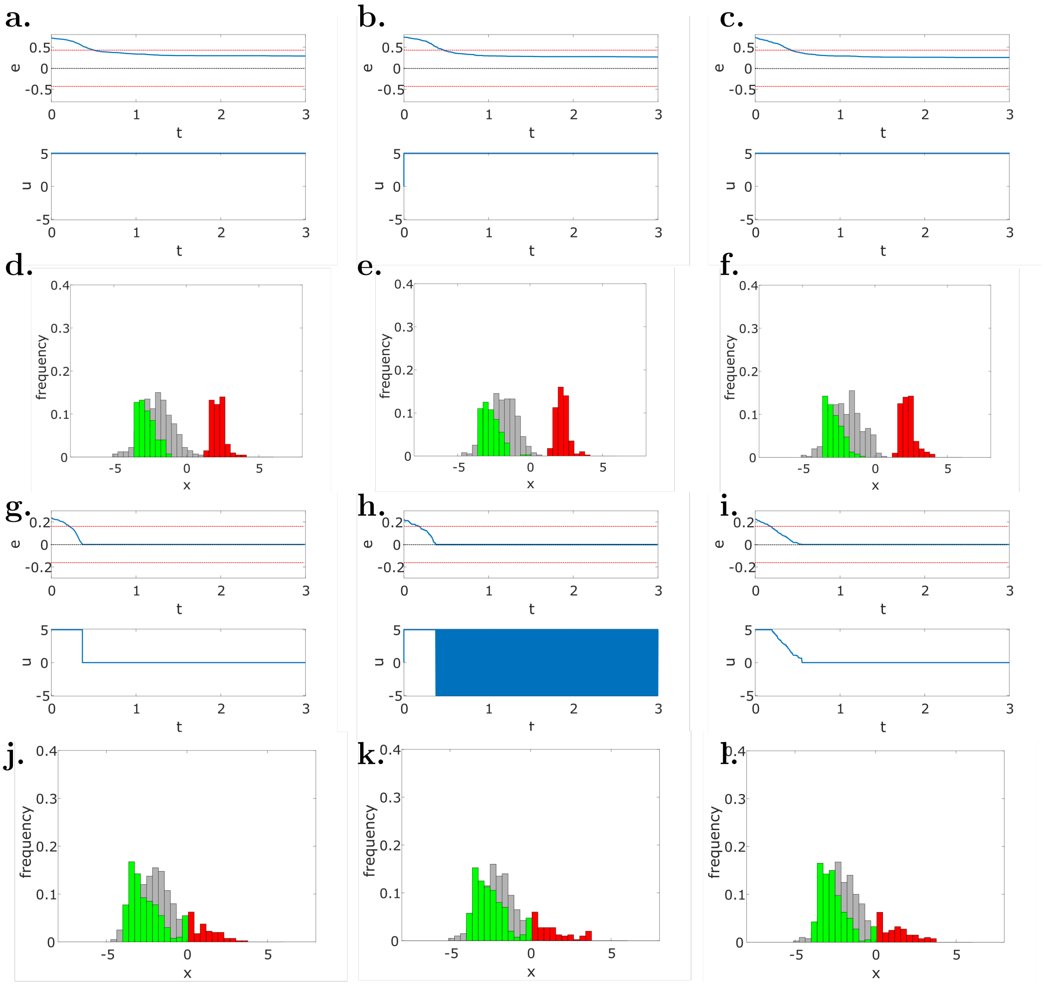


Figure A-9: **Balance to *asymmetric* ratios is achieved with a residual steady-state error in the presence of uncontrollable cells.** (**a-c**, **g-i**) Evolution of the error signal $e(t)$ and of the control input $u(t)$ for (**a**, **g**) the first implementation of the relay controller (A-6), (**b**, **h**) the second implementation of the relay controller (A-8), (**c**, **i**) the PI controller (A-10)-(A-11). (**d-f**, **j-l**) Distribution of the cells state at the beginning of the simulation ($t = 0\,\text{a.u.}$, grey histogram) and at steady state ($t = 3.0\,\text{a.u.}$, green and red histograms), for (**d**, **j**) the first implementation of the relay controller, (**e**, **k**) the second implementation of the relay controller, (**f**, **l**) the PI controller. Desired ratios are set to (**a-f**) $r_\text{d} = 0.75$ (1:3 ratio) and (**g-l**) $r_\text{d} = 0.25$ (3:1 ratio). The green and red bars in panels **d-f** and **j-l** correspond to cells in the region of attraction of A_i and B_i, respectively. The maximum control input is set to $\bar{u} = 5$ and the gains of the PI controller are set to $k_\text{P} = 30$ and $k_\text{I} = 10$. All cells ($N = 400$) have initial conditions $x_i(0)$ drawn from the normal random distribution $\mathcal{N}(-2, 1)$, and the parameters η_i are drawn with uniform distribution from the interval $[-1, 14]$, therefore both monostable ($\underline{\eta} < 0$) and unswitchable ($\overline{\eta} > \hat{\eta} \approx 5.53$) cells can be present in the population. The steady-state errors observed in the *in silico* experiment are equal to **(a)** 0.265, **(b)** 0.235, **(c)** 0.265, and **(g-i)** 0. Note that all the observed errors are below the theoretical upper bound on the control error estimated using (6-5) as $e_r = e_r^0 + e_r^u \approx 0.07 + 0.31 = 0.38$. (depicted in the panels **a-c** and **g-i** as red dashed lines). The lower steady-state error observer in panels **g-i** with respect to that in panels **a-c** is due to the initial conditions of the uncontrollable cells, already exhibiting the right phenotype and therefore they are already belonging to the set they should be in, and do not affect the steady-state error.

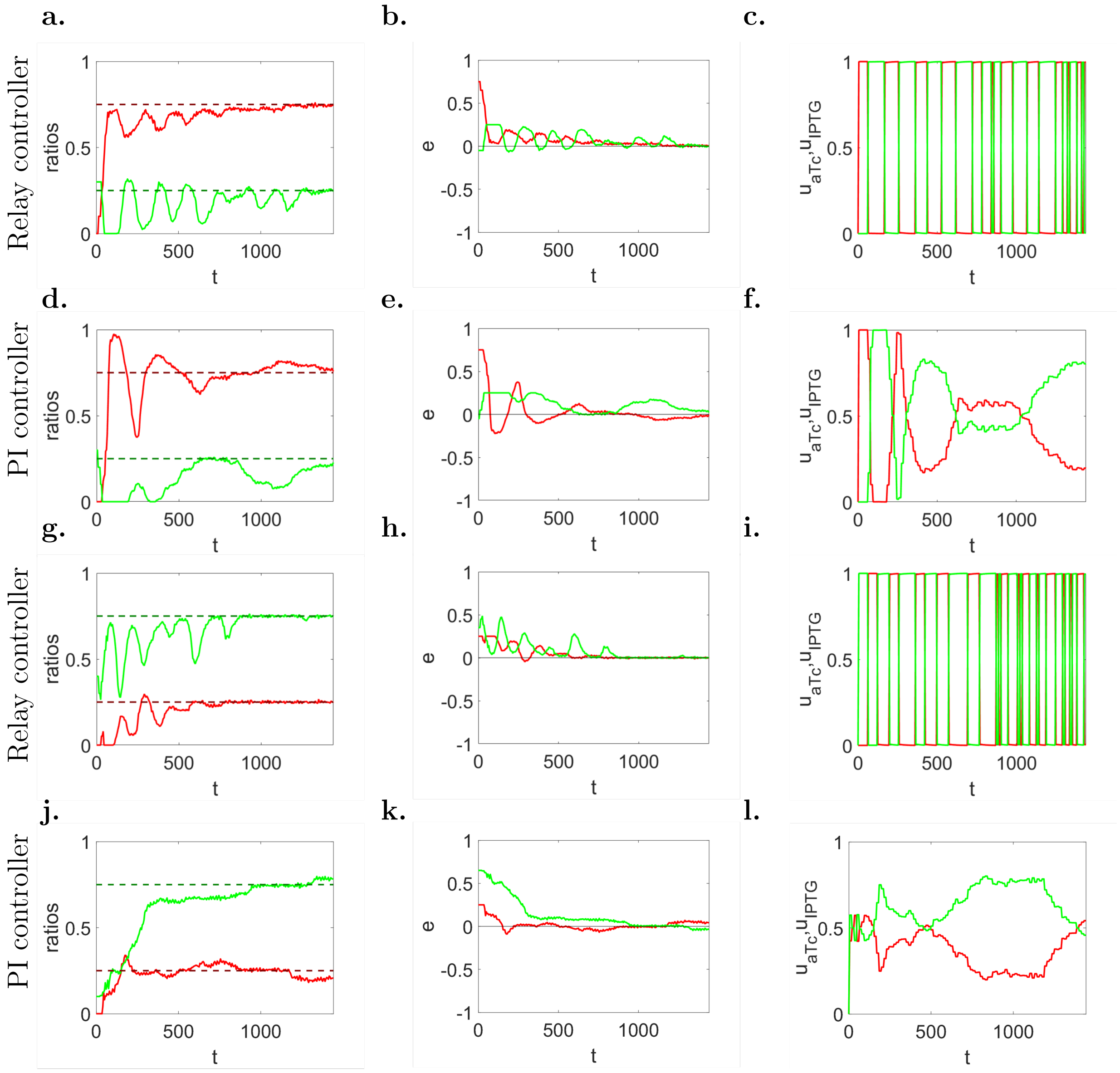


Figure A-10: **Cooperative production of two monomers to *asymmetric* population ratios can be achieved by means of feedback ratiometric controllers in microfluidics.** (**a, d, g, j**) Evolution in time of populations' ratio r_A (solid green line) and r_B (solid red line) with their respective desired reference values in dashed lines, (**b, e, h, k**) of the error signals e_A (solid green line) and e_B (solid red line), and (**c, f, i, l**) inducer control signals u_{aTc} (solid red line) and u_{IPTG} (solid green line), normalised to their maximum values U_{aTc} and U_{IPTG}, respectively. Desired ratios are set to (**a-f**) $r_{\mathrm{d}} = 0.75$ (1:3 ratio) and (**g-l**) $r_{\mathrm{d}} = 0.25$ (3:1 ratio). (**a-c, g-i**) Parameters of the relay control (A-41): $U_{\mathrm{aTc}} = 60\,\mathrm{ng/mL}$, $U_{\mathrm{IPTG}} = 0.5\,\mathrm{mM}$. (**d-f, j-l**) Parameters of the PI controller (A-43): $U_{\mathrm{aTc}} = 60\,\mathrm{ng/mL}$, $U_{\mathrm{IPTG}} = 0.5\,\mathrm{mM}$, $k_{\mathrm{P},1} = 60$, $k_{\mathrm{P},2} = 0.75$, $k_{\mathrm{I},1} = 1.5$, $k_{\mathrm{I},2} = 0.05$. Cells (about 200) in the simulated microfluidic chamber (with dimensions $40\,\mu\mathrm{m} \times 50\,\mu\mathrm{m} \times 1\,\mu\mathrm{m}$) have the same parameters' value, and their evolution has been obtained using the agent-based simulator BSim [122, 123].

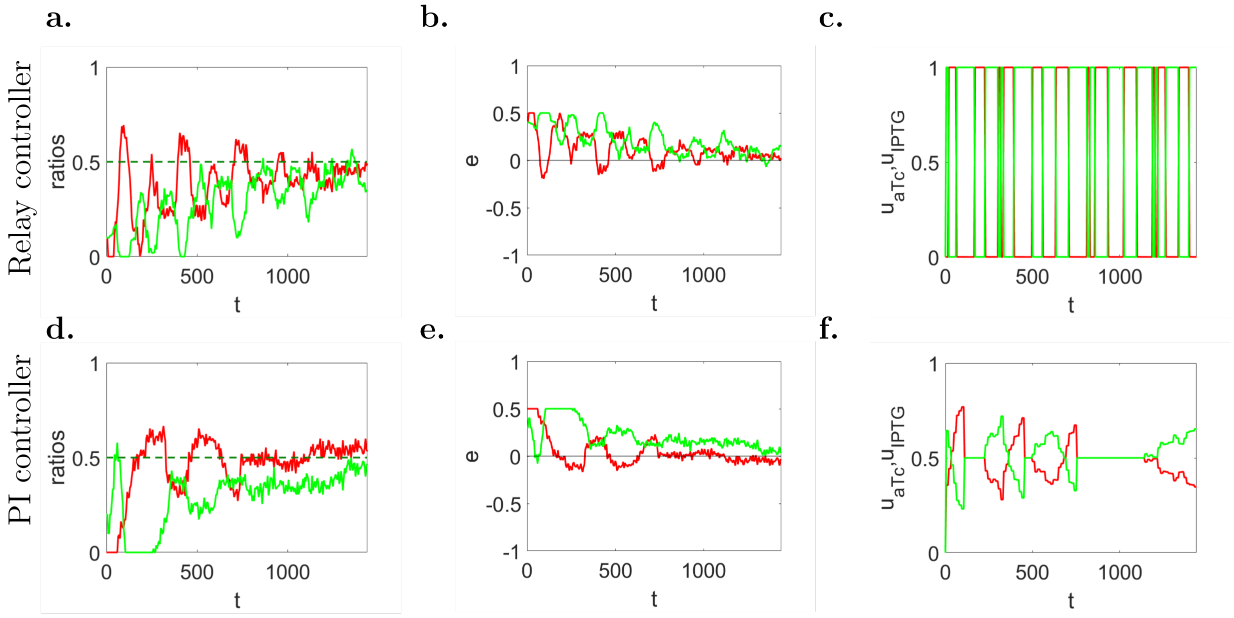


Figure A-11: **Cooperative production of two monomers to 1:1 population ratio can still be achieved in *smaller* growth chambers.** Larger noise fluctuations due to cells flowing out of the microfluidic chambers are expected in this situation. (**a,d**) Evolution in time of populations' ratio r_A (solid green line) and r_B (solid red line) with their respective desired reference values in dashed lines ($r_\mathrm{d} = 0.5$), (**b,e**) of the error signals e_A (solid green line) and e_B (solid red line), and (**c,f**) inducer control signals u_aTc (solid red line) and u_IPTG (solid green line), normalised to their maximum values U_aTc and U_IPTG, respectively. (**a-c**) Parameters of the relay control (A-41): $U_\mathrm{aTc} = 60\,\mathrm{ng/mL}$, $U_\mathrm{IPTG} = 0.5\,\mathrm{mM}$. (**d-f**) Parameters of the PI controller (A-43): $U_\mathrm{aTc} = 100\,\mathrm{ng/mL}$, $U_\mathrm{IPTG} = 1\,\mathrm{mM}$, $k_\mathrm{P,1} = 100$, $k_\mathrm{P,2} = 1.5$, $k_\mathrm{I,1} = 1.5$, $k_\mathrm{I,2} = 0.05$. The simulated microfluidic chamber (with dimensions $13.3\,\mu\mathrm{m} \times 16.7\,\mu\mathrm{m} \times 1\,\mu\mathrm{m}$) can host a cell population of about 30 cells. All simulations were obtained using the agent-based simulator BSim [122, 123]. (See also Supplemental Video 2.)

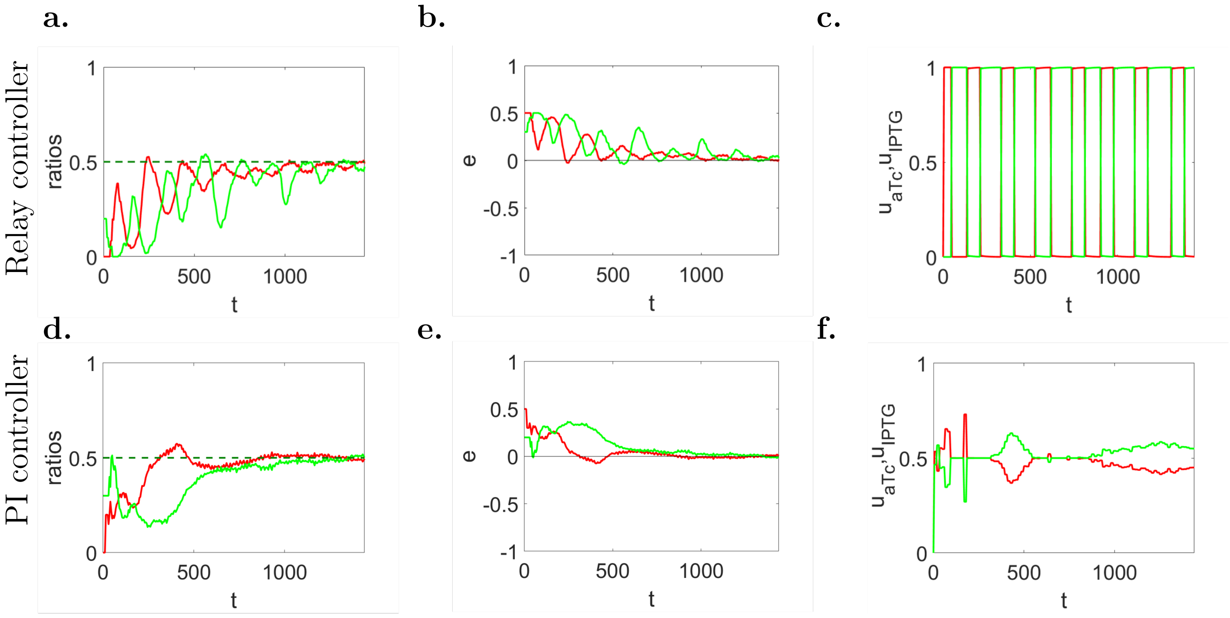


Figure A-12: **Cooperative production of two monomers to 1:1 population ratio can be achieved in a *heterogeneous* cell population by means of feedback ratiometric controllers in microfluidics.** (**a,d**) Evolution in time of populations' ratio r_A (solid green line) and r_B (solid red line) with their respective desired reference values in dashed lines ($r_\mathrm{d} = 0.5$), (**b,e**) of the error signals e_A (solid green line) and e_B (solid red line), and (**c,f**) inducer control signals u_aTc (solid red line) and u_IPTG (solid green line), normalised to their maximum values U_aTc and U_IPTG, respectively. (**a-c**) Parameters of the relay control (A-41): $U_\mathrm{aTc} = 60\,\mathrm{ng/mL}$, $U_\mathrm{IPTG} = 0.5\,\mathrm{mM}$. (**d-f**) Parameters of the PI controller (A-43): $U_\mathrm{aTc} = 100\,\mathrm{ng/mL}$, $U_\mathrm{IPTG} = 1\,\mathrm{mM}$, $k_{\mathrm{P},1} = 100$, $k_{\mathrm{P},2} = 1.5$, $k_{\mathrm{I},1} = 1.5$, $k_{\mathrm{I},2} = 0.05$. The simulated microfluidic chamber (with dimensions $40\,\mu\mathrm{m} \times 50\,\mu\mathrm{m} \times 1\,\mu\mathrm{m}$) can host a cell population of about 200 cells. Each time a cell divides, the parameters of the daughter are drawn from a normal distribution centred in the nominal value, and with variance equal to its 20%. All simulations were obtained using the agent-based simulator BSim [122, 123].

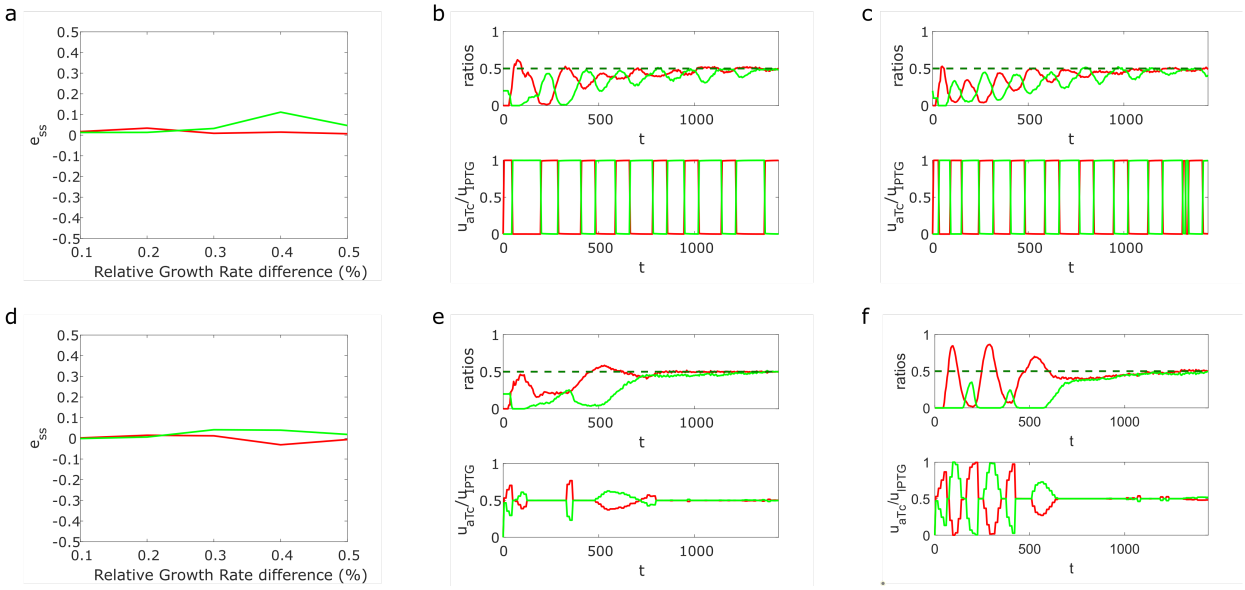


Figure A-13: **Cooperative production of two monomers to 1:1 population ratio can also be achieved when the expression of the two proteins is associated to different metabolic burdens.** (**a,d**) Average steady-state error in dependence of the relative growth rate difference when expressing different phenotypes. The steady-state error e_{ss} is computed as the average error over the last 60 minutes of the simulation (lasting 1440 min). The *relative growth rate difference* is computed as $(g_n - g_p)/g_n$, where g_n is the nominal growth rate and g_p is the reduced growth rate induced by assuming additional metabolic burden posed by overexpressing LacI. (**b,c,e,f**) Top panel: evolution in time of populations' ratios r_A (solid green line) and r_B (solid red line), with the desired reference values depicted with a dashed line ($\text{r}_d = 0.5$); bottom panel: inducer control signals u_{aTc} (solid red line) and u_{IPTG} (solid green line), normalised to their maximum values U_{aTc} and u_{IPTG}, respectively. (**a-c**) Parameters of the relay control (A-41): $U_{\text{aTc}} = 60\,\text{ng/mL}$, $U_{\text{IPTG}} = 0.5\,\text{mM}$. (**d-f**) Parameters of the PI controller (A-43): $U_{\text{aTc}} = 100\,\text{ng/mL}$, $U_{\text{IPTG}} = 1\,\text{mM}$, $k_{\text{P},1} = 100$, $k_{\text{P},2} = 1.5$, $k_{\text{I},1} = 1.5$, $k_{\text{I},2} = 0.05$. Cells (about 200) in the simulated microfluidic chamber (with dimensions $40\,\mu\text{m} \times 50\,\mu\text{m} \times 1\,\mu\text{m}$) have the same parameters' value, but their growth rate changes when LacI is expressed. Specifically, relative growth rate differences equal to 0.1 and 0.5 have been chosen for the *in silico* experiments reported in panels (**b,e**) and (**c,f**), respectively. The cells' evolution has been obtained using the agent-based simulator BSim [122, 123].

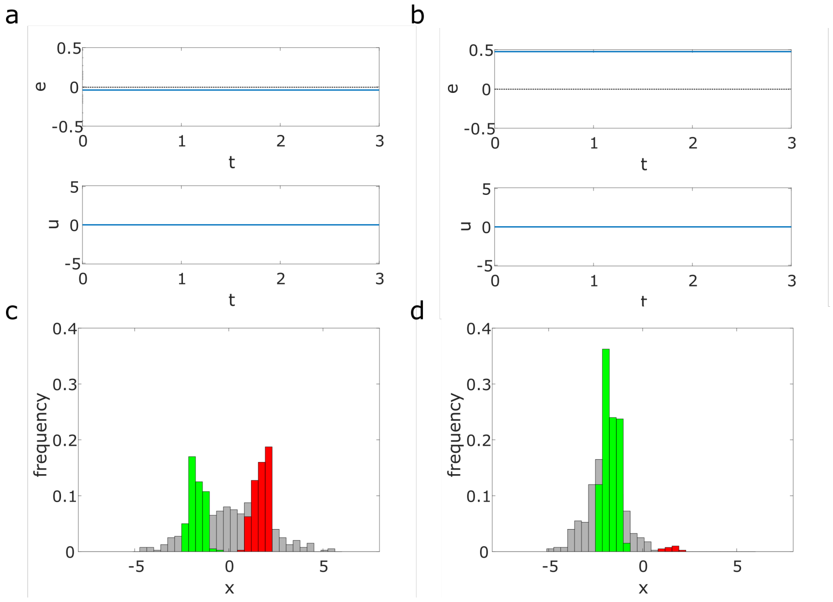


Figure A-14: **In the absence of any control input, the cells described by equation (6-1) converge to some bimodal distribution whose relative numbers of cells in each mode depend on the initial conditions of the cells.** (**a,b**) Evolution in time of the error signal $e(t)$ and the control input fed to the population. The desired set-point for the error ($e = 0$) is depicted as a dashed grey line. (**c,d**) Distribution of the cells state at the beginning of the simulation ($t = 0$ a.u., grey histogram) and at steady state ($t = 3$ a.u., green and red histograms). The initial conditions $x_i(0)$ of all cells ($N = 400$) are drawn from the normal distributions $N(0, 4)$ and $N(-2, 1)$, respectively for panels **a,c** and **b,d**, while the parameters η_i are drawn from a uniform distribution in the interval $[1, 5]$.

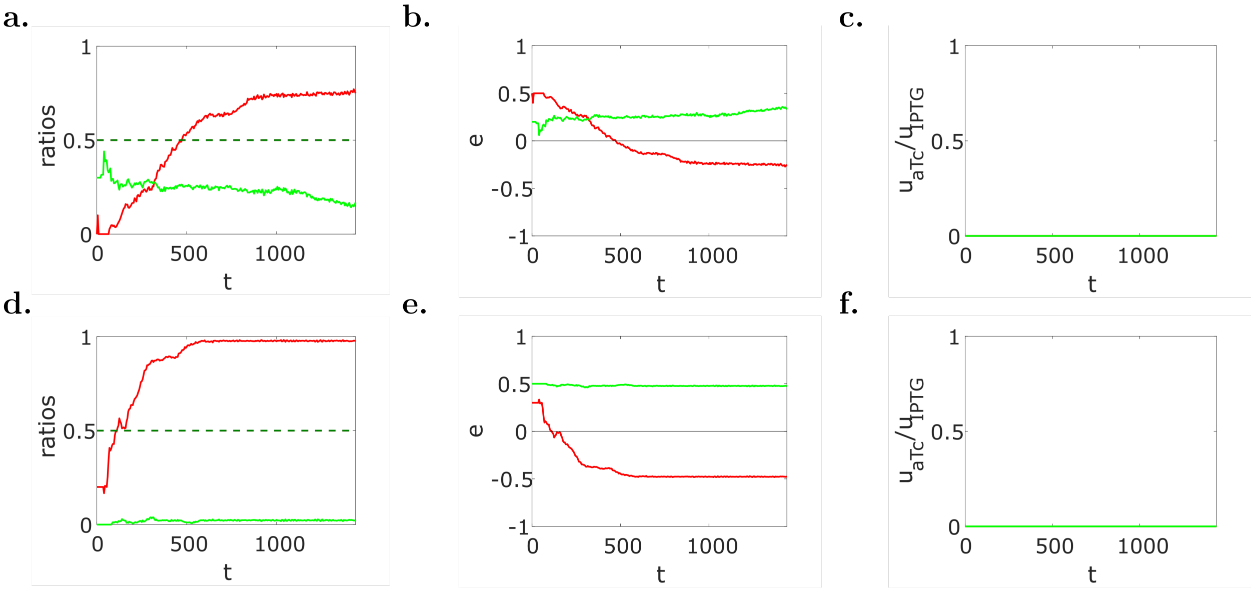


Figure A-15: **In the absence of any feedback control action the cell population does not converge to the desired ratio, but to some other value that depends on the initial conditions.** (**a,d**) Evolution in time of populations' ratio r_A (solid green line) and r_B (solid red line), with the desired reference values depicted with a dashed lines (r_d=0.5). (**b,e**) Evolution in time of the control errors e_A (solid green line) and e_B (solid red line). (**c,f**) Inducer control signals u_{aTc}(solid red line) and u_{IPTG} (solid green line), normalised to their maximum values U_{aTc} and u_{IPTG}, respectively. Cells (about 200) in the simulated microfluidic chamber (with dimensions 40 μm × 50 μm × 1 μm) have the same parameters' value, and their evolution has been obtained using the agent-based simulator BSim. The initial conditions of the cells, that is, the values of $mRNA_{LacI}$, $mRNA_{TetR}$, $[LacI]$ and $[TetR]$ at the beginning of the *in silico* experiments were drawn from a uniform distributions in the intervals $[3, 9]$, $[3, 9]$, $[150, 450]$, $[200, 600]$ for panels **a-c**, and $[6, 18]$, $[3, 9]$, $[300, 900]$, $[200, 600]$ for panels **d-f**. The concentration of the inducer molecules at the beginning of the simulation inside each cell was set to 0.

A.4 Supplementary material for TES

A.4.1 Derivation of the aggregate model

The first part of the model, equations (6-14)-(6-17), follows immediately from the assumption that the average concentrations follow the same dynamics described in (6-8)-(6-11) for all cells in the population.

Equation (6-18) can be obtained by noticing that (6-12) can be rewritten as

$$\eta\left(Q_i - Q_e\right) = \rho_0 - \gamma q_i - \frac{dq_i}{dt}, \tag{A-45}$$

that substituted into (6-13) gives

$$\frac{dq_e}{dt} = \sum_{i=1}^{M}\left(\rho_0 - \gamma q_i - \frac{dq_i}{dt}\right) - \gamma_e q_e, \tag{A-46}$$

which in turns can be rewritten as

$$\frac{d}{dt}\left(q_e + \sum_{i=1}^{M} q_i\right) = \rho_0 M - \left(\gamma_e q_e + \sum_{i=1}^{M}\gamma q_i\right). \tag{A-47}$$

Now, noting that, by definition, $q := q_e + \sum_{i=1}^{M} q_i$, i.e., q is the total amount of molecule Q into the external environment and into all the cells, and that, when considering a constant volume V, only degradation must be considered, that is, $\gamma = \gamma_e$, we obtain

$$\frac{dq}{dt} = \rho_0 M - \gamma_e q, \tag{A-48}$$

that, by dividing both sides by V and by using $N = M/V$, yields (6-18). Finally, equation (6-19) follows by assuming the typical logistic growth dynamics for the cell population in a limited environment, with the term $-dPN$ capturing the slowdown effect of the inhibitor protein P on the growth rate of the cells, similarly to what was reported in Ref [154].

A.4.2 Nondimensional model

Model (6-14)-(6-19), with the first-order approximation of the Hill function, can be nondimensionalized to obtain a simpler model with less parameters by rescaling time as $\tau = \gamma t$ and introducing the following dimesionless variables:

$$\begin{aligned} &\bar{S} = \frac{S}{S_0}, \quad \bar{T} = \frac{T}{T_0}, \quad \bar{C} = \frac{C}{C_0}, \\ &\bar{P} = \frac{P}{P_0}, \quad \bar{Q} = \frac{Q}{Q_0}, \quad \bar{N} = \frac{N}{N_m}, \end{aligned} \tag{A-49}$$

with

$$S_0 = T_0 = C_0 = \frac{\gamma k}{\kappa_P d}, \quad P_0 = \frac{k}{d}, \quad Q_0 = \frac{N_m \rho_0}{\gamma}. \tag{A-50}$$

Table A-4: Nominal vaules of the systems' parameters.

Parameter	**Value**	**Unit**	**Reference**
α_0^S	5.316	tr h^{-1}	[157]
α^S	variable	tr h^{-1}	-
α_0^T	13.4	tr h^{-1}	[157]
α^T	11400	tr h^{-1}	[157]
κ_{C+}	1.542	co tr^{-1} h^{-1}	[157]
κ_{C-}	0.4032	co^{-1} tr h^{-1}	[157]
κ_P	5	nM $\mathrm{co}^{-1}\mathrm{h}^{-1}$	[154]
ρ_0	595.2	nM $\mu\mathrm{m}^3$ h^{-1}	-
γ	2	h^{-1}	[154]
γ_e	2	h^{-1}	[154]
η	120	$\mu\mathrm{m}^3$ h^{-1}	-
θ	20	nM	-
n	2	-	[157]
k	0.97	h^{-1}	[154]
d	4 10^{-3}	nM^{-1} h^{-1}	[154]
N_m	0.145	CFU $\mu\mathrm{m}^{-3}$	-
V_c	5.5	$\mu\mathrm{m}^3$	-

The units of measurement tr, co and CFU stand for transcripts, complexes, and colony-forming units, respectively.

Thus, a *nondimensional model* of (6-14)-(6-19) can be obtained as:

$$\frac{d\bar{S}}{d\tau} = \bar{\alpha}^S - \bar{\kappa}_{C+}\bar{T}\bar{S} + \bar{\kappa}_{C-}\bar{C} - \bar{S} \tag{A-51}$$

$$\frac{d\bar{T}}{d\tau} = \bar{\alpha}^T\bar{Q} - \bar{\kappa}_{C+}\bar{T}\bar{S} + \bar{\kappa}_{C-}\bar{C} - \bar{T} \tag{A-52}$$

$$\frac{d\bar{C}}{d\tau} = \bar{\kappa}_{C+}\bar{T}\bar{S} - \bar{\kappa}_{C-}\bar{C} - \bar{C} \tag{A-53}$$

$$\frac{d\bar{P}}{d\tau} = \bar{C} - \bar{P} \tag{A-54}$$

$$\frac{d\bar{Q}}{d\tau} = \bar{N} - \frac{1}{\bar{\gamma}}\bar{Q} \tag{A-55}$$

$$\frac{d\bar{N}}{d\tau} = \bar{k}\bar{N}(1-\bar{N}) - \bar{d}\bar{P}\bar{N} \tag{A-56}$$

where

$$\begin{aligned} \bar{\alpha}^S = \frac{\alpha^S}{\gamma S_0}, \quad \bar{\alpha}^T = \frac{Q_0\alpha^T}{2\gamma\theta S_0}, \quad \bar{\gamma} = \frac{\gamma}{\gamma_e}, \\ \bar{\kappa}_{C+} = \frac{S_0\kappa_{C+}}{\gamma}, \quad \bar{\kappa}_{C-} = \frac{\kappa_{C-}}{\gamma}, \quad \bar{d} = \bar{k} = \frac{k}{\gamma}, \end{aligned} \tag{A-57}$$

are positive nondimensional parameters.

A.4.3 Logistic growth dynamics in BSim

The typical rod-shaped geometry of *E. coli* bacteria is modeled in BSim by considering for each cell a cylinder with radius $r = 0.5\,\mu$m and length L having a semi-sphere with the same radius r attached to both bases. Cells have a random initial length drawn with uniform distribution from the interval $[0.7,\, 1.3]$ μm, and their growth is implemented by assuming that the length L of each cell grows by following a logistic dynamics of the form:

$$\frac{dL}{dt} = k_L L \left(1 - \frac{L}{L_{\max}}\right), \tag{A-58}$$

where $L_{\max}$ denotes the maximum cell length, set to $5.625\,\mu$m, and k_L denotes the growth rate. Cell division occurs when the length of the cell reaches the critical threshold[123] $L_{\text{th}} = 4.5\,\mu$m. As a consequence, the time needed for the division to occur, that is, the doubling time of the population, given the initial length of a bacterium L_0, is defined as:

$$t_L^{\text{d}} = -\frac{1}{k_L} \ln\left[\left(\frac{L_{\max}}{L_{\text{th}}} - 1\right) \frac{1}{\frac{L_{\max}}{L_0} - 1}\right]. \tag{A-59}$$

Note that, in the original implementation in BSim, t_L^{d} does not depend on the population density N, hence population will grow exponentially at a constant rate. Therefore, in order for the population density N to follow a logistic growth rate in BSim, it is necessary to make k_L depending directly on N. To this aim, we assume that the doubling time of the cell length t_L^{d} is equal to the doubling time of the population density, whose dynamics is described by:

$$\frac{dN}{dt} = kN\left(1 - \frac{N}{N_m}\right) - dPN, \tag{A-60}$$

so that the doubling time of the population density can be derived as:

$$t_N^{\text{d}} = \frac{\ln(2)}{k\left(1 - \frac{N}{N_m}\right) - dPN}. \tag{A-61}$$

Finally, by equating the two doubling times, t_N^{d} and t_L^{d}, we obtain the desired relationship between k_L and N as:

$$k_L(N,P) := \frac{k\left(1 - \frac{N}{N_m}\right) - dPN}{\ln(2)} \ln\left[\frac{L_{\text{th}}(L_{\max} - L_0)}{L_0(L_{\max} - L_{\text{th}})}\right], \tag{A-62}$$

which was implemented to carry out the agent-based simulations reported in our research. Notice that, according to (A-62), the growth rate k_L can also become negative due to the effect of the inhibitor protein P, causing the cell length to shrink. When the length becomes smaller than $0.5\,\mu$m, the cell is assumed to be dead, and it is therefore removed from the chamber.

Appendix B

Agent-based simulations of bacteria in BSim

B.1 Introduction

Having a realistic environment for the simulation of bacterial populations and their growth environment is essential for testing *in silico* proof of concept designs for sythetic biological circuits, partially filling the gap between modeling and *in vivo* experiments. BSim was desveloped in [122] and updated with new functionalities in [123] as an agent-based simulator explicitly designed for the simulation of bacterial populations. In this environment we can accurately model the environment where bacteria grow, simulating diffusion and decay processes of all the chemical species present. In addition, it is possible to explicitly account for cells' bio-mechanics, including their movement and division.

We extended the environment original implementation, presented in [123], to include the noise deriving from the intrinsic stocasticity of biochemical processes taking place inside the cells. Specifically, we implemented the SDE-based algorithm presented in [174]. Formally, for each cell we solved a system of equations of the form

$$d\mathbf{x} = \mathbf{S} \cdot \mathbf{F}(\mathbf{x}, t)dt + \mathbf{S} \cdot \mathrm{diag}(\mathbf{F}(\mathbf{x}, t))d\mathbf{w} + \Delta(\mathbf{x})dt, \tag{B-1}$$

where $\mathbf{x}$ is the state vector, containing all the species constituting the GRN hosted in the cell, $\mathbf{w}$ is a vector of independent standard Wiener processes, $\mathbf{S} = [\mathbf{s}_1, \ldots, \mathbf{s}_6]$ is the stoichiometry matrix associated to the Gene Network and $\mathbf{F}(\mathbf{x}) = [a_1(\mathbf{x}), \ldots, a_6(\mathbf{x})]^\top$, where $a_j(\mathbf{x})$ is the propensity function of the j-th pseudo-reaction. Notice that there is an additional term $\Delta(\mathbf{x})dt$ that models the the diffusion dynamics of chemicals across the cell membrane, assumed to be deterministic.

The growth environment we simulated consists of a microfluidic device, a microscopy, a computer and an actuation system. The microfluidic device contains a chamber where cells are trapped and grow, which is connected to two perfusion channels that bring nutrients and inducers to the cellular population. The flow of the fluids sent to the chambers is governed by two syringes, each one containing a mixture of growth media and one inducer. By adjusting the relative height of the syringes, it is possible to change the amount of inducers delivered to the cells. The architecture is completed by a computer, controlling the syringes depending on the output of a control algorithm, and a microscopy, necessary to take images whose analysis results in the computation of the error.

The nature of this platform imposed some technological constraints on the simulations:

1. The state of the cells cannot be measured more often than 5 min, to avoid excessive photo-toxicity;
2. There is a time delay of 40 s on the actuation of the control inputs due to the time that the flow of the chemical inducers takes to reach the chambers in the microfluidic chip where cells are hosted;
3. The minimum time interval between two consecutive control inputs cannot be smaller than 15 min, to limit excessive osmotic stress on the cells;

In this appendix we describe three case studies where the BSim extension we developed was used both to validate proof of concept designs for *external* control of bacterial populations [142, 19] and to support the design and implementation of *in vivo* feedback control of an error comparation module in *E. Coli* [18].

B.2 Balancing cell populations endowed with a synthetic toggle switch via adaptive pulsatile feedback control

The problem of controlling cells endowed with a genetic toggle switch around its unstable equilibrium has been recently highlighted as a benchmark problem in synthetic biology. It has been shown that a carefully selected periodic forcing can balance a population of such cells in an undifferentiated state. The effectiveness of these control strategies, however, can be hindered by the presence of stochastic perturbations and uncertainties typically observed in biological systems and is therefore not robust.

To solve this problem and balance a toggle switch population, modelled as described in section A.3.10, we designed a feedback (closed-loop) control approach that is based on using two mutually exclusive periodic inputs of aTc and IPTG of period T and duty cycle d_k, where k indices the inputs in time. The algorithm is able to select their amplitudes and adapt and change online their duty-cycle to achieve the desired control goal.

In particular, we adopted an external control strategy [25] that can be implemented in microfluidics via a fluorescence microscope and actuated syringes as described in the previous section. The *feedback* control algorithms adapt the duty-cycle d_k of the periodic inputs as a dynamic function of the current cell behavior, measured via the fluorescence microscope. Specifically, we implemented two feedback control actions:

- a proportional-integral (PI-PWM) controller that drives a pulse width modulation block;
- a Model Predictive Controller (MPC) that selects the input duty-cycle dynamically by optimizing a desired cost function.

Feedback control strategies

PI-PWM controller. The controller evaluates, at every time instant $t_k = k\,T$, $k \in \mathbb{N}_{>0}$, the duty-cycle d_k by adding a correction δd_k to the reference value d_{ref} returned by a Model Based Inversion algorithm, based on analytical approximations made in [126] (Figure B-1a). Specifically, d_{ref} is computed using the curve of equilibria $\Gamma_{\bar{u}_{\mathrm{aTc}},\bar{u}_{\mathrm{IPTG}}}$ of the average model, which is parameterized in terms of the duty cycle and the amplitude pulse waves. The Model Based Inversion algorithm also uses a projector Π to project the desired setpoint and the current state on $\Gamma_{\bar{u}_{\mathrm{aTc}},\bar{u}_{\mathrm{IPTG}}}$. (for more details see [142]).

Thus, the duty-cycle is evaluated online by the PI controller as $d_k = d_{\mathrm{ref}} + \delta d_k$, $k \in \mathbb{N}_{>0}$, where δd_k is computed as:

$$\delta d_k = k_{\mathrm{P}}\, e_k^{\pi} + k_{\mathrm{I}} \sum_{j=0}^{k} e_j^{\pi}$$

with initial condition $d_0 = d_{\mathrm{ref}}$, and where k_{P}, k_{I} are the gains of the PI controller and e_k^{π} is the *error* computed projecting the current stae on the equilibria cureve using Π.

Specifically, the nonlinear projector Π computes at the end of each period the error e_k^{π} as the length of the arc on the curve of equilibria $\Gamma_{\bar{u}_{\mathrm{aTc}},\bar{u}_{\mathrm{IPTG}}}$ between the projections on the curve of the desired mean value (or setpoint) $\bar{x}_{\mathrm{ref}}$ and the current measured mean value over a time period T computed as

$$\langle x_k \rangle = \frac{1}{T} \int_{kT}^{(k+1)T} x(\tau)\, d\tau.$$

The projections of these two points are denoted as $\widehat{x}_{\mathrm{ref}}$ and $\langle \widehat{x}_k \rangle$, respectively.

Because of the highly nonlinear and uncertain nature of the model, the tuning of the PI gains was carried heuristically via extensive numerical simulations in MATLAB.

Note that the projected error e_k^{π} being equal to 0 (i.e. $\|\langle \widehat{x}_k \rangle - \widehat{x}_{\mathrm{ref}}\| = 0$) does not necessarily correspond to zero regulation error of the mean state value $\langle x_k \rangle$ onto $\bar{x}_{\mathrm{ref}}$, that is $\|\langle x_k \rangle - \bar{x}_{\mathrm{ref}}\| = 0$. Indeed, at steady-state the line connecting these two points is orthogonal to the curve $\Gamma_{\bar{u}_{\mathrm{aTc}},\bar{u}_{\mathrm{IPTG}}}$ but its length may not be zero. This residual error at steady-state can be made smaller by computing more curves $\Gamma_{\bar{u}_{\mathrm{aTc}},\bar{u}_{\mathrm{IPTG}}}$ in the database.

Model Predictive Controller (MPC). MPC strategies have been widely applied in the field of synthetic biology [175], demonstrating their effectiveness also for *in-vivo* experiments [140]. The controller chooses, at each sampling time $t_k = k\,T$, the duty-cycle d_k to be applied in the next control cycle (of duration T), by solving an online optimization problem on a finite prediction horizon interval of length $T_{\mathrm{p}} = N\,T$, with $N \in \mathbb{N}$ (Figure B-1b). Specifically, at each step k, the algorithm finds the sequence of duty-cycles $\{d_0, d_1, \ldots, d_{N-1}\}$ of length N that minimize the cost function J_k, defined as

$$J_k(\{d_0, d_1, \ldots, d_{N-1}\}) = \sum_{i=0}^{N-1} \int_{(k+i)T}^{(k+1+i)T} e(t, d_i)\, dt,$$

where $e(t, d)$ is the weighted relative squared error defined as

$$e(t, d_i) = K_{\mathrm{LacI}} \left(\tfrac{LacI(t,d_i) - LacI_{\mathrm{ref}}}{LacI_{\mathrm{ref}}}\right)^2 + K_{\mathrm{TetR}} \left(\tfrac{TetR(t,d_i) - TetR_{\mathrm{ref}}}{TetR_{\mathrm{ref}}}\right)^2,$$

in which $LacI(t, d_i)$ and $TetR(t, d_i)$ are the solution of the *deterministic* dynamical model (A-33) when PWM pulsatile signals with duty-cycle d_i are applied as inputs to the system. In order to make the control action robust to uncertainties and noise, only the first element d_0 of the sequence is used by the controller, that is, $d_k = d_0$.

The optimization problem is solved at each step by using the *genetic algorithms*[176] toolbox available in MATLAB, which also generates the sequences of duty-cycles $\{d_0, d_1, \ldots, d_{N-1}\}$, such that $d_i \in [0, 1]$.

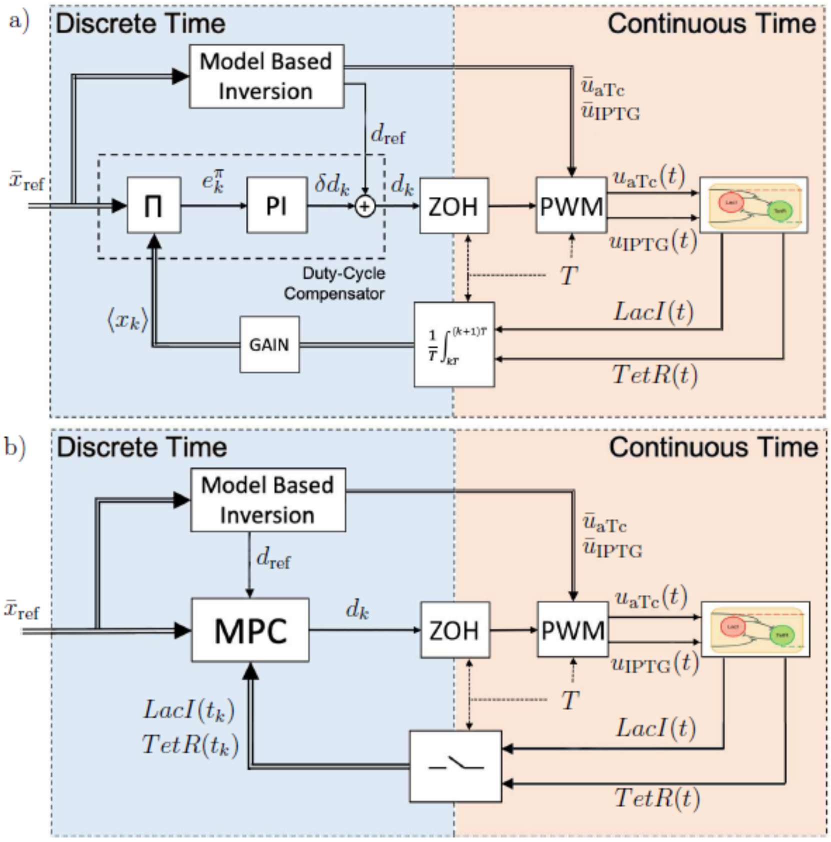


Figure B-1: Block diagrams of proposed feedback control strategies. (a) PI-PWM control strategy. Given the setpoint for the average model $\bar{x}_{\mathrm{ref}}$, two actions regulate the parameters of the PWM inputs that feed the system. The feedforward action is composed by the Model Based Inversion that evaluates the amplitudes $\bar{u}_{\mathrm{aTc}}$ and $\bar{u}_{\mathrm{IPTG}}$ and the nominal value of the duty-cycle d_{ref}. The nonlinear projector Π and a proportional-integral controller compose the feedback loop. At each time period $t_k = kT$, the nonlinear projector Π evaluates the projection error e_k^π that is minimized by a PI controller that evaluates the correction δd_k to be added to d_{ref}. (b) MPC control strategy. At each discrete time step k, the MPC finds the sequence of duty-cycles $\{d_0, d_1, \ldots, d_{N-1}\}$ that minimizes the cost function J_k over the prediction horizon $T_{\mathrm{p}} = NT$. Then, only the first element is selected ($d_k = d_0$) and the corresponding pulsatile control signals are applied to the biological system in the time interval $[t_k,\, t_k + T]$.

Agent-based validation of the control algorithms

In this context, with a view towards the *in-vivo* implementation, we validated via agent-based simulations the performance of both control strategies. To this aim we conducted *in-silico* experiments with BSim 2.0, an advanced agent-based simulator of bacterial populations [122, 123] that is able to replicate realistic phenomena such as cell growth, spatial diffusion, variability in cell geometry, and flush-out of the cells from the microfluidic chamber (see Section B.1 for further details).

Figure **B-2** shows the results of the agent-based simulations confirming the effectiveness and viability of the strategies for *in-vivo* experiments. We see that both strategies are effective in controlling the mean value of the toggle switch response to the desired value despite the presence of perturbations and uncertainties.

In general, we find that in the *deterministic* case the MPC – often used in control applications in synthetic biology [140] – guarantees better performance in terms of dynamic regulation; the overshoot and transient duration being significantly lower than those observed with the other strategy. Moreover, the MPC achieves also better steady-state regulation of the setpoint with respect to the PI-PWM controller over 18 periods (corresponding to 72 hours, considered to be realistic for *in-vivo* control applications).

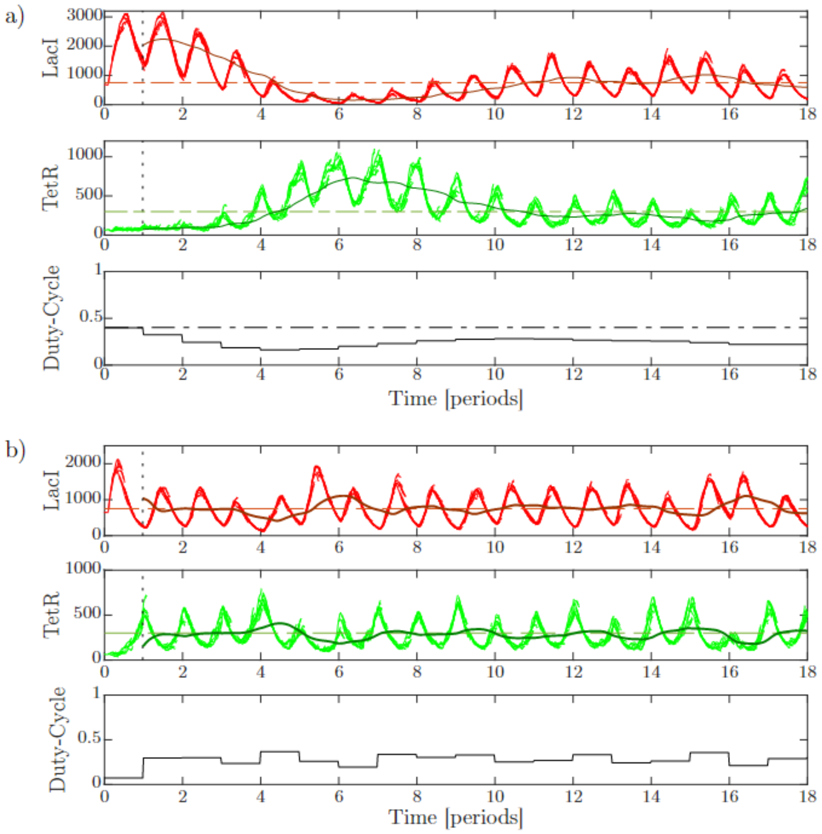


Figure B-2: Agent-based simulation in BSim 2.0. (a) Response of the cells controlled by using the PI-PWM strategy. PI gains were set to $k_{\mathrm{P}} = 0.0101$ and $k_{\mathrm{I}} = 0.0401$. Top panel: evolution over time of $LacI$; the dashed line representing the setpoint $LacI_{\mathrm{ref}} = 750$, while lighter lines the evolution of the state for each cell in the simulation, and the darker solid line the mean trajectory computed over the population, evaluated through a moving window of period T. Middle panel: evolution over time of $TetR$; the dashed line representing the setpoint $TetR_{\mathrm{ref}} = 300$, lighter lines are the evolution of the state for each cell in the simulation, while the dark solid line represents the evolution of the mean trajectory across the population in the period, evaluated using a moving window of period T. Bottom panel: evolution of the duty-cycle over time. (b) Response of the cells controlled by using the MPC strategy. MPC cost function parameters were set to $K_{\mathrm{LacI}} = 1$, $K_{\mathrm{TetR}} = 4$, while the prediction horizon is $N = 3$. We considered *E. coli* cells growing in a single chamber of a "mother machine"-like microfluidic device [177]: the simulations start with a single cell located at the bottom of the chamber; as the cell grows and duplicates, it pushes outside of the chamber new cells exceeding the maximum capacity of the chamber (around 10 cells). The total simulation time is 72 hours with $T = 240\,\mathrm{min}$. The reference values provided by the Model Based Inversion algorithm and used in both simulations are $\bar{u}_{\mathrm{aTc}} = 35$, $\bar{u}_{\mathrm{IPTG}} = 0.35$, and $d_{\mathrm{ref}} = 0.4$.

B.3 Control-Based Continuation of an inducible Toggle switch

Mathematical modelling is widely used within synthetic biology design-build-test-learn cycles. In the context of engineered gene regulatory networks, models can both support the design phase (indicating the parameter space which allows the emergence of the desired dynamics, such as oscillations), and the testing upon experimental implementation. Moreover, modelling is currently the only way in synthetic biology to study the relationship between physical parameters variations and bifurcations; the latter represent stability boundaries where qualitative and quantitative changes to the system's dynamics occur.

The derivation of biochemical models can however be challenging, both in terms of model structure (which depends on underlying hypothesis on the system), and parameter identification and validation (which can be troublesome in the case of incomplete/noisy experimental datasets). Model uncertainties will inevitably result in misleading conclusions [178, 179]. As a consequence, the design and testing of engineering synthetic biochemical circuits that perform as intended is extremely difficult, unless various design, build, test and learn iterations are performed [76]

Control-Based Continuation (CBC) is a general and systematic method to carry out the bifurcation analysis of physical experiments that does not rely on a mathematical model and thus overcomes the uncertainty introduced when identifying bifurcation curves indirectly through modelling and parameter estimation.

Control Based Continuation algortithm

CBC retrieves stable and unstable invariant solutions of a dynamical system through the application of an external control action. In order to acquire the bifurcation diagram, the controller must not modify the position in the parameter space of the uncontrolled system's invariant solutions. For example, the equilibria of a controlled system are in general different from the one of the uncontrolled system and, to recover the response of the underlying uncontrolled system of interest, CBC seeks a control signal that decays asymptotically to zero, i.e.,

$$\lim_{t\to\infty} u(x(t), x_{ref}) \equiv 0 \tag{B-2}$$

where $x \in \mathcal{R}$ is the system's state, $x_{ref} \in \mathcal{R}$ is the control reference signal and t is time. Although the control signal is asymptotically converging to zero, it is synthetized in order to stabilise local dynamics of the system such that unstable equilibria become stable and hence detectable forward in time. A control strategy that satisfies (B-2) is called noninvasive and does not modify the system's equilibrium positions in the parameter space. Finding a noninvasive control signal usually requires finding the right reference input (x_{ref}) for the controller such that (B-2) is satisfied, while the solution of interest is stabilised.

When the control input can be chosen as the bifurcation parameter of interest, the methodology can be significantly simplified as the control signal is only required to settle to a constant value. Indeed, when this condition is achieved, the non-zero constant control signal can be viewed as a mere shift in the bifurcation parameter value. In this section, Control

Based Continuation will be used to infer the bifurcation diagram of a Genetic Toggle switch described by the model introduced in section A.3.10 using IPTG as a control input and bifurcation parameter.

The algorithm applied for control-based continuation can be briefly summarized in the following steps:

1. **Set a new control reference**. A reference value $x_{ref} = TetR^*$ is used to evaluate the error (the difference between the measured $TetR$ and the reference target), and it is given to the controller. The range where the reference is picked normally covers the minimum and maximum expression of the protein of interest, but it can be modified according to the need.

2. **Compute the control action**. Depending on the control strategy, the proportional controller evaluates the control action depending on the state of the system.

3. **Feed the control action to the system**. The computed control action u is fed to the system for 5 minutes continuously, as it is considered to be the minimum sampling time for measurements and actuation in an hypothetical microfluidics/microscopy experimental set-up. After 5 minutes, the system output is measured and a new error is computed. Steps 2 and 3 are repeated until the process is considered at steady state.

4. **Check for steady state**. The initial implementation of the algorithm considers the system at steady state after a fixed amount of time (set to $9h$ and $55min$). The algorithm allows a variable time for steady state computation. In this case, to compute the steady state of $TetR$ and $IPTG$, the slope of the linear curve fitting the last 12 samples after 3 hours of simulation is calculated for both the error and the control signals (where the slope corresponds to the angle the curve makes with respect to the $x - axis$). If the computed value is below a user-defined tolerance, the system is assumed at steady state. Otherwise, the simulation continues and the steady state is checked every new sample, until the slope is sufficiently small. A threshold is set on both the error and the control slopes, which may vary between a deterministic and a stochastic simulation.

5. **Acquire the steady state**. The steady state values of $TetR$ and $IPTG$ ($\overline{TetR}$ and $\overline{IPTG}$) are saved as the average value computed over the last 12 samples of each signal, and the algorithm goes back to step 1.

All steps are repeated until enough steady states are collected, or until the range of output and parameter values has been covered. An initial guess is given as initial state for the algorithm to start with.

The control algorithm chosen to complement the CBC algorithm were a proportional controller and an MPC algorithm.

The proportional controller, formally described as

$$u = K_{\mathrm{p}}(x_{\mathrm{ref}}(t) - x(t)), \tag{B-3}$$

delivers a control action proportional to the mismatch between the measured output and the reference set point. For the toggle switch, the error signal is computed by subtracting the measured output to the control reference ($TetR^* - TetR$), while the control input delivered is the $IPTG$ concentration.

The Model predictive control is a control scheme based on two main features: prediction and optimization. At each step, a model reproducing the process behaviour is used to predict the process outputs to given input signals. The input minimizing a cost function is then computed using an optimization algorithm and fed to the controlled process. We chose a linear model to reproduce behaviours we expect from the real process. A further requirement for the prediction procedure is the process current state which, if not fully measurable, can be estimated using a Kalman filter. Further details on the system used for the preditction, the estimation process and the optimization algoritm used can be found in [19].

Agent based validation

To validate the CBC algorithm in a relistic simulation environment we used BSim, the Java-based bacteria simulator described in section B.1. We implemented single cell control considering the stochastic model described by Equations (A-39).

In figures **B-3** B, **B-4** B representative examples of $TetR$ and $IPTG$ trajectories are shown. The bifurcation curves that are obtained, shown in Fig. **B-3** A and **B-4** A, are comparable with the bifurcation curve numerically computed using the nominal values of the system. However, we observed worse control performances with respect to the MPC for the proportional controller in BSim. This is mainly due to additional factors introduced in simulations. Specifically, the cells biomechanics and their flush out from the microfluidic chamber are simulated, as well as the chemicals' spatial distribution and diffusion that introduce additional delay in the control input delivery. Additionally, in BSim we explicitly simulated an actuation delay due to the time the media takes to be delivered to the cells within the microfluidic device. All these factors might contribute to the performances exhibited by the proportional controller. A similar result when using more realistic simulation environments was also found in [17], where the authors showed that the performance of a PI controller considerably deteriorated when the algorithm was tested on an agent based model, while an MPC algorithm was able to maintain similar performance.

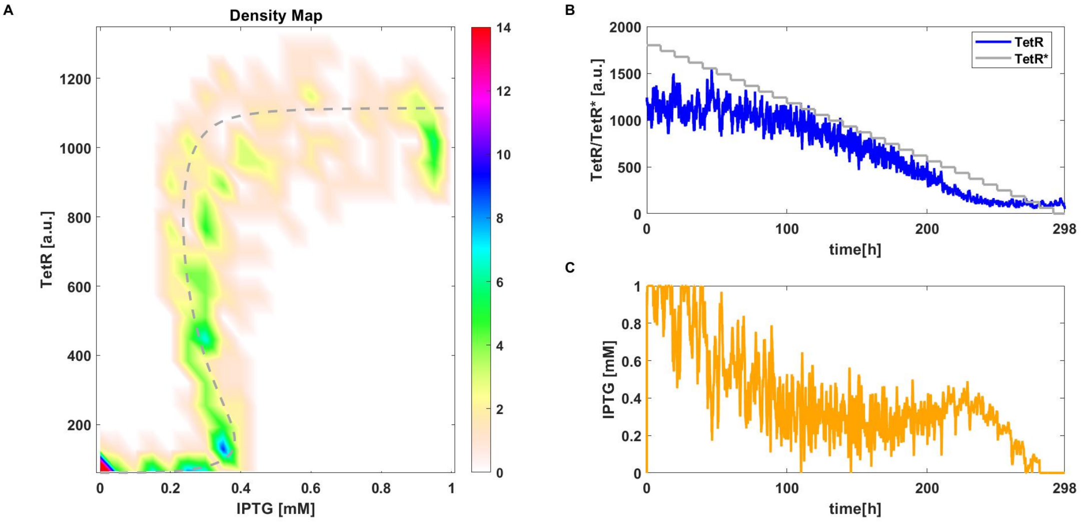


Figure B-3: CBC with P controller applied to the stochastic toggle switch model (A-39) in BSim. A) Density plot of equilibrium curve measured using CBC. Reference equilibrium curve obtained using numerical continuation (- -). B) Time evolution of one simulation of $TetR$ and the control reference signal $TetR^*$ (–). C) Time evolution of $IPTG$ (i.e. control signal). Parameter values: $k_p = 0.0016$ and $aTc = 25\, ngmL^-1$.

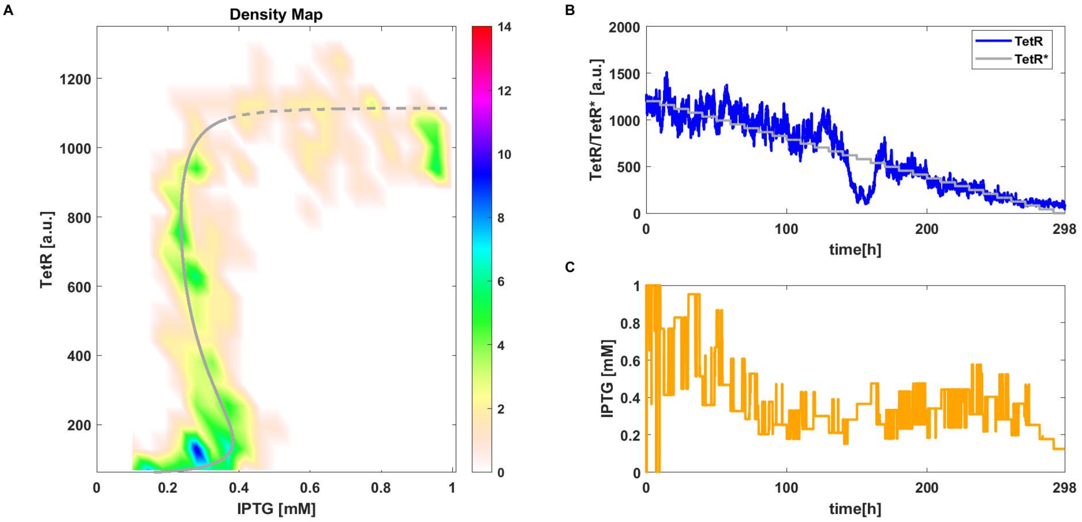


Figure B-4: CBC with MPC applied to the stochastic toggle switch model (A-39) in BSim. A) Density plot of equilibrium curve measured using CBC. Reference equilibrium curve obtained using numerical continuation (- -). B) Time evolution of one simulation of $TetR$ and the control reference signal $TetR^*$ (–). C) Time evolution of $IPTG$ (i.e. control signal). Parameter values: $\gamma = 0.3$ and $aTc = 25\, ngmL^-1$.

B.4 *In Vivo* Feedback Control of an Antithetic Molecular Titration Motif in *Escherichia coli* Using Microfluidics

Molecular titration modules have been highlighted as a fundamental component for the implementation of feedback controllers in living cells. Sequestration mechanisms can be used as effective error computation modules for the implementation of in vivo feedback control of living cells and have been previously evaluated in silico [55] and confirmed experimentally [105]. More specifically, these modules can be used to compare some desired reference input of interest with a signal related to the phenotype one wishes to control, so as to provide an error signal that the molecular control device can use to regulate its behaviour.

The crucial role played by molecular titration motifs for feedback control of cells necessitates a more thorough study of their properties and dynamic response. In this section we present the in vivo external feedback control of an engineered *E. coli* population endowed with the gene regulatory network shown in Figure **B-5**a. The controller is implemented in microfluidics by closing the loop through an inducible promoter controlled by an antithetic σ/anti-σ module that is driven by an external reference input (see Figure **B-5**b).

In this context we developed a mathematical model that supported the development of a relay controller guarantee reliable and tunable regulation of gene expression levels.

Benchmark molecular titration module

We implemented our previously designed biological system [105] as a benchmark to test the dynamic response of an error computation module - a comparator - based on molecular titration [112]. The system is a multi-input, single output device that uses two independent signals to regulate the expression level of a target protein of interest. It is based around an orthogonal σ/anti-σ pair able to specifically regulate the activity of the p_20992 promoter through antithetic behaviour[112] (Figure **B-5**A). The cell-cell communication molecule 3-O-C6-HSL (AHL) [180] and the chemical inducer Isopropyl β-D-1-thiogalactopyranoside (IPTG) were used as external inputs to control the system. The p_20992 promoter was placed upstream of the superfolder GFP (sfGFP) protein. sfGFP was chosen for its fast maturation time of in the order of minutes [181]. Production of σ_20_992 was controlled by the AHL inducible promoter plux, whilst the production of the anti-σ_20_992 was controlled by the IPTG inducible promoter *plac-UV5*. This system works on the basis that free anti-σ can bind to σ, preventing it from recruiting RNA polymerase to the p_20992 promoter for the expression of GFP. Therefore, the level of GFP expression will be proportional to the amount of free σ available, and to the difference between the two signals (AHL and IPTG). All proteins were fused to a degradation tag (ssrA tag) to ensure the fast dynamics of expression from the system [182].

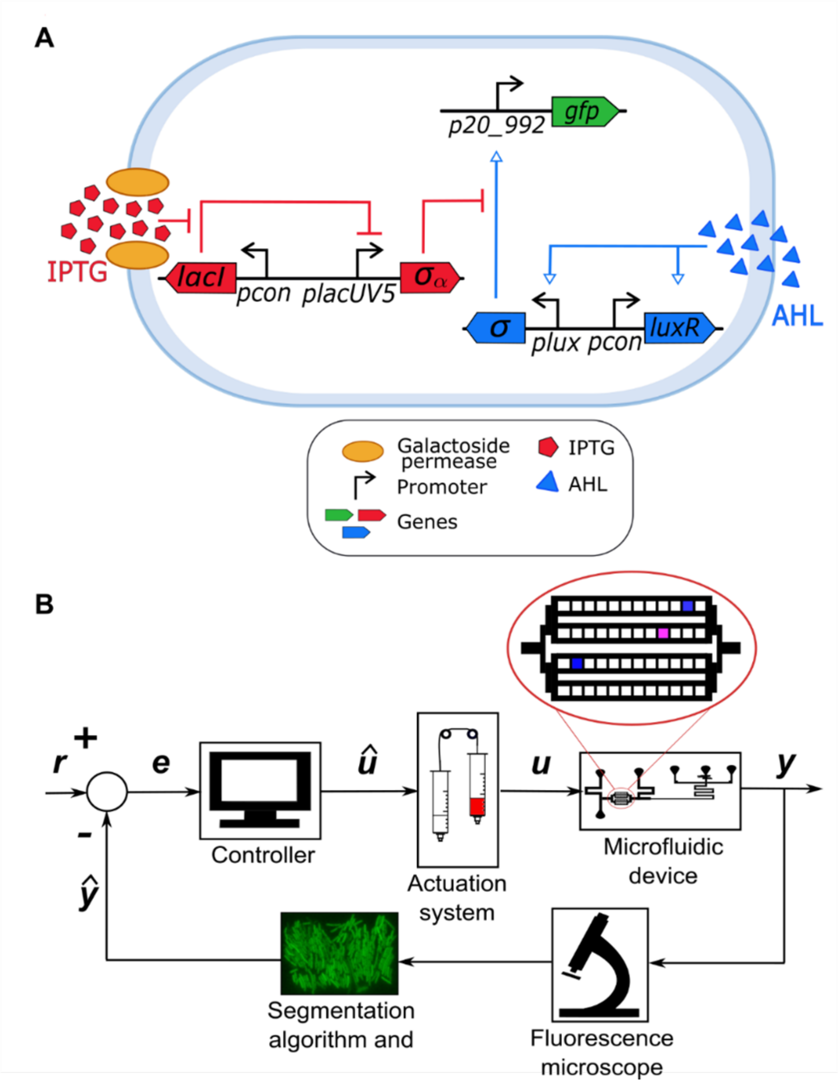


Figure B-5: A) Biological design of the molecular titration system. GFP output from the system is dependent on the concentrations of IPTG and AHL, which control the relative levels of anti-σ and σ respectively. This in turn determines the level of p20_992 activity and GFP expression. B) Schematic of the control platform. An inverted fluorescence microscope is used to capture the fluorescence (y) of the cells. Images are segmented to quantify the average fluorescence (y). The error (e) between the desired fluorescence value and the actual output is then evaluated and used to decide if the input should be provided to the cells or removed ($\hat{u}$). The actuation system is then instructed to either deliver or remove inducers(s) from the microfluidic device accordingly. A schematic of where the cells are cultured inside the microfluidics device is also included. Pink boxes indicate cells inside the *controlled* chamber, blue boxes indicate the locations of cells in *uncontrolled* chambers. The exact positioning of these chambers varied between different experiments.

Open-loop system response and in silico modelling

Our existing mathematical model previously developed for batch culture simulations[105] was parameterized to reflect the conditions of our microfluidics experiments. We started by characterizing experimentally the response of the molecular titration system to time-varying external inputs. We showed that the microfluidic platform can maintain bacterial population growth, and that the GFP fluorescence of the cells can be varied as a function of the external IPTG and AHL signals. The experimental data was used to parameterize a computational model of the titration module proposed in our earlier work[105]. Model fitting and predictions were found to match the *in vivo* observations, confirming the viability of the computational model we derived for control system design and in silico validation. The model is described as a set of Ordinary Differential Equations (ODE) as follows:

$$
\begin{aligned}
[\dot{GFP}] &= \chi_{0,GFP} + \frac{\chi_{1,GFP}[\sigma]^{n_\sigma}}{K_\sigma^{n_\sigma} + [\sigma]^{n_\sigma}} - f(X)[GFP] - \gamma_{P,GFP}[GFP] \\
[\dot{\sigma}] &= \chi_{0,\sigma} + \frac{\chi_{1,\sigma}[A]^{n_A}}{K_A^{n_A} + [A]^{n_A}} - k^{+}_{\sigma:\sigma_a}[\sigma][\sigma_a] + k^{-}[\sigma:\sigma_a] - f(X)[\sigma] - \gamma_\sigma[\sigma] \\
[\dot{\sigma_a}] &= \chi_{0,\sigma_a} + \frac{\chi_{1,\sigma_a}[I]^{n_I}}{K_I^{n_I} + [I]^{n_I}} - k^{+}_{\sigma:\sigma_a}[\sigma][\sigma_a] + k^{-}[\sigma:\sigma_a] - f(X)[\sigma_a] - \gamma_{\sigma_a}[\sigma_a] \\
[\sigma \dot{:} \sigma_a] &= k^{+}_{\sigma:\sigma_a}[\sigma][\sigma_a] - k^{-}[\sigma:\sigma_a] - f(X)[\sigma:\sigma_a] - \gamma_{\sigma:\sigma_a}[\sigma:\sigma_a]
\end{aligned}
\tag{B-4}
$$

whose dependent variables model the concentrations in time of mature GFP [GFP], free σ factor [σ], free anti-σ [σ_a] and the σ:anti-σ complex [σ:σ_a]. Their dynamics are determined by production, dilution, and degradation terms. In the model, the dynamics of mRNA transcription was assumed to be faster than the translation process, and a quasi-steady-state approximation of mRNA dynamics was therefore applied. Also, the dynamics of intracellular mRNAs are described through saturating Hill curves parameterized by the concentration of an input, AHL ([A]) or IPTG ([I]), or an RNA polymerase sigma subunit [σ], a half-maximal activation concentration (K_σ,K_A, and K_I) and a Hill coefficient (n_σ, n_A and n_I). Parameters $\chi_{0,i}$ and $\chi_{1,i}$ are the basal and maximal rate of production of protein species i; $\gamma_{P,GFP}$, γ_σ, γ_{σ_a} and $\gamma_{\sigma:\sigma_a}$ are the rates of dilution of GFP, σ, anti-σ and the σ:σ_a complex, respectively; $k^{+}_{\sigma:\sigma_a}$ and $k^{-}_{\sigma:\sigma_a}$ are the association and dissociation rate of the σ:σ_a complex formation.

In Equation (4), the function $f(X) = \frac{Y_D}{c_e+X}$ captures the competition between available ssrA tag for available proteases, where X=[σ]+[σ_a]+[σ:σ_a]+[GFP] is the total number of ssrA tagged proteins present in the system, Y_D defines the maximum value of enzymatic degradation, and c_e defines the half activation threshold.

Closed loop control experiments

Using the in silico model, we designed external closed loop control experiments in which the aim was to keep GFP fluorescence at a fixed set-point or track a time-varying signal. These experiments were first performed in silico and then experimentally in vivo. AHL (10^{-2}mM)

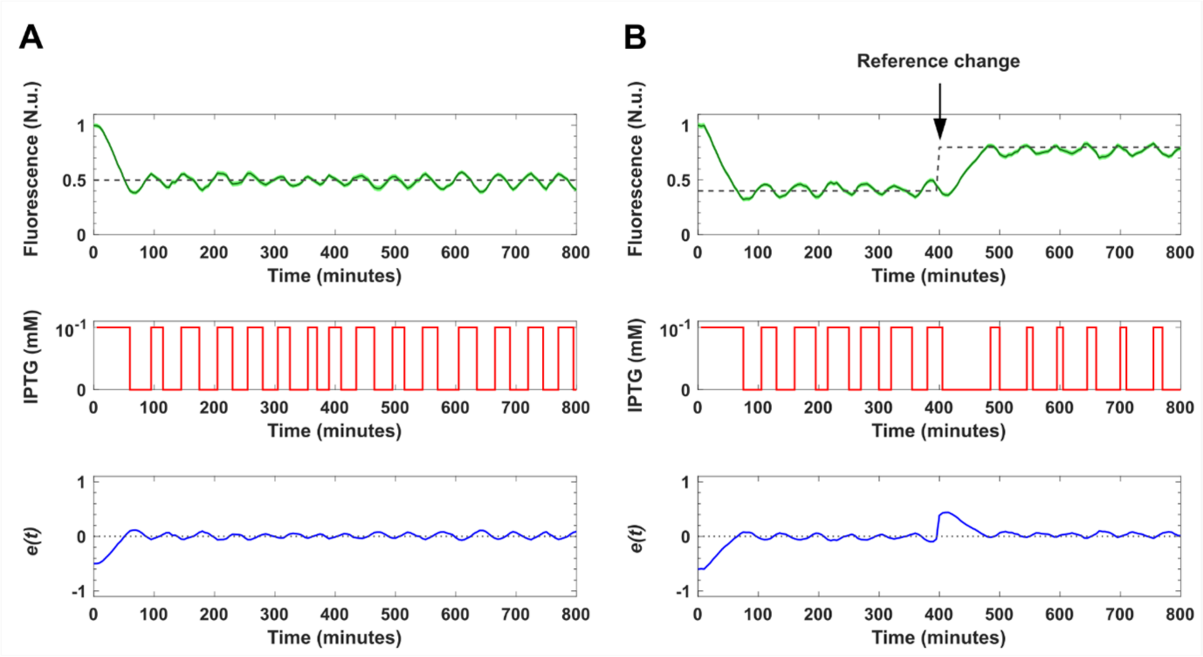


Figure B-6: GFP expression across the cell population (y in green). The simulations were performed with a constant input of 10-2 mM AHL. The desired GFP fluorescence was set to A) 50% of the dynamic range of fluorescence (yref in dashed grey) for the length of the control experiment. B) For the first 400 mins of the experiment, the desired GFP fluorescence was set to 40% of the dynamic range of fluorescence, which was then switched to 80% for the remaining 400 mins (yref in dashed grey). The input signal (10-1 mM IPTG in red), was delivered to the cells when the error (e, in blue) between the desired GFP value and the actual GFP was negative. The control input was removed when the calculated error was positive. A calibration phase was performed prior to control start, see [18]. The integral square error (ISE) for experiment A = 7.79 and for experiment B = 18.15. The standard error of the mean of y (SEM) is represented by the shaded region in lighter green.

was continually supplied to the cells to maintain production of the σ factor, while IPTG was used as a control input (for the generation of anti-σ factor) by switching between 0 mM and 10^{-1} mM.

Prior to the control start in both *in silico* experiments and in vivo experiments, a calibration phase was performed to calculate the minimum and maximum relative fluorescence values exhibited by the cells at steady state for each of the conditions. To implement feedback control, we used a simple yet effective relay (ON/OFF) control strategy to modulate the amount of IPTG provided to the cells as a function of the mismatch, say e(t), between the reference signal (r) and the system fluorescence output (GFP)[183]. Specifically, we set:

$$u = \begin{cases} U_{max} & \text{if} e(t) < 0 \\ U_{min} & \text{if} e(t) > 0 \end{cases} \tag{B-5}$$

where u(t) is the IPTG input to cells set to $U_{max} = 10^{-1}$ mM or $U_{min} = 0$mM. We chose a relay control strategy as its simplicity enabled ease of implementation and was sufficient to provide a proof of concept that our device could be controlled in real-time using a microfluidics platform.

To determine the feasibility of our control strategy, we first performed *in silico* stochastic simulations of the planned *in vivo* closed loop experiments. We embedded the ODE model into the agent-based simulator BSim[123] to recapitulate dynamics in the microfluidic device and account for cell division, input diffusion and cell-to-cell variability. Two different closed loop control experiments were simulated: 1) set-point regulation of GFP to 50% of its dynamic range (Figure **B-6**a), and 2) signal tracking, where the reference value changed from 40% to 80% mid-way through the experiment (Figure **B-6**b). We used ISE (the integral square error, see [18] for further details) as a metric to evaluate the performance of the control strategy; ISE penalises big variations in time of the mismatch providing a reliable assessment of control performance. For both control goals, we observed that the GFP output was driven to the desired level after the first 50 minutes of the experiment and successfully tracked the desired value for the rest of the experiment, with oscillations around the set-point due to the chosen control algorithm. The in silico experiments also showed that the ISE for both regulation and tracking experiments remained relatively low (ISE = 7.79 for regulation and ISE = 18.15 for tracking), confirming the effectiveness of the control strategy.

Proven the feasibility of the control algorithm, *in vivo* control experiments were conducted to demonstrate that the performances predicted *in silico* were met in microfluidic experiments. Further details on the *in vivo* experiments can be found in [18].

Appendix C

A network model of Italy shows that intermittent regional strategies can alleviate the COVID-19 epidemic

Regionalism is an integral part of the Italian constitution. Each of Italy's twenty administrative regions is independent on Health and oversees its own share of the Italian National Health service. The regional presidents and their councils can independently take their own actions, strengthening or, at times, weakening national containment rules. Previous studies have modelled the spread of the epidemics and its evolution in the country at the national level[184, 185, 186], and some have looked at the effects of different types of containment and mitigation strategies [187, 188, 189]. Limited work [190, 191, 192, 193, 194, 195, 196] has taken into account the spatial dynamics of the epidemic but, to the best of our knowledge, no previous paper in the literature has explicitly taken into consideration the pseudo-federalist nature of the Italian Republic and its strong regional heterogeneity when it comes to health matters, hospital capacity, economic costs of a lockdown and the presence of inter-regional people's flows.

In this study, we investigate the whole of the country as a network of regions, each modelled with different parameters. The goal is to identify if and when measures taken by the Italian government had an effect at both the national, but most importantly, at the regional level. Also, we want to uncover the effects on the epidemic spread of regional heterogeneity and inter-regional flows of people and use control theoretic tools to propose and assess differentiated interventions at the regional level to reopen the country and avoid future recurrent epidemic outbreaks.

As aggregate models of the COVID-19 epidemic cannot capture these effects, to carry out our study we derived and parameterized from real data a network model of the epidemics in the country (see Figure **C-1**a), where each of the 20 regions is a node and links model both proximity flows and long-distance transportation routes (ferries, train, planes).

In this context we will describe an *ad hoc* identfication algorithm used to parameterize the model from real data. The parametrized model was used in [20] to evaluate the effectiveness of the national lockdown strategy implemented by the Italian government and to show that

b

Index	Region	Index	Region
1	Abruzzo	11	Lombardy
2	Aosta Valley	12	Marche
3	Apulia	13	Molise
4	Baslicata	14	Piedmont
5	Calabria	15	Sardinia
6	Campania	16	Sicily
7	Emilia	17	Tuscany
8	Friuli-Venezia Giulia	18	Trentino-Alto Adige
9	Lazio	19	Umbria
10	Liguria	20	Veneto

Figure C-1: Schematic diagram of the network model structure and representative regional parameters. **a.** Representative graph of the network model structure used in the paper. Only a subset of all links is shown for the sake of clarity. Solid lines represent proximity links, dashed lines long distance transportation routes (air, train, road), dotted lines show major ferry routes between insular regions and the Italian mainland. **b.** Table of the Italian region names and their positions in the graph.

regional feedback interventions, where each of the twenty regions strengthens or weakens local mitigating actions (social distancing, inflow/outflow control) as a function of the saturation of their hospital capacity, can be beneficial in mitigating possible outbreaks and in avoiding recurrent epidemic waves while reducing the costs of a nationwide lockdown.

C.1 Model formulation

To capture the regional diversity of the response to the epidemic in Italy, we derive a network model of Italy where each node represents a different region and links capture fluxes of people among the regions (see Fig. C-1a). Using a data-driven compartmental modelling approach, a set of ODEs is obtained describing the dynamics of six different compartments in each region (Susceptibles, Infected, Hospitalized, Quarantined, Deceased and Recovered); data analysis being used to define flows among compartments.

As a regional model of the COVID-19 epidemic spread, we use the compartmental model, which we found from data analysis and identification trials to be the simplest model structure able to capture the real data. Specifically, we constructed the model by testing how different configurations of the links between its compartments affected the model ability to capture the available data. We found that the structure we propose was the best compromise between model simplicity and its ability to capture the data.

The full model equations describing the dynamics of susceptible (S_i), infected (I_i), quarantined (Q_i), hospitalized (H_i), recovered (R_i) and deceased (D_i) are:

$$\begin{aligned}
\dot{S}_i =& -\rho_i \beta \frac{S_i I_i}{N_i} \\
\dot{I}_i =& \rho_i \beta \frac{S_i I_i}{N_i} - \alpha_i I_i - \psi_i I_i - \gamma I_i \\
\dot{Q}_i =& \alpha_i I_i - \kappa_i^H Q_i - \eta_i^Q Q_i + \kappa_i^Q H_i \\
\dot{H}_i =& \kappa_i^H Q_i + \psi_i I_i - \eta_i^H H_i - \zeta_i H_i - \kappa_i^Q H_i \\
\dot{D}_i =& \zeta_i H_i \\
\dot{R}_i =& \gamma I_i + \eta_i^Q Q_i + \eta_i^H H_i
\end{aligned} \tag{C-1}$$

where β and γ are the infection and recovery rate, respectively, which are assumed to be the same for all regions as COVID-19 is transmitted from person to person and there is no parasite vector or evidence of environmental parameters significantly altering its infection rate, $\rho_i \in [0, 1]$ is a parameter modelling the effects of social distancing measures in the i-th region, α_i is the rate of infected that are detected and quarantined, ψ_i is the rate of infected that needs to be hospitalized, η_i^Q is the rate of quarantined who recover, η_i^H is the fraction of hospitalized who recover, κ_i^Q is the rate of hospitalized that is transferred to home isolation, κ_i^H is the rate of quarantined who need to be hospitalized, and ζ_i is the mortality rate that was shown from data analysis to be a function of the ratio between H_i and the maximum number, say T_i^H, of patients that can be treated in ICU at the hospitals in i-th region. N_i is the actual population in the i-th region, i.e. the resident population without those removed because quarantined, hospitalized, deceased or recovered.

Extending previous approaches for modelling Dengue fever in Brazil [197], we obtain the national network model of the COVID-19 epidemic in Italy as a network of twenty regions (see Figure **C-1**) interconnected by links modelling commuter flows and major transportation routes among them.

The network model of Italy we adopt in this study is, for $i = 1, \ldots, 20$,

$$
\begin{aligned}
\dot{S}_i &= -\sum_{j=1}^{M}\sum_{k=1}^{M} \rho_j \beta \phi_{ij} S_i \frac{\phi_{kj}(t) I_k}{N_j^p} \\
\dot{I}_i &= \sum_{j=1}^{M}\sum_{k=1}^{M} \rho_j \beta \phi_{ij} S_i \frac{\phi_{kj}(t) I_k}{N_j^p} - \alpha_i I_i - \psi_i I_i - \gamma I_i \\
\dot{Q}_i &= \alpha_i I_i - \kappa_i^H Q_i - \eta_i^Q Q_i + \kappa_i^Q H_i \\
\dot{H}_i &= \kappa_i^H Q_i + \psi_i I_i - \eta_i^H H_i - \zeta \left(\frac{H_i}{T_i^H}\right) H_i - \kappa_i^Q H_i \\
\dot{D}_i &= \zeta \left(\frac{H_i}{T_i^H}\right) H_i \\
\dot{R}_i &= \gamma I_i + \eta_i^Q Q_i + \eta_i^H H_i \\
N_i^p &= \sum_{k=1}^{M} \phi_{ki}(t) \left(S_k + I_k + R_k\right)
\end{aligned}
\tag{C-2}
$$

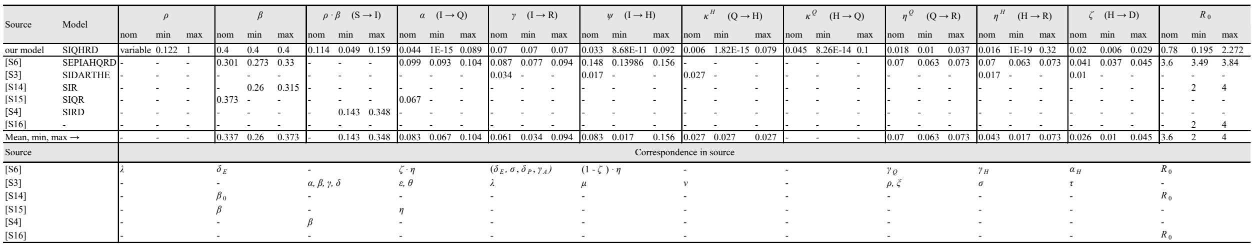
where in addition to the parameters and states described above, we included the fluxes $\phi_i j(t)$ between regions; $\phi_i j(t) : \mathcal{R} \rightarrow [0, 1]$ denoting the ratio of people from region i interacting with those in region j at time t, such that $\sum_{j=1}^{M} \phi_{ij}(t) = 1$. More details on the fluxes definition and estimation can be found in [20]. Note that, as a result of the identification procedure illustrated in Section C.2, in equations (C-2) the mortality rate ζ is expressed as a function of the saturation of the regional health systems whose expression is given in Section C.2.

C.2 Identification

The resulting model is then parameterized using a predictor-corrector algorithm applied to both a national aggregate model and to each of the twenty regional models, identifying the time points at which parameter values present significant changes. Specifically, we divided the model parameterization into two stages. Firstly, we estimated from the available data the parameters of each of the twenty regional models; then, we use publicly available mobility data in Italy to estimate the fluxes among the regions (see [20] for details on fluxes estimates).

The identification procedure assume that, for each region, the parameters of the model remain constant over n time windows, but neither their number n nor their durations $\delta_1, \ldots, \delta_n$ are assumed to be known a priori. Therefore, the identification procedure detects at the same time the breakpoints $t_1, \ldots, t_(n-1)$ when notable parameters' changes are detected and, within each time-window, estimates the parameter values as those that best capture the trend of the available data.

The model used to carry out the identification of regional or national parameters is a discretized version of the model of the epidemic spread in each area of interest, which can be rewritten from model (C-2) for each region (i = 1,. . . ,20) as (dropping the subscripts to the parameters for notational convenience):



Source	Model	ρ			β			$\rho\cdot\beta$ (S → I)			α (I → Q)			γ (I → R)			ψ (I → H)			κ^H (Q → H)			κ^Q (H → Q)			η^Q (Q → R)			η^H (H → R)			ζ (H → D)			R_0		
		nom	min	max	nom	min	max	nom	min	max	nom	min	max	nom	min	max	nom	min	max	nom	min	max	nom	min	max	nom	min	max	nom	min	max	nom	min	max	nom	min	max
our model	SIQHRD	variable	0.122	1	0.4	0.4	0.4	0.114	0.049	0.159	0.044	1E-15	0.089	0.07	0.07	0.07	0.033	8.68E-11	0.092	0.006	1.82E-15	0.079	0.045	8.26E-14	0.1	0.018	0.01	0.037	0.016	1E-19	0.32	0.02	0.006	0.029	0.78	0.195	2.272
[S6]	SEPIAHQRD	-	-	-	0.301	0.273	0.33	-	-	-	0.099	0.093	0.104	0.087	0.077	0.094	0.148	0.13986	0.156	-	-	-	-	-	-	0.07	0.063	0.073	0.07	0.063	0.073	0.041	0.037	0.045	3.6	3.49	3.84
[S3]	SIDARTHE	-	-	-	-	-	-	-	-	-	-	-	-	0.034	-	-	0.017	-	-	0.027	-	-	-	-	-	-	-	-	0.017	-	-	0.01	-	-	-	-	-
[S14]	SIR	-	-	-	-	0.26	0.315	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	2	4
[S15]	SIQR	-	-	-	0.373	-	-	-	-	-	0.067	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
[S4]	SIRD	-	-	-	-	-	-	-	0.143	0.348	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
[S16]	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	2	4
Mean, min, max →		-	-	-	0.337	0.26	0.373	-	0.143	0.348	0.083	0.067	0.104	0.061	0.034	0.094	0.083	0.017	0.156	0.027	0.027	0.027	-	-	-	0.07	0.063	0.073	0.043	0.017	0.073	0.026	0.01	0.045	3.6	2	4

Source	Correspondence in source											
[S6]	λ	δ_E	-	$\zeta\cdot\eta$	$(\delta_E, \sigma, \delta_P, \gamma_A)$	$(1 - \zeta)\cdot\eta$	-	-	γ_Q	γ_H	α_H	R_0
[S3]	-	-	$\alpha, \beta, \gamma, \delta$	ε, θ	λ	μ	ν	-	ρ, ξ	σ	τ	-
[S14]	-	β_0	-	-	-	-	-	-	-	-	-	R_0
[S15]	-	β	-	η	-	-	-	-	-	-	-	-
[S4]	-	-	β	-	-	-	-	-	-	-	-	-
[S16]	-	-	-	-	-	-	-	-	-	-	-	R_0

Figure C-2: Comparison between the parameter's values we used in our work and those used in other papers proposing national models for the COVID-19 epidemic in Italy that recently appeared in the literature. Notice that, unfortunately, often it is not possible to pin down a specific parameter in a different model that clearly corresponds to one of ours, and vice-versa. This is because of the different meaning that compartments have in the models and the dynamics of people between the compartments which do not always overlap in an unambiguous manner. When we had to use a time constant, say τ, to determine the value of a parameter, say k, we set $k = 1/\tau$.

$$\begin{aligned}
\hat{S}(t+1) =& \hat{S}(t) - \rho\beta\frac{\hat{S}(t)\hat{I}(t)}{N_i(0) - \tilde{Q}(t) - \tilde{H}(t) - \tilde{D}(t)} \\
\hat{I}(t+i) =& \hat{I}(t) + \rho\beta\frac{\hat{S}(t)\hat{I}(t)}{N_i(0) - \tilde{Q}(t) - \tilde{H}(t) - \tilde{D}(t)} - \tau\hat{I}(t) - \gamma\hat{I}(t) \\
\hat{C}(t+i) =& \tilde{C}(t) + \tau\hat{I}(t) \\
\hat{Q}(t+1) =& \tilde{Q}(t)\alpha\hat{I}(t) - \kappa^H\tilde{Q}(t) - \eta^Q\tilde{Q}(t) + \kappa^Q\tilde{H}(t) \\
\hat{H}(t+1) =& \tilde{H}(t) + \kappa^H\tilde{Q}(t) + \psi\hat{I}(t) - \eta^H\tilde{H}(t) - \zeta\tilde{H}(t) - \kappa^Q\tilde{H}(t) \\
\hat{D}(t+1) =& \tilde{D}(t) + \zeta\tilde{H}(t) \\
\hat{R}^O(t+1) =& \tilde{R}^O(t) + \eta^Q\tilde{Q}(t) + \eta^H\tilde{H}(t)
\end{aligned} \tag{C-3}$$

where measured quantities are denoted by a tilde and estimated state variables by a hat and $\tau := \alpha + \psi$. Here, $C_i = Q_i + H_i + D_i + R_i^O$ represents the total number of cases detected in region i as daily announced by the Protezione Civile.

As also noted in other previous work[185, 198], identification of SIR and SIR-modified models is highly non-convex and hence the optimization landscape is scattered with local minima that must be avoided as not being admissible. To mitigate this problem, we identified from the literature admissible intervals for the parameter values (see Figure C-2) in order to reduce the feasible search space for the optimization algorithm and provide it with reasonable initial guesses. Note that, β and ρ always appear as a product in the model, therefore, only their product can be identified. Hence, we fixed β to an intermediate value of 0.4 from those reported in the literature [185], where the estimates for β range from 0.301 [185], 0.373 [199], and 0.315 [187] to a much higher value of 1.12 [200]. When β models the transmission rate for documented infections, it can become larger than 1, capturing the fact that the new infected can get the disease from undetected infected individuals [200]. Setting $\beta = 0.4$

in our model roughly scales the parameter ρ between 0 and 1, making apparent the effect of the social distancing rules imposed during the lockdown. Moreover, since γ represents the flux of infected people that recover without having any symptoms, thus connecting two compartments that cannot be measured, we chose to fix this rate at $1/14$ days^{-1} . This choice constitutes a conservative approximation of the value reported in recent work[192] where the longest estimate of the infectious period for an asymptomatic infected (counting from as soon as the contagion occurs) is 12.93 days. Note that in other works[185] an even higher choice of 29.41 days is used. Altogether, the unknown parameters left to be estimated are $[I(0), \rho, \tau, \alpha, \eta^Q, \eta^H, \kappa^H, \kappa^Q, \zeta]$ both at the national and regional level.

It is possible to observe that, exploiting the available data, the predictor can be split into two parts so that two different algorithms can then be used to estimate the parameters of each part. An ad hoc identification algorithm estimates the parameters $\{\tau, I(0), \rho\}$, considering the equations describing S, I and C, and automatically detects the breakpoints where notable parameter changes occur. Instead, an ordinary least squares method is used to identify the parameters of the remaining Equations. Note that, as the actual number of infected is not known, we include the number of infected at the beginning of each time window as a parameter to be estimated by the algorithm used for the nonlinear part. The identification described next is repeated for each of the 20 regions and, for the sake of completeness, to parameterize a national aggregate model.

Step 1: Online Identification of the estimation breakpoints and the parameters $\tau, I(0), \rho$ in each time window. We start by identifying the parameters' vector $\theta := [I(0), \rho, \tau]$ exploiting equations describing S, I and C in Equation (C-3) of the main text and the time series of the number of cases $\tilde{C}$ collected for T_{tot} consecutive days, starting from the day when 10 deceased and 10 recovered were first reported in the area of interest. In particular, an *ad hoc* optimization algorithm (described below and implemented in MATLAB) is used to find breakpoints t_j and the values of the parameters' vector $\hat{\theta}$ that minimize the cumulative squared prediction error in each window, defined as:

$$SSE(\hat{\theta}, t_j, t_{j+1}) = \sum_{t=t_j}^{t_{j+1}} \left\| \tilde{C}(t, \theta)) - \hat{C}(t, \theta)) \right\|^2 \tag{C-4}$$

with $j = 0, 1, 2, \ldots, n-1$.

We use the following recursive procedure:

1. Set the initial time t_0 as the first day in which the first 10 deaths and 10 recovered were reported in the area (region or nation).

2. Assume the initial guess for the width of the window to be $T = [(2p+1)/d]$, where p is the number of parameters to be identified and d the number of measured variables.

3. Estimate the parameters over the entire window to obtain

$$\hat{\theta} = \min_{\hat{\theta}} SSE(\hat{\theta}, t_0, t_0 + T)$$

4. Divide the window into two intervals and estimate the parameters over each subsample obtaining the two estimates

$$\hat{\theta}_a = \min_{\hat{\theta}_a} SSE(\hat{\theta}_a, t_0, t_0 + T/2) \quad \text{and} \quad \hat{\theta}_b = \min_{\hat{\theta}_b} SSE(\hat{\theta}_b, t_0 + T/2, t_0 + T)$$

   subject to the continuity constraints reported In Table C-1.

5. Perform a Chow statistical test:

$$F = \frac{(T - 2p)(\sigma - (\sigma_a + \sigma_b))}{p(\sigma_a - \sigma_b)} \approx \mathcal{F}_{\{p,T-2p\}}$$

   where

$$\sigma = SSE(\hat{\theta}, t_0, t_0 + T),$$
$$\sigma_a = SSE(\hat{\theta}_a, t_0, t_0 + T/2),$$
$$\sigma_b = SSE(\hat{\theta}_b, t_0 + T/2, t_0 + T),$$

   with null hypothesis $H_0 : \{\hat{\theta}_1 = \hat{\theta}_2\}$ and critical p-value $p^* = 10^{-5}$.

6. Then:

   6.1. if $\mathcal{F}_{\{p,T-2p\}} > p^*$, the null hypothesis cannot be rejected, and the parameters are considered constant in the time-window T. Then, the length of the current window is increased by setting T=T+1, and steps 3, 4 and 5 are repeated;

   6.2. if $\mathcal{F}_{\{p,T-2p\}} \leq p^*$ the null hypothesis is rejected, and then the next breakpoint t_1 is selected as

$$t_1 = \text{argmax}\mathcal{F}_{\{p,T-2p\}}(F)$$

   and the parameter set $\hat{\theta}$ that minimizes $SSE(\hat{\theta}, t_0, t_1)$ is selected as the set that best fits the data over the window (t_0, t_1), whose duration is therefore $\delta_1 := t_1 - t_0$.

7. If $t_1 = T_{tot}$ the algorithm is stopped, otherwise starting from t_1 steps 2-7 are repeated to find the next breakpoint and the new set of parameters best estimating the data in the next window until the end of the available datapoints.

Step 2: Offline refinement of the identification process. At the end of the process we will have the set of breakpoints t_j and the parameters set in each of the windows $(t_j, t_(j+1))$ best fitting the data. As the number of windows can be large given the variability in the available data, we refine the estimation results as follows to estimate the minimal number of windows able to capture qualitatively the trend of the real data.

In particular, once the window breakpoints are obtained at the end of step 1, any two consecutive windows of duration say $\delta_j, \delta_(j+1)$ are merged into one larger window of size $\delta_j + \delta_(j+1)$ if one of the two following conditions is verified:

1. The window size of the first window is less than 5 days.

Constraint	Description
$0.9\hat{I}^j \leq \hat{I}^{j+1} \leq 1.1\hat{I}^j$	Continuity constraint on the number of infected at the national level from window j to window j+1.
$LB \leq \hat{I}_i^{j+1} \leq UB$ $LB = 0.9\hat{I}_i^j - 0.1\hat{I}^j$ $UB = 1.1\hat{I}_i^j + 0.1\hat{I}^j$	Continuity constraint on the number of infected at the regional level. The constraints are relaxed by 10% of the national estimated infected to account for the fact that in estimating the region parameter we are neglecting the influx of infected from other regions.
$\hat{\kappa}^Q \leq 0.1$ $\hat{\kappa}^H \leq 0.1$	We assume that the daily number of people hospitalized from quarantine and discharged but still positive (and vice versa) is no higher than 10% of the total.
$0.7\tau^j \leq \alpha^j + \psi^j \leq 1.3\tau^j$	We assume $\hat{\tau}_i = \hat{\alpha}_i + \hat{\psi}_i$ does not differ from the national estimate $\hat{\tau}$ of more than 30%.
$\eta^{Q,j+1} = \eta^{Q,j}$	We assume the recovery rate of those quarantined at home remains the same from a time window to the next as this parameter is likely to be time-invariant. In any case, removing this constraint, we observed no significant change of this parameter from a time window to the next.

Table C-1: Set of parameters constraints enforced by the ordinary least square algorithm used for Step 3 and in the nonlinear identification procedure (Steps 1 and 2).

2. The relative variation of the sum $\tau + \rho$ as estimated in each of the two windows is less than 5%, i.e.
$$\frac{(\tau+\rho)^{j+1} - (\tau+\rho)^j}{(\tau+\rho)^j} \leq 0.05$$
where j denotes the window to which the parameter estimates refers to.

If two windows are merged, then the parameters are estimated again on the entire merged window and the procedure is iterated once more in case condition 2 is still verified.

As a final refinement step, we heuristically explore the effect on the fitting of perturbing the breakpoints within ±5 days from their estimated value. As a representative example, the results of the fitting procedure for the national aggregate model are shown in Table C-2 and depicted in Figure C-3. The same procedure is repeated to parametrize each of the 20 regional models.

As the key use of the identified model is to validate the intermittent mitigation strategies we propose in the paper, it is crucial to check whether the proposed method overfits the data, thus worsening the model prediction ability. To provide a representative validation of our estimation approach, we report in Figure S10 the time evolution of the total number of detected cases at the national level predicted by the model in each time window. It is possible to see that using just 30% of the datapoints from all the available data (shown as red circles), the model predictions (solid bluelines) fit well the rest of the data in each time window both before and after the windows are merged as a result of step 2 with a maximum prediction error of 10000 units.

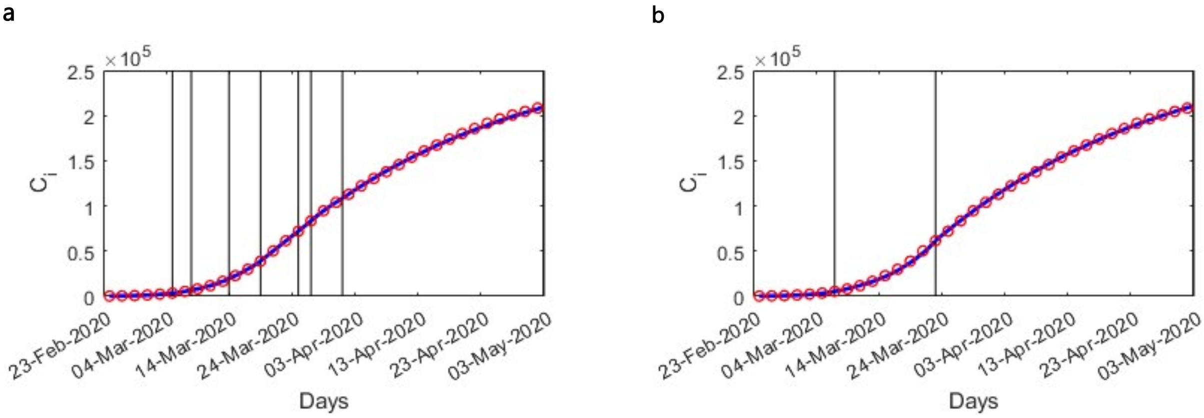


Figure C-3: **a.** Comparison between model estimates and data collected with time widows identified at the end of Step 1 of the parameter identification process. **b.** Comparison between model estimates and data collected with the merged time widows obtained after step 2. In both panels the estimated number of cases estimated by the model $\hat{C}$(blue solid line) is compared with the available datapoints $\tilde{C}$(shown as red circles).

Step 3: Identifying the parameters $\eta^H, \eta^Q, \psi, \alpha, \kappa^H, \kappa^Q, \zeta$**.** For each of the time windows identified in Step 2, using the time series $I(t)$ estimated from the equations parametrized in Step 1, and considering that equations describing Q, H, D, R in model (C-3) of the predictor are linear with respect to the parameters, we use an ordinary constrained least squares method, with constraints given in Table **C-1** to compute the remaining parameters.

The comparison between the model predictions and the available data is depicted in Figure **C-4** and Figure **C-5**. Values of all estimated parameters at the end of the process are given in Table **C-4** for each region where in the last column the regional net reproduction numbers are computed as $R_{0,i} = \rho_i\beta/(\alpha_i+\psi_i+\gamma)$ in each time window. Already in the earliest windows it is possible to see the effects of the first measures taken by the government that date back to February 23rd and March 4th. Indeed, estimates of $R_(0, i)$ in the first windows are lower than the value between 3 and 4 estimated in the literature. Taking Lombardy as a representative example and carrying out our identification procedure on the first 7 days (from February 24th to March 1st), yields an estimate of $R_0 = 3.37$, for that region, confirming that the social distancing measures adopted by the government started taking effect around the beginning of March. Note that we enforce continuity of the trajectory between different time windows by imposing soft constraints in the optimization problems (see Table **C-1**) so that the result of the simulation in the previous time window constrains the dynamics in the next one. The apparent discontinuity between time-windows that can be observed in the parameter values given in Table **C-4** is typical of all predictor-corrector algorithms where after a certain number of prediction steps the actual data points are used to correct the final predictions and restart over the next period. This is standard in systems and control theory (e.g. Kalman predictor and n-step ahead predictors).

Figure **C-6** shows the distribution of the regional social distancing parameters over time depicting the effects of the national lockdown at the regional level.

ρ	I_0	I_f	η_Q	η_H	ζ	α	ψ	κ_H	κ_Q	t_i	$R_{0,i}$
Step 1											
*0.965	1200	11580	0.028	0.045	0.023	0.020	0.050	0.000	0.000	24-2-20	2.76
*0.958	11580	18805	0.028	0.013	0.024	0.018	0.059	0.027	0.049	05-3-20	2.61
†0.774	18805	43814	0.028	0.017	0.029	0.034	0.042	0.000	0.000	08-3-20	2.11
†0.607	43814	65971	0.028	0.019	0.029	0.048	0.025	0.100	0.000	14-3-20	1.70
†0.331	65971	56726	0.028	0.008	0.032	0.053	0.037	0.099	0.100	19-3-20	0.83
‡0.533	56726	59229	0.028	0.002	0.029	0.061	0.042	0.045	0.054	25-3-20	1.24
‡0.234	59229	44109	0.028	0.000	0.027	0.049	0.043	0.056	0.100	27-3-20	0.58
‡0.374	44109	17980	0.028	0.000	0.018	0.023	0.085	0.001	0.088	01-4-20	0.84
Step2											
*0.937	1200	16639	0.028	0.029	0.022	0.018	0.053	0.000	0.000	24-2-20	2.64
†0.646	16639	78621	0.028	0.014	0.032	0.018	0.074	0.000	0.100	07-3-20	1.59
‡ 0.298	78621	26392	0.028	0.000	0.020	0.019	0.055	0.000	0.080	23-3-20	0.82

Table C-2: Parameters of the aggregate national model before and after Step 2. Parameters values are given before and after merging the time windows. Symbols at the beginning of each row denote parameters from windows that are then merged in Step 2 of the identification procedure. Note that, because of the nonlinear nature of the model, parameter values in the merged windows (after step 2) can exceed the ranges of the separate windows obtained in step 1.

ζ_i as a function of the occupancy of ICU beds in each region. Observing the data and the identified parameters in Table C-4, we noticed a significant correlation between the mortality rate in each time window and the congestion of the ICU system in that region.

Specifically, we found that:

$$\zeta_i = f(\bar{H}_i) = \zeta_0^{IC} + \zeta_1^{IC} \bar{H}_i$$

where ζ_0^{IC} and ζ_1^{IC} are coefficients to be estimated, while $\bar{H}_i$ is the estimated average congestion of the hospitals in each time-window, defined as the average ratio between the number of hospitalized subjects and the number of available beds in ICU in that region, say T_i^H

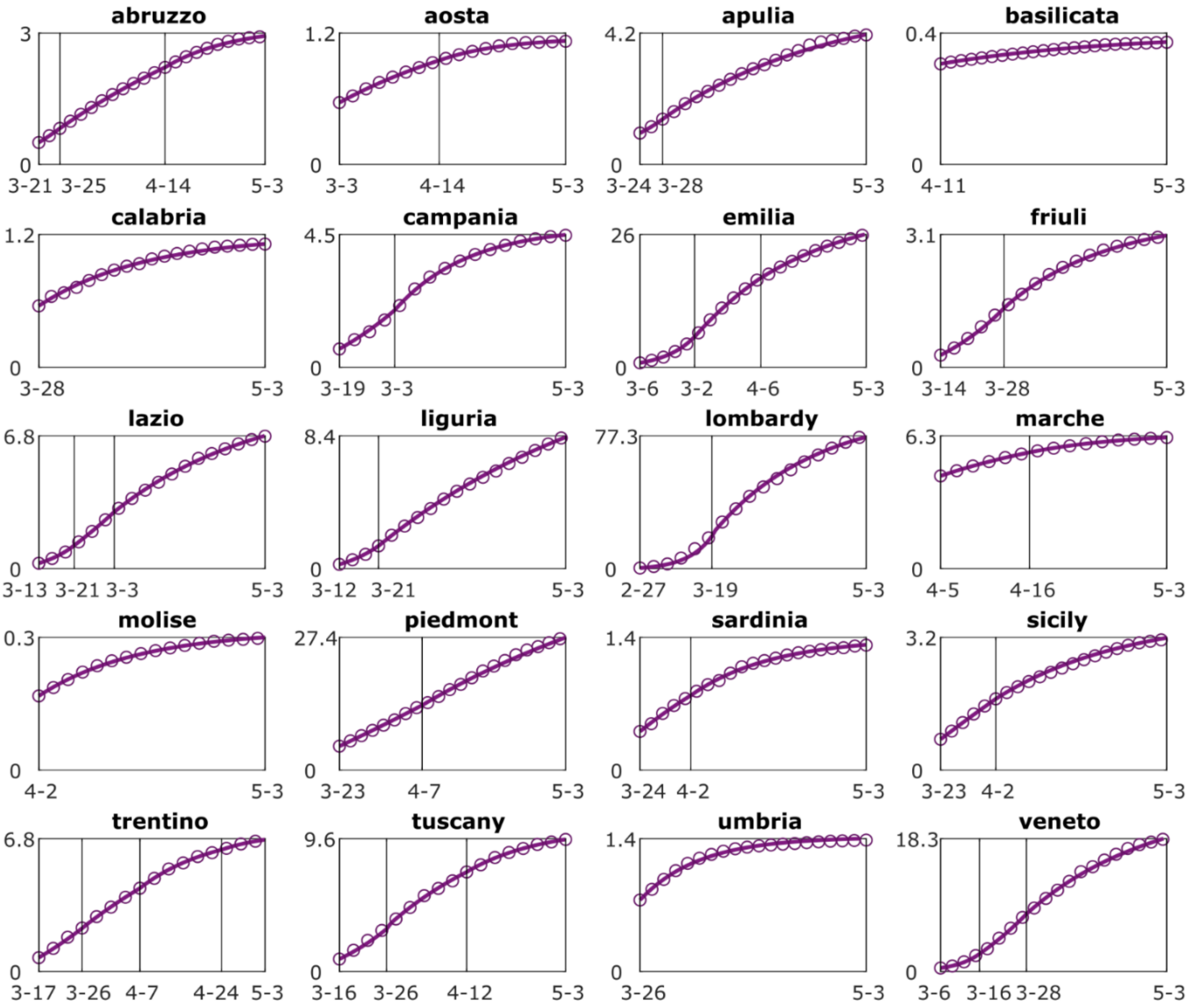


Figure C-4: Identification of the regional models - Steps 1,2. Comparison of each of the regional model predictions for the total number (expressed in thousands of people) of detected cases in each region (solid magenta line) against the available data points. Parameters are set to the values estimated at the end of Steps 2 carried out for each region. Vertical black lines denote the breakpoints from one estimation window to the next.

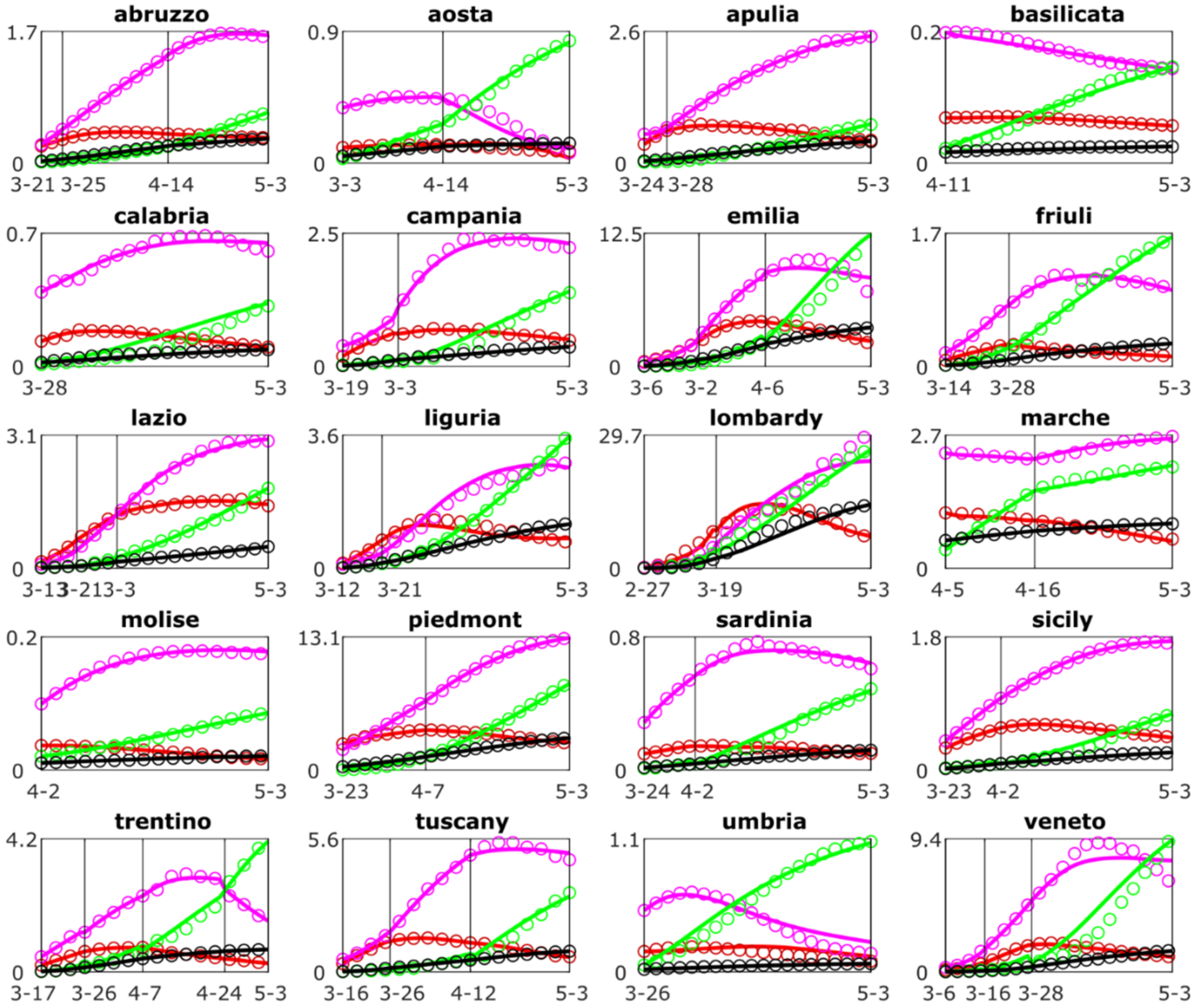


Figure C-5: Comparison of each of the regional model predictions of the total number (expressed in thousands of people) of recovered (green), quarantined (magenta), hospitalized (red), deceased (black) and recovered (green) in each region against the available data points (plotted as circles of the same colour). Parameters are set to the values estimated at the end of Step 3 carried out for each region. Vertical black lines denote the breakpoints from one estimation window to the next.

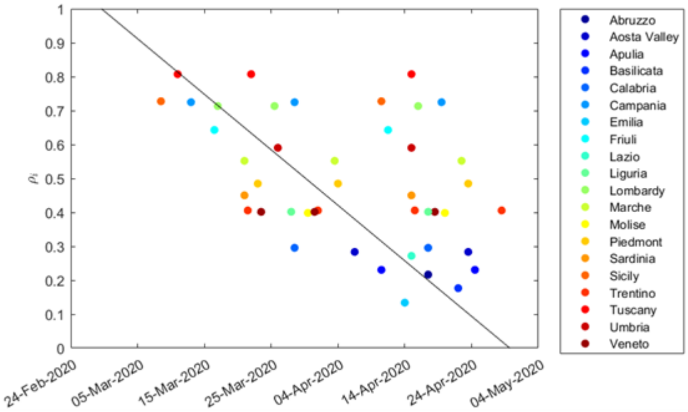


Figure C-6: Distribution of the social distancing parameter ρ_i in the different detected estimation windows. The black line is the LS-interpolant of the National data reported in Table C-4

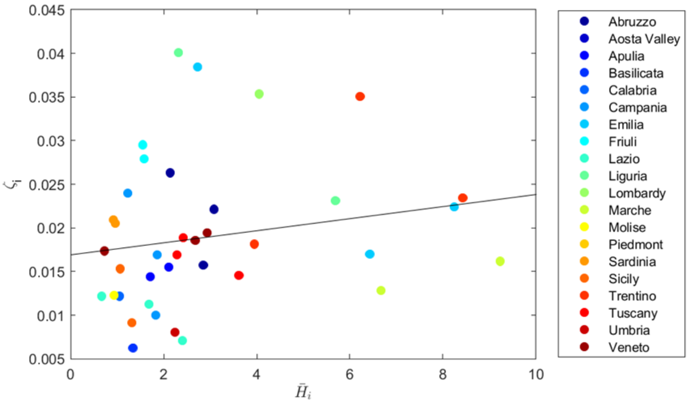


Figure C-7: Each point is the value of ζ_i estimated for a given region in each time window. It is plotted against the average number of hospitalized in each time window over the total number of ICU beds available in the same time window. Here, $R^2 = 0.280$, $p = 2.47 \cdot 10^{-4}$, 45 observations, and 43 DOF. Normality of the residuals has been tested with the Lilliefors test (p-value 0.12). The function is saturated at both ends, i.e. at 0 and at the value 0.029 for $\bar{H}_i \geq 10$.

(obtained by linearly interpolating the number of ICU beds at the beginning of the year, reported on the web-page of the Italian Ministry of Health, and those reported by the Italian Government at the end of the national lockdown). From a qualitative viewpoint, this assumption can be explained by the fact that the quality of the care hospitals can provide degrades as the health system becomes more and more under stress. Each point in Figure C-7 is the value of ζ_i estimated for a given region in each time window, plotted against the number of $\bar{H}_i$ in the corresponding region averaged in the time window. As illustrated in Figure C-7, a least square linear fitting yields $\zeta_0^{IC} = 0.016, \zeta_1^{IC} = 0.00069$. The function is then saturated at both ends, i.e. at 0 and at the value 0.029 for $\bar{H}_i \geq 10$.

Remark on the ad-hoc identification method in the context of the existing literature As in other available methods in the literature (quasi-linearization, Finite Differences, Integration of Data and Smooth-the-Data methods) that cannot rely on analytical solutions of the model equations, we carried out the identification by using some approximating solution of the model predictions. Here, we integrated numerically the model using a Runge-Kutta algorithm and took the solution as a piecewise differentiable approximation of its true solutions (rather than other approximating solutions adopted in existing methods that we found unsuitable in our case). We then incorporated techniques from the literature to find the breakpoints, that is, the points in time where the parameters change significantly, as the Chow test. Overall, our method is therefore in line with the others available in the literature but is fine tuned and adapted to the specific case of interest, also to render its use possible online, i.e. as new datapoints become available.

Region	ρ	I_0	I_f	η^Q	η^H	ζ	α	ψ	κ^H	κ^Q	t_i	R_0
Abruzzo	0,485	944	1083	0,010	0,003	0,026	0,029	0,051	0,005	0,099	21-3-20	1,29
	0,321	1083	747	0,010	0,000	0,022	0,025	0,049	0,000	0,087	25-3-20	0,89
	0,194	847	210	0,010	0,019	0,014	0,078	0,003	0,005	0,000	14-4-20	0,51
Aosta	0,283	425	262	0,010	0,096	0,036	0,062	0,016	0,028	0,000	30-3-20	0,77
	0,122	262	41	0,010	0,260	0,011	0,062	0,000	0,079	0,000	14-4-20	0,37
Apulia	0,590	1300	1732	0,010	0,000	0,016	0,028	0,047	0,100	0,100	24-3-20	1,62
	0,278	1732	530	0,010	0,004	0,015	0,022	0,051	0,002	0,078	28-3-20	0,78
Basilicata	0,177	90	26	0,010	0,060	0,006	0,037	0,021	0,025	0,025	11-4-20	0,55
Calabria	0,272	384	49	0,015	0,000	0,012	0,042	0,059	0,005	0,079	28-3-20	0,63
Campania	0,467	1231	1816	0,018	0,000	0,022	0,014	0,064	0,000	0,100	19-3-20	1,26
	0,221	2231	234	0,018	0,000	0,011	0,067	0,019	0,006	0,040	30-3-20	0,57
Emilia	0,725	1418	7246	0,029	0,000	0,038	0,020	0,089	0,000	0,100	06-3-20	1,62
	0,400	7246	4467	0,029	0,000	0,023	0,059	0,062	0,000	0,045	20-3-20	0,84
	0,362	4467	1881	0,029	0,031	0,017	0,063	0,050	0,000	0,017	06-4-20	0,79
Friuli	0,450	900	1717	0,028	0,022	0,028	0,032	0,034	0,000	0,100	14-3-20	1,32
	0,202	1717	376	0,028	0,049	0,029	0,044	0,007	0,004	0,000	28-3-20	0,67
Lazio	0,713	722	1689	0,015	0,003	0,013	0,018	0,081	0,000	0,062	13-3-20	1,69
	0,483	1689	1995	0,015	0,012	0,012	0,029	0,076	0,055	0,100	21-3-20	1,11
	0,330	1995	732	0,015	0,008	0,007	0,024	0,066	0,039	0,100	30-3-20	0,82
Liguria	0,643	900	1933	0,037	0,010	0,040	0,030	0,070	0,100	0,062	12-3-20	1,52
	0,398	2126	1053	0,037	0,010	0,023	0,012	0,092	0,000	0,100	21-3-20	0,92
Lombardy	0,727	1799	28900	0,010	0,053	0,033	0,009	0,092	0,000	0,040	27-2-20	1,69
	0,303	28900	6731	0,010	0,029	0,024	0,018	0,056	0,000	0,027	19-3-20	0,84
Marche	0,231	1906	1206	0,010	0,080	0,016	0,022	0,047	0,009	0,000	05-4-20	0,66
	0,133	1178	311	0,010	0,007	0,011	0,000	0,057	0,002	0,068	16-4-20	0,42
Molise	0,217	120	13	0,013	0,000	0,012	0,067	0,018	0,000	0,043	02-4-20	0,56
Piedmont	0,398	6527	7244	0,022	0,000	0,019	0,010	0,073	0,000	0,100	23-2-20	1,05
	0,363	7244	4588	0,022	0,014	0,021	0,021	0,071	0,000	0,100	07-4-20	0,90
Sardinia	0,296	618	487	0,013	0,000	0,022	0,064	0,017	0,026	0,100	24-3-20	0,78
	0,216	487	60	0,013	0,038	0,021	0,066	0,017	0,015	0,063	02-4-20	0,56
Sicily	0,402	1025	952	0,015	0,000	0,016	0,048	0,055	0,034	0,100	23-3-20	0,93
	0,293	952	271	0,015	0,000	0,009	0,017	0,068	0,012	0,100	02-4-20	0,75
Trentino	0,406	2305	2867	0,029	0,000	0,035	0,048	0,024	0,001	0,000	17-3-20	1,14
	0,291	2867	2261	0,029	0,000	0,032	0,032	0,038	0,004	0,100	26-3-20	0,83
	0,226	2261	802	0,029	0,035	0,023	0,081	0,006	0,002	0,000	07-4-20	0,58
	0,201	802	368	0,029	0,320	0,018	0,073	0,017	0,060	0,100	24-4-20	0,50
Tuscany	0,552	1675	2666	0,012	0,000	0,017	0,031	0,086	0,014	0,100	16-3-20	1,18
	0,353	2932	1690	0,012	0,000	0,014	0,046	0,062	0,000	0,093	26-3-20	0,79
	0,317	1690	482	0,012	0,064	0,019	0,063	0,055	0,001	0,008	12-4-20	0,68
Umbria	0,134	794	19	0,010	0,141	0,008	0,089	0,000	0,052	0,000	26-3-20	0,34
Veneto	0,807	848	3538	0,031	0,000	0,018	0,062	0,052	0,000	0,100	06-3-20	1,75
	0,520	3538	5089	0,031	0,000	0,019	0,078	0,044	0,000	0,039	16-3-20	1,08
	0,336	5089	1749	0,031	0,000	0,019	0,054	0,048	0,002	0,100	28-3-20	0,78

Table C-4: Values of estimated parameters for each region at the end of the identification process. Dates are given corresponding to breakpoints between estimation windows that are automatically detected by the estimation procedure we proposed. Regional net reproduction number are reported in the last column clearly showing the increasing effect of the national lockdown measures taken by the government on March 8th, 2020.