Università degli Studi di Napoli
"Federico II"
PhD program in Computational Biology and Bioinformatics
27th cycle

Identification and control of gene networks in living cells

Candidate:
Gianfranco Fiore

Tutors:

Dr. Diego di Bernardo

Prof. Mario di Bernardo

Coordinator:

Prof. Sergio Cocozza

Academic Year
2014-2015

Abstract

System identification is a branch of control engineering aimed at developing computational approaches to derive, from measurement data, a quantitative dynamical model of a physical system able to predict its future behaviour. There is a long tradition in the successful application of system identification approaches to medicine and physiology, however, in molecular biology, only few attempts have been made to infer a quantitative model of gene regulation due to experimental limitations of current techniques. Indeed, whereas in engineering it is now common to measure thousands of time-points at a desired sampling rate for a physical system to be modelled, this has been very difficult in biology, where time-series data consist of very few samples.

In order to overcome the current limitations, I devised an experimental platform based on a microfluidic device, a time-lapse microscopy apparatus and, a set of automated syringes all controlled by a computer, that allows to provide a time varying concentration of any molecule of interest (input) to a population of cells, and to measure the single-cell response in the form of the fluorescence level of a reporter protein, at a sufficiently high sampling rate, thus making it possible to evaluate the dynamics of the process of interest. I tested the experimental platform to implement and compare different linear and nonlinear system identification approaches to a transcriptional network in the yeast S. cerevisiae. The results I obtained confirm that the experimental system identification platform I developed can successfully be used to infer quantitative models of a eukaryotic promoter in a rapid and efficient manner. Moreover I have used the same experimental set up for the study and the *in-vivo* implementation of

novel feedback control strategies meant to precisely regulate the level of expression of a protein from the *GAL1* endogenous promoter and from a complex synthetic transcriptional network in yeast cells. The proposed effective control approach, allows to generate custom time profiles of a desired protein, and it can be exploited to study trafficking or signalling pathways and the endogenous control mechanisms of a cell.

To my Son, my Wife, my Parents, my whole Family and,
to all those who have always believed that
I would have deserved and achieved this goal.

Contents

1

Introduction

1.1 Background and motivation

The aim of System Identification, an important branch of Systems and Control Theory, is to derive a dynamical input-output model of a physical system of interest from measurement data. The model can then be used to predict the behaviour of the system to an unknown input or to derive and implement control strategies to steer at will its dynamic behaviour towards a predefined goal. This field is quite advanced and a well established theory has been developed in the case of Linear Systems (i.e. systems described by a set of ordinary linear differential equations) (1). Nevertheless nonlinearity is generic in nature and many practical examples of nonlinear dynamic behaviour have been reported in the engineering literature (2); modelling and identification of nonlinear dynamic systems is a challenging task because the principle of linear superposition does not (generally) apply to nonlinear systems, therefore heuristic methods are required for their identification (3).

In biology, system identification can be used with a dual purpose in mind: (1) as in the case of control engineering, to design feedback control strategies to steer the biological system towards a desired goal (i.e. a desired protein concentration) (4, 5, 6, 7, 8, 9); (2) to understand the biological mechanisms underlying the biological process. In both cases the dynamical nature of biological processes is a crucial feature that needs to be captured by system identification.

Current experimental techniques are, in general, suitable to assess steady-state behaviour of molecular pathways, or to measure a few time points during a time-course experiment, thus making these approaches unsuitable for System Identification purposes. Hence, only few attempts have been made to infer models of transcriptional regulation using system identification techniques (10, 11). Indeed, the majority of the models in systems biology have been built using *a priori* knowledge of the underlying chemical and genetic mechanisms (12, 13, 14, 15).

In order to overcome the current limitations, I devised an experimental platform based on a microfluidic device, a time-lapse microscopy apparatus and, a set of automated syringes all controlled by a computer. Microfluidics allows to grow cells and to precisely change their environmental conditions in real-time; moreover the cells in the device can be imaged with the microscope at high sampling rate (thus overcoming limitations of standard techniques), in order to evaluate the effects of the input provided to the system. This is achieved by measuring over time the fluorescence of a reporter used to track the output of the phenomenon of interest. The number of measured outputs relies on the number of different colours that can be tracked at the same time by fluorescent microscopy (up to 4 can be easily quantified).

On the other hand, Control engineering has been applied as a powerful theoretical framework to elucidate the underlying principles driving gene networks(16, 17, 18, 19), to predict their dynamics and their robustness to noise(20, 21), and to theoretically demonstrate the possibility of steering gene network dynamics(22, 23, 24).

More recently, other groups have reported experimental applications of control engineering to drive gene expression from artificial inducible promoters by means of external stimuli (e.g. light or osmotic pressure) either in single cells, or across a cell population(4, 5, 6, 7, 8, 9)

Here I propose the study and the *in-vivo* implementation of novel feedback control strategies meant to precisely regulate the level of expression of a protein from the *GAL1* endogenous promoter and from a complex synthetic transcriptional network in yeast cells. This control approach, namely the ability of generating custom time profiles of a certain protein, can be exploited to study trafficking or signalling pathways and the endogenous control mechanisms of a cell.

1.2 Thesis outline

This manuscript is organised as follows:

1. **Chapter 2:** I provide an overview of the disciplines, and of their fundamental concepts, used to complete this study
2. **Chapter 3:** I introduce the state of the art of the application of Control Theory principles, to the analysis and the control of biological systems.
3. **Chapter 4:** I provide details of the experimental platform for the external intervention on living cells, that I have designed and developed during this study.
4. **Chapter 5:** I describe the results achieved by using the devised experimental set up for System Identification purposes. The procedure for the inference of several mathematical description for the *GAL1* promoter in *Saccharomices cerevisiae* is reported together with the models of a synthetic gene network embedded in yeast cells (11) called IRMA and, of a synthetic circuit integrated in mammalian cells (25) both analysed and used in this study.
5. **Chapter 6:** I propose the design and the implementation of *in-vivo* feedback control strategies to regulate in real-time gene expression in populations of living cells, from endogenous promoters as well as complex synthetic gene networks.
6. **Chapter 7:** I discuss the design of a feedback control strategy for an inducible synthetic circuit in mammalian cells.
7. **Chapter 8:** I detail all the materials employed and the methods developed in this work.

8. **Chapter 9:** I discuss the results of this study, together with proposing directions for their future extensions and applications.

2

Preliminary notions

In this chapter, I provide an overview of the disciplines and of the tools adopted to complete this study. I introduce concepts of biological networks such as motifs and modularity and, the founding principles of System Identification (used to infer models of biological systems and networks) and of Control Theory (employed to devise strategy meant to regulate transcriptional processes towards desired targets and behaviours).

2.1 Transcriptional network motifs

The minimal unit of life, the cell, is a very complex and dynamical environment. At the molecular level, all chemical reactions and physical interactions are determined by the laws of thermodynamics and, by stochasticity. Moreover tens of thousands of genes encode for even more proteins and RNAs, all having the potential to diffuse and interact dynamically with each other. One way of dealing with this complexity is to operate a classification of these interactions, according to arbitrary, logic criteria such as the nature of the molecule (e.g. protein, or gene etc.), its function (e.g. kinase, or transcription factor etc.) or the cellular function that it acts upon, i.e. its associated pathway.

The instructions of life are written in the genetic code, and the key to the transmission of these instructions lies in the interaction between the code itself and all the molecules that are able to read it: transcription factors. Transcription factors are proteins that are able to recognise specific DNA sequences, i.e.

elements, and either modify the accessibility of the chromatin or recruit the apparatus that activates or represses gene expression. Information processing through signaling events inside the cell transduce a plethora of external and internal signals into modifications of transcription factors, thus deciding whether, when and how a gene has to be expressed.

Exploiting graph theory, biological interactions can be represented as networks, in which each molecule (protein, DNA, RNA, metabolite) is a node, and each association among molecules is an edge. These networks can be built according to different criteria; protein interaction networks are realised by considering physical interactions among proteins; metabolic networks are built by taking into account chemical transformations among metabolites. Furthermore the associations and interactions of large molecules that determine gene expression are instead represented by transcription regulatory networks. In these networks, each node is a transcription factor or a gene, and there are two types of edges, representing either activation or repression; an edge between two nodes indicates a direct interaction. We can observe different levels of complexity inside transcription regulatory networks: the first, basic level is composed by the direct interaction between a transcription factor and its target; at the next, but still local, level, there are motifs; then we observe modules, which are the first entities of the network to have some sort of functional independence; and finally there is the whole network (Figure 2.1). These networks rely primarily on protein-protein and protein-DNA interaction, although chemical modifications and small molecules take part in these regulations as well. These interactions cause the colocalisation of specific components that, together, are able to modify the transcriptional state of a gene.

2.1.1 Motifs

The transcriptional network can be locally divided into regulatory motifs. A motif is the unit of network architecture; each motif is characterised by a specific pattern of regulation (i.e. edges) among transcription factors and targets (i.e. nodes). It is important to notice, however, that a motif is not a functionally independent

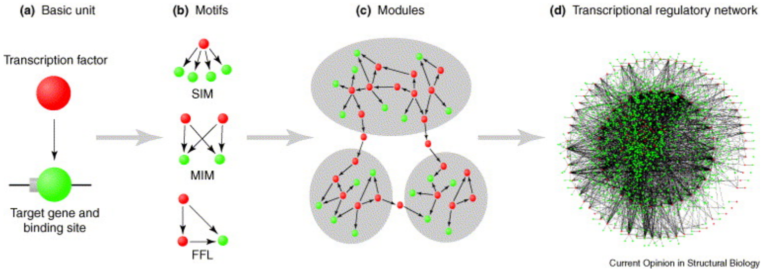

(a) Basic unit
(b) Motifs
(c) Modules
(d) Transcriptional regulatory network
Transcription factor
Target gene and binding site
SIM
MIM
FFL
Current Opinion in Structural Biology

Figure 2.1: Levels of complexity of transcription regulatory networks. The first level is composed by the simple transcription factor - target gene interaction (a); the second level of complexity is represented by motifs (b); the third level is composed of modules that carry out specific functions (c); the final level is the whole cellular network (d). From (26).

unit; nonetheless, it has been demonstrated that each motif is characterised by particular kinetic properties that regulate the temporal expression and function of the nodes in their network (reviewed in (27)). The most common examples of motifs are represented in Figure 2.2.

Autoregulation, or the Feedback Loop : Autoregulation is the ability of a transcription factor to act on its own expression, and can be either positive or negative. It was shown that roughly 10% of *S. cerevisiae* genes that encode for regulators are subject to autoregulation (28). Negative autoregulation, or Negative Feedback Loop (NFL) occur when a transcription factor represses its own expression, and has been shown to have several dynamical properties.

The NFL speeds up the response time of a circuit: comparing two expression cassettes, one with a NFL and one simple transactivation, in order to get to the same steady state level of expression, the NFL allows the use of a strong promoter, which cause a steeper increase in the production rate of the mRNA; on the contrary, a simple transactivation must use a weaker promoter to get to the same level of expression. It can be mathematically demonstrated that the rise-time of a negatively regulated system is significantly faster than in a non-regulated system, even with a stable gene product (29). Another important consequence of the NFL is that it reduces variability in cell-to-cell protein levels: a high production rate will self-limit, while a low production will allow more

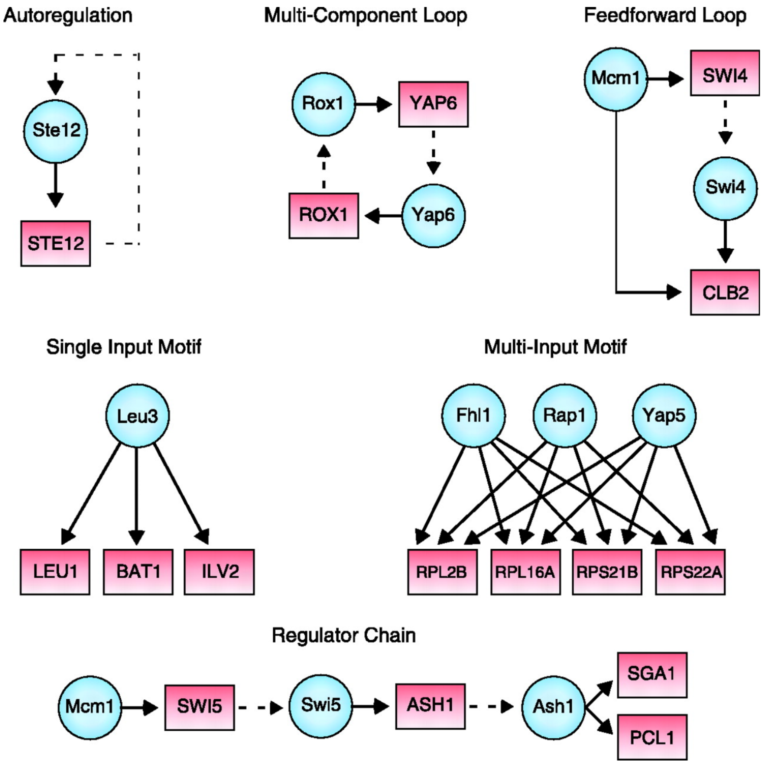


Figure 2.2: Examples of the most common transcriptional motifs. The activity of a transcription factor on its own gene constitutes a Feedback Loop, or Autoregulation, and it can be either positive or negative. The multi-component loop provides a feedback as well, but through two or more factors. In the Feedforward Loop, a gene regulates a target, and together they target another gene. The Single Input Motif consists of a transcription factor regulating multiple targets, while in a Multi-Input Motif, multiple transcription factors regulate the same targets. Finally, we have a Regulator Chain when a transcription factor regulates downstream genes through a cascade of regulations. From (28).

transcription, making the distribution of protein levels among cells more narrow (30).

The Positive Feedback Loop (PFL) occurs when a transcription factor activates its own transcription. The dynamical properties of the PFL are inverse to those of the NFL: usually it slows down the response time and increases cell-to-cell variability. This system is slow because, for low levels of the transcription factor, the protein is not able to overcome its own activation threshold and the gene tends to be switched off. However, any perturbation that allows the protein to overcome its threshold rapidly translates into autocatalytic activation. This property is also responsible for the increase in variability among a population of cells, that often results in bi-modality if the PFL is very strong, as was demonstrated in *E. coli* (31) and more recently in mammalian cells (32) .

The Multi-Component Loop : When a transcription factor regulates its own transcription through one or more other regulators, we have the Multi-Component Loop. It has been predicted that 10% of yeast genes also have this type of regulation. The effects of this kind of motif are similar to those of the PFL and NFL (33).

The FeedForward Loop : A FeedForward Loop (FFL) is a motif in which a transcription factor regulates another one, and together they also regulate a third factor. If they both cause the same effect on their common target, we have a Coherent FeedForward Loop (C-FFL); otherwise, there is an Incoherent FeedForward Loop (I-FFL). The C-FFL can behave as a persistence detector, showing a delayed response after stimulus addition, but no delay after stimulus removal, thus acting as a sign-sensitive filter; the I-FFL is able to accelerate the response time, and to ultimately produce a pulse-like profile response of the target gene (27).

Single- and Multi-Input Motifs : Single- and Multi-Input Motifs (SIMs and MIMs) both regulate a set of target genes. This kind of regulation occurs usually when a whole pathway or biological function has to be activated. SIM is the kind of regulation that is used when a single signal is sufficient to activate the pathway, for example the activation of the Leu3 gene in yeast to induce the set of genes responsible for Leucine biosynthesis (28), or the Transcription Factor EB inducing lysosomal biogenesis in mammalian cells (34). When multiple signals activate the same pathway, the MIM is used, such as in the well-studied pathway of the heat shock response, which integrates different stress signals into the transcription of the same set of Heat Shock Proteins (35).

Regulator Chain : The regulator chain motifs are three or more transcription factors that activate one another in a sequential order. This kind of regulation is the most used when there is the need to temporally regulate a process; famous examples are muscle differentiation (36) and the cell cycle regulation.

2.1.2 Modularity of networks

An important property of networks is their modularity. Since the genetic code is universal and conserved among almost all taxa, a gene can be easily transferred from an organism to a different one (provided that the receiving cell has the whole apparatus for correctly expressing and modifying the gene product). A whole network, carrying for example the components to accomplish a specific function, can be transferred from an organism to another, or different components can be rewired to obtain novel networks and different functions. Horizontal gene transfer in bacteria is a natural-occurring example of the former case, in which bacteria obtain whole new function (such as the ability of metabolise different molecules, or invade new hosts) by absorbing exogenous DNA; thus modularity is also a powerful tool for evolution, allowing life to experiment on itself and to preserve only those new interactions that provide a selection advantage. The tools of Genetic Engineering and Synthetic Biology allow the experimenter to modify and displace whole sets of genes as well: this can be advantageous both to isolate and study functions out of their natural context, and to transfer them to new organisms (such as pesticide resistance in OGM crops).

2.2 System Identification

System Identification is a discipline dealing with the problem of developing computational approaches to derive, from measurement data, a quantitative dynamical model of a dynamical physical system able to predict its future behaviour.

The word *"system"* defines a process in which variables of different types, interacting each other, generate observable signals. The interesting observable variables are called *outputs*. Those external stimuli, affecting system dynamics, that are manipulable are called *inputs*, others, not controllable, are named *disturbances*, these can be measured directly or they can be estimated from the output (Figure 2.3 Panel A). A model is a set of mathematical equations η that can be defined as follows:

$$y = \eta(u, d, k_i) . \tag{2.1}$$

Where u represents the input(s), and d the disturbances acting on the system , y is the system output (Figure 2.3 Panel B). The general aim of System Identification, is to estimate the function η and the values of the parameters k_i, starting from measurement data (i.e u and y).

The model inference procedure consists of three main stages (1):

- **The data**: input-output data are measured in a specifically designed identification experiment. The user has to decide the measure of which variables and with which sampling rate it is necessary to carry out, in order to have data as much informative as possible.

- **The set of models or the model structure**: the user can define a set of candidate models within which the identification process has to look for a suitable one. Model sets with adjustable parameters with physical interpretation are defined as *grey boxes* (their inference is the so called "grey box identification"); whereas models whose parameters are interpreted only as a mean to fit measured data are named *black box* (the "black box identification" is carried out to estimate their parameters).

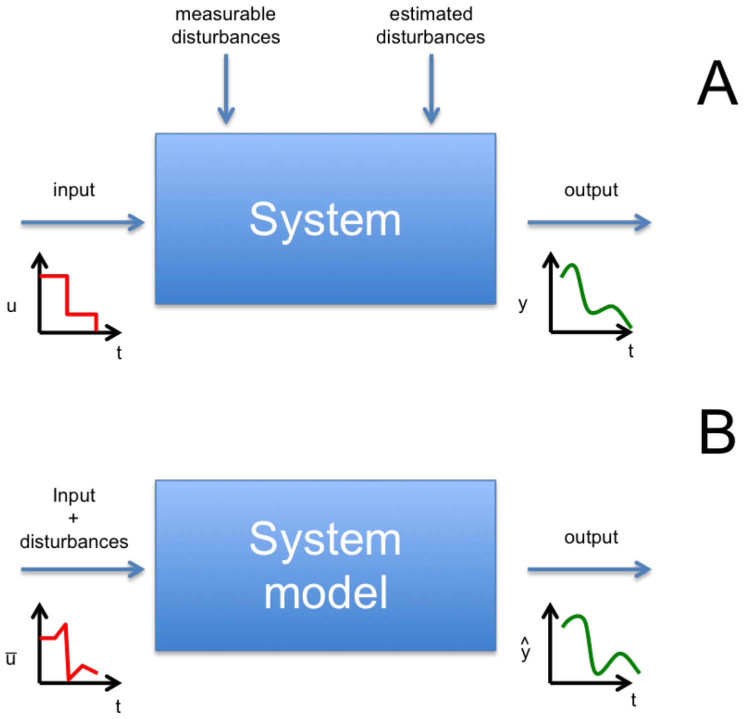


Figure 2.3: System Identification paradigm. *(A) The interaction of external stimuli (input u), with measurable and estimated disturbances, lets the system produce observable variables (output y). (B) The system model, inferred starting from the available measured data, has the property of reproducing an output $\hat{y}$ that is as close as possible to system output y, when it is stimulated with the same signals (inputs and disturbances, $\bar{u}$) acting on the system itself.*

- **Determining the "best" model in the set, according to the data**: the assessment of model quality is usually performed evaluating how they can predict future values of the output from the past values of the input and output.

Once the best model has been identified, among all the candidates, it has to be tested to check whether it is valid for its purpose. This *model validation* step, can be performed according several criteria relying on the physics of the process being modelled or on the capability of reproducing data sets different from those used for model inference.

In Biology, System Identification can be applied to acquire new insights in the biological mechanisms underlying the biological process under exam. The crucial feature, which needs to be captured by System Identification procedures, is the dynamical nature of biological processes hence, the inferred models should be able to reproduce the dynamic behaviours over the time of all the variables of interest belonging to the modelled system.

2.3 Control Engineering

2.3.1 Negative feedback

Control Engineering is a discipline whose aim is to control a dynamical system so that its output follows a desired behaviour, by appropriately choosing its input. The approach employed to accomplish this task is the *negative feedback* (37); the variable to be controlled (system output y) is measured and its value is subtracted from the desired value (control reference r). The quantity thus obtained, the feedback error e, is minimised by the controller by choosing an input u in order to guarantee that the output y matches the desired reference r (Figure 2.4). This control approach is the "'closed loop control"', its counterpart is the "'open loop control"' where the control action is exerted without any measurment of the system output, therefore the input u is pre-computed and not calculated on the basis of the control error e.

One of the simplest and most famous examples of engineered feedback control systems, is the the thermostat; this device measures the temperature in a building, compares it with the desired temperature and uses the resulting control to decide whether to turn heat on, if the temperature is too low or, to turn it off, if the temperature is too high.

Negative feedback is an intrinsic mechanism highly exploited and conserved in Nature. Ecosystems, due to the complex interactions among animals and plants, show a plethora of examples of feedback, as well as, global climate dynamics which, depends on the feedback between the atmosphere, the oceans, the land and the sun. Another significative example, at a smaller scale, is the regulation of glucose in the bloodstream through the production of insulin and glucagon by the pancreas. The body attempts to maintain a constant concentration of glucose,

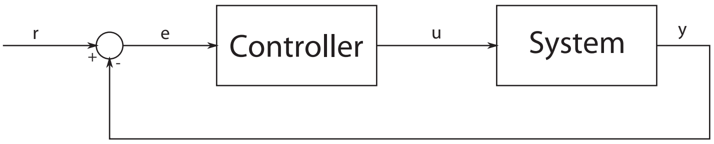


Figure 2.4: Negative feedback paradigm. *The founding principle of Control Engineering is the negative feedback paradigm. The quantity to be regulated (system output y) is fed back and subtracted from the reference signal r. This difference results in the feedback error e, which is the quantity to be minimised by the controller to ensure that the output y matches the reference r. To this end the controller calculates an appropriate signal u to steer the system output towards control reference*

which is used by the body's cells to produce energy. When glucose levels rise, the insulin is released and causes the body to store excess glucose in the liver. Whereas, when glucose levels are low, the pancreas secrets the hormone glucagon, which has the opposite effect. Insulin and glucagon secretions throughout the day helps to keep the blood-glucose concentration constant to physiological levels.

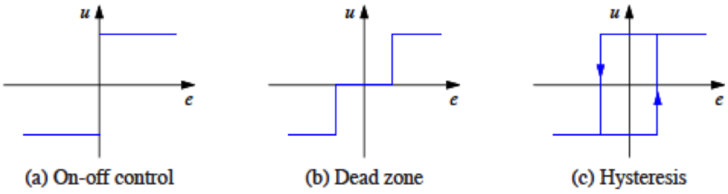
2.3.2 Simple Feedbacks

The principle of negative feedback relies on determining correcting actions (signal u) on the basis of the difference between desired (r) and actual output (y). This task can be accomplished in different ways. The benefits of feedback can be obtained by very simple feedback laws such as on-off control (relay control), proportional control and proportional - integral - derivative (PID) control.

On-off control: This simple feedback control strategy can be expressed as follows:

$$u = \begin{cases} u_{max} & \text{if } e > 0 \\ u_{min} & \text{if } e < 0 \end{cases}$$

where the *control error* (feedback error) $e = r - y$ is the difference between the reference signal r and the actual system output y, u is the control input. This



(a) On-off control (b) Dead zone (c) Hysteresis

Figure 2.5: On-off controllers, input-output characteristics. *On-off controllers characteristics with the feedback error e on the horizontal axis and the control input u on the vertical axis. The ideal on - off controller is depicted in (a), a modified versions, with the introduction of a dead zone (b) and hysteresis (c) are also represented. From (37)*

control law implies that maximum corrective action is always used (Figure 2.5 (a)). This control strategy, although very simple, is effective in keeping the system output very close to the reference; typically the controlled variables oscillate around the reference, and this oscillations are acceptable if they are sufficiently small. This control action is employed by the thermostat controlling the temperature of a room. To reduce the number of switches of the control input (physical controllers such the thermostat are designed to last over a defined number of commutations), a dead zone (Figure 2.5) or a hysteresis (Figure 2.5) can be added to the on-off controller. In the case of the introduction of a hysteresis the control input depends on the past values of the control error.

PID control: On-off controllers could often lead to oscillations of the controlled variable, due to the over actuation performed by this control law; even a small variation in the control error results in a variation of the control variable u over its entire range. One of the solutions that can be adopted to sort out this issue is the *proportional control*, where the control input u is proportional to the control error for small errors. This can be achieved with the control law:

$$u(t) = \begin{cases} u_{max} & \text{if } e \geq e_{max} \\ u_p & \text{if } e_{min} < e < e_{max} \\ u_{max} & \text{if } e \leq e_{min} \end{cases}$$

where k_p is the controller gain, $e_{min} = u_{min}/k_p$ and $e_{max} = u_{max}/k_p$. The behaviour of the controller is linear when the error is is in the interval (e_{min}, e_{max}):

$$u_p(t) = K_p(r - y) = K_p e(t) \quad \text{if } e_{min} \leq e \leq e_{max}. \tag{2.2}$$

The major drawback of proportional control, is that it is not able to guarantee that $e = 0$, but only that e is bounded. To overcome this limitation it is necessary to take into account the entire "history" of the control error, namely the control input has to be proportional to the integral of the error:

$$u_i(t) = K_i \int_0^t e(\tau) d\tau. \tag{2.3}$$

This control form is called *integral control*, and K_i is the integral gain. It can be demonstrated that a controller with integral action has zero steady-state error. The catch is that there may not always be a steady state because of the time varying reference or due to oscillations in the system. To further improve control performances, the controller can be provided with the predictive feature of a term proportional to the derivative of the error:

$$u_d(t) = K_d \frac{de(t)}{dt}. \tag{2.4}$$

Putting together proportional, integral and derivative control, the result is a controller expressed as follows:

$$\hat{u}(t) = u_p(t) + u_i(t) + u_d(t) = K_p e(t) + K_i \int_0^t e(\tau) d\tau + K_d \frac{de(t)}{dt}. \tag{2.5}$$

The control action is thus a sum of three terms: the past as represented by the integral of the error, the present as represented by the proportional term and the future as represented by the derivative term. This form of feedback is called a *proportional-integral-derivative (PID) controller* and its functioning is illustrated in Figure 2.6

A PID controller is very useful and is capable of solving a wide range of control problems, indeed more than 95% of all industrial control problems are solved by PID control (37).

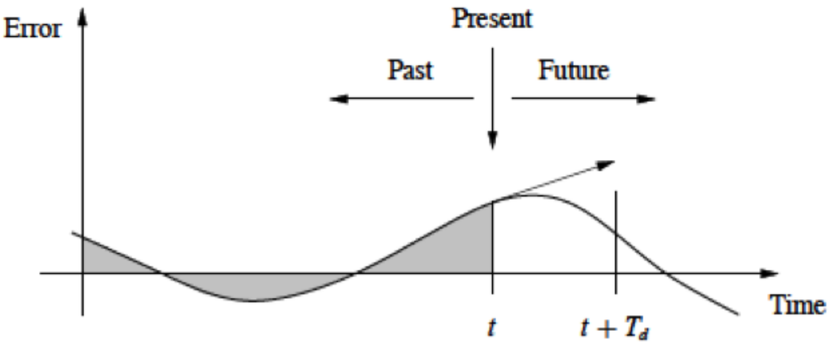


Figure 2.6: PID controller functioning. *At each sampling time the control input $\hat{u}$ is calculated as the weighted sum of three contributions: a) the proportional term (P), that at time t depends on the instantaneous value of the error, b) the integral term (I) that is based on the integral of the error up to time t (shaded area) and c) the derivative term (D) that provides an estimate of the error trend over the time. From (37)*

.

Actually, in the majority of the cases, the derivative action is not used either because the control references usually vary slowly, so that a prediction term is not necessary and also because in presence of noisy measurements, the control action would become noise due to the high pass filter nature of the derivative term. Moreover, although the proportional controller can be used alone, it is never applied together the derivative control without the integral term; a proportional derivative controller would not reach the control reference since it cannot guarantee the stability of the closed - loop system (37).

Integrator windup In practical control implementations the control signal $\hat{u}(t)$ is fed to the process being regulated by means of actuators (i.e. motors, valves, pumps). These components have physical limitations: motors have limited speed and acceleration, valves cannot be more than fully opened or fully closed and pumps cannot go slower than stopped. Thus the control signal acting on the system is saturated between the minimum and the maximum values achievable with the actuator used.

If the control variable $\hat{u}$ passes the saturation limits, the actuator will con-

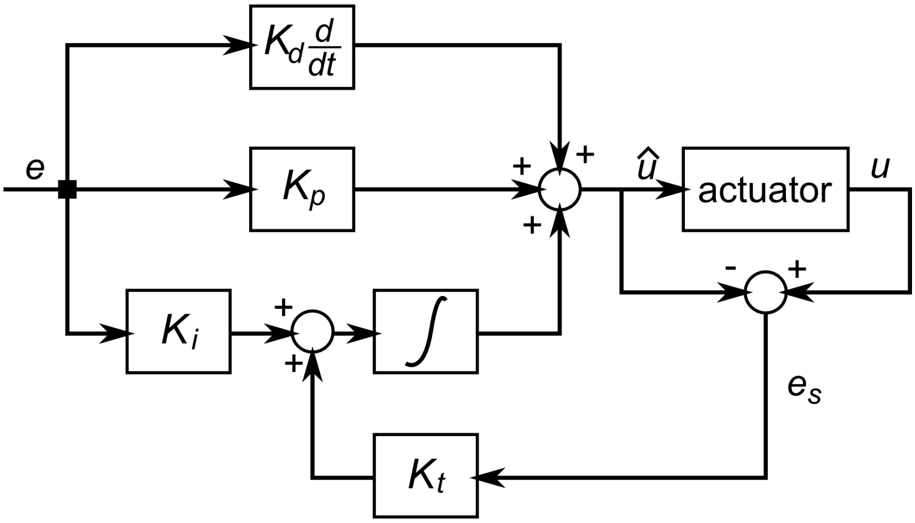


Figure 2.7: PID controller with anti-windup. *The input to the integrator consists of the error term plus a "reset" based on input saturation. If the actuator is not saturated $e_s = u - \hat{u} = 0$, otherwise e_s will decrease the integrator input to prevent windup*

.

stantly run at its saturation limits despite system output, thus the feedback loop is broken and the error is nonzero. In the case of PI or PID feedback control strategy, the error is integrated and the integral term may become very large, hence the control signal remains saturated even if the error changes and, it may take a long time before the integrator and the controller output come inside the saturation range. This situation is called *integrator windup* and leads to large transient in system response.

To avoid windup it is possible to modify the control scheme as proposed in (37), by adopting an additional feedback from the actuator output to the integrator (Figure 2.7): e_s, namely the difference between the actuator output u and the output of the controller $\hat{u}$, is fed to the input of the integrator through a gain $K_t = \frac{1}{T_t}$. When the actuator saturates, the signal e_s is different from zero and it decreases the integrator input with a time constant T_t, thus resetting the integrator output and leading the signal $\hat{u}$ within the saturation limits.

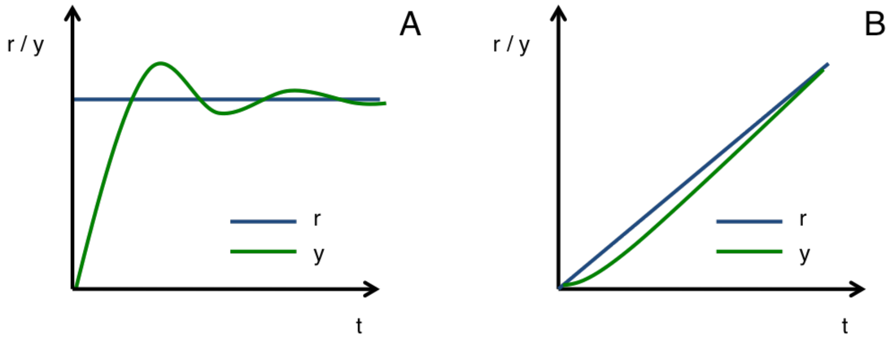


Figure 2.8: Control objectives. Panel A: *In the* **set - point control**, *the reference signal r (blue line) is a fixed value over the time and the output y (green line) is regulated to reach and maintain that level.* Panel B: *in the* **signal tracking control** *the reference r (blue line) is a time varying signal and the output y (green line) is regulated to follow his trend over the time.*

2.3.3 Control Objectives

Control Engineering principles can be applied with two main control objectives:

- **Set - point control:** the reference signal r is a constant value over the time and the output y is regulated to reach and maintain that level (Figure 2.8 A); one example of this control objective is found when, using a thermostat, the temperature of a room is kept constant over time.

- **Signal tracking control:** the reference r is a time varying signal and the output y (green line) is regulated to follow his trend over the time (Figure 2.8 B); a typical application of signal tracking control is represented by vehicles driven by an autopilot that has to follow a given trajectory over time.

3

Control Theory and Biology

A crucial feature of biological systems, is their ability to maintain homeostasis despite fluctuations in their environment. The compensation of changes in external conditions is achieved by means of naturally embedded negative feedback loops in the cell. Synthetic Biologists, aiming at building novel functions and biological circuits within cells, are exploiting principles of System Identification and Control Engineering to wire synthetic control feedback loops, in order to regulate the behaviour of cellular processes.

Here I propose an overview of feedback regulation systems either embedded within cells or implemented as external controllers.

3.1 Synthetic embedded control schemes

The ability of building synthetic negative feedback loop control scheme within cells, can be instrumental to confer robustness and to increase the yield of metabolic processes. This for instance, is the case of biosynthesis from microbial cells populations. These processes are often limited by metabolic imbalances, that could be dynamically regulated by the use of molecular feedback controllers, able to adapt their products on the basis of the state of the hosting cell.

Furthermore, in the future, engineered feedback loops could be delivered to host organisms via bacterial or viral infection as a therapeutic approach for genetic diseases or metabolic disorders.

To this end the synthetic circuit should comprise a block that can sense changing in intra-cellular or extra-cellular products or conditions (i.e. the sensor), a regulator that can dynamically adjust gene expression according to bio-signals produced by the sensor (i.e. the controller) and by a gene regulatory network, or a signalling cascade, able to set and store the control reference.

Although endogenous sensing and actuation mechanisms are already available and well described, very few attempts have been performed, so far, to compose them in feedback controllers. This is manly due to the fact that endogenous regulators evolved to control natural pathways with low fluxes hence they are not easily modified to work on engineered pathways.

One impressive application was carried out by Zhang and colleagues who developed a dynamic sensor-regulator system (DSRS) for the efficient production of biodiesel in the form of fatty acid ethyl ester (FAEE) by *E. coli* (38).

They took advantage of a FAEE biosynthetic (39) and integrated in its genome a control system to improve its production yield, by controlling the expression of enzymes involved in the synthesis process. Zhang *et al.* engineered a biosensor for free fatty acids which are key intermediates in the FAEE biosynthetic pathway, and a system of DNA regulatory elements to produce FAEE when fatty acids accumulate. In Figure 3.1 a complete scheme of the engineered pathway is depicted. The dynamic sensor-regulator system contains the repressor gene *fadR* and two promoters P_{modB} and P_{modC}; when there are not fatty acid accumulated, then the expressed FadR represses P_{modB} and P_{modC}, thus inhibiting the synthesis of ethanol and acyl-CoA. When fatty acids accumulate, they are activated to acyl-CoA by FadD and then acyl-CoA binds to FadR and activates the biosynthesis of ethanol and more acyl-CoA and the expression of wax-ester synthase that converts ethanol and acyl-CoA to FAEE.

Authors demonstrated that this ON-OFF control strategy improves the yield of production of FAEE in comparison with the biosynthesis achieved with the AEE strain (39) since the production of FAEE from glucose is activated only upon a sufficient accumulation of fatty acids.

Another interesting feature, that can be implemented in biological system, is the ability to track given biomolecular signals in response to external stimuli.

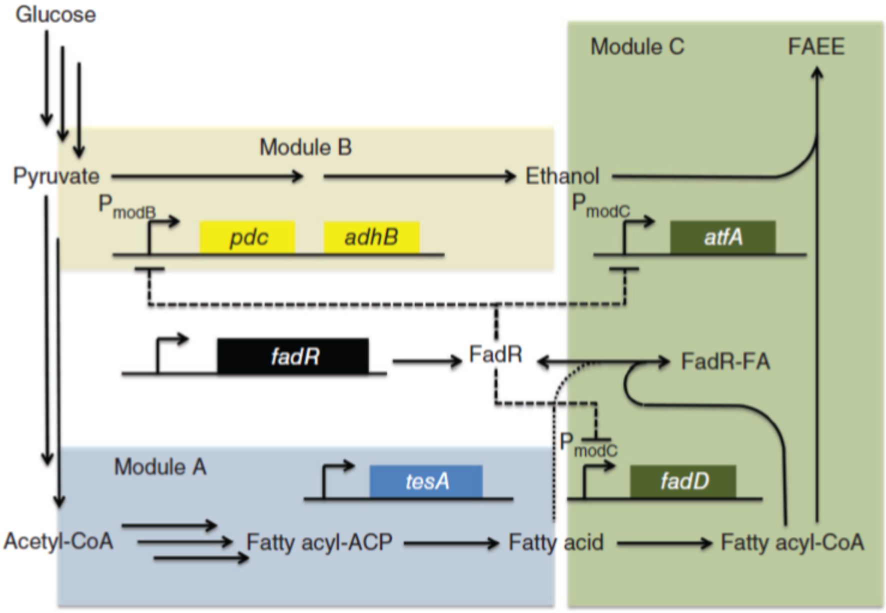


Figure 3.1: Dynamic sensor-regulator system for biofuels production *The engineered pathway was built starting from a FAEE biosynthetic pathway already developed and described in (39). With the addition of the repressor gene fadR and of the two FadR controlled promoters P_{modB} and PmodC, the starts producing FAEE only upon a sufficient accumulation of fatty acids. From (38)*

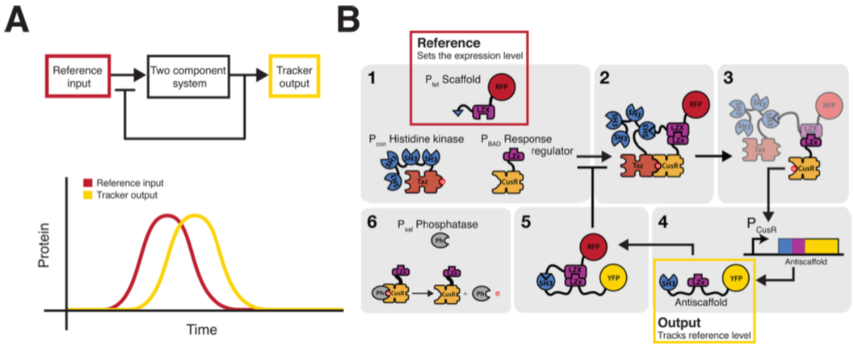


Figure 3.2: Protein concentration tracker *(A) Conceptual scheme of the concentration tracker as a negative feedback loop. (B) Negative feedback loop implementation via the two-component scaffold/anti scaffold protein system. From (40)*

Hsiao *et al.* implemented a synthetic biomolecular concentration tracker in bacteria, able to robustly reproduce a reference signal generated in response to an external stimulus (40).

The circuit designed and implemented, employs a negative feedback control scheme to accomplish the signal tracking task (Figure 3.2 A); it is based on a two-component scaffold/anti scaffold protein system: it regulates the production of the amount of a target protein (anti scaffold - YFP) with respect to the reference protein (scaffold - RFP) (Figure 3.2 B). YFP expression depends on the amount of the free scaffold available since the control target contains domains to sequester free scaffold thus implementing a negative feedback loop. The scaffold protein, the reference, is under the control of an inducible promoter whose activation is used to set the control objective.

Authors described the dynamics of the system deriving a dynamical mathematical model able to predict circuit behaviour and confirmed, with *in-vivo* experiments, the effectiveness of the signal tracker to reproduce different reference signals. The advantage of the proposed strategy resides in the fact that the small size of the scaffold and anti scaffold proteins allows this two-component system to be used to dynamically regulate the concentration of larger proteins. Moreover this strategy can be applied to mimic endogenous biological clocks, by clamping this two-component system to naturally oscillating genes.

3.2 External control of intra-cellular dynamics

The control of cellular environmental conditions, by means of external intervention, has been intensively adopted to optimise and guarantee living conditions to cells. This can allow, for instance, to regulate, at least indirectly, the yield of bio-reactors and fermentators, in which the amount of the produced outcome is proportional to the fitness and the viability of the cells growing in them.

On the contrary, it is only recently that the application of Control Engineering principles to the regulation of endogenous and synthetic biological circuits has been achieved.
To accomplish this task, the availability of a sensing apparatus, able to measure intracellular dynamics (e.g. gene expression and protein localisation) and to quantify their deviations from desired values, is crucial; moreover the control feedback loop has to be equipped with a system able to exert corrective actions on the biological system of interest on the basis of those deviations.

Many attempts to control cellular dynamics have been performed and, they all rely on the measure of a fluorescent proxy of the variable to be controlled whereas, they differ for the methodology adopted to feed the control input to cells. Microfluidic devices, allowing a tight control of cellular growing medium and administration of inducer molecules, have been successfully employed to investigate synchronization properties of synthetic biological clocks in bacterial cells (4) and, to control the transcription from an endogenous osmostress promoter in yeast *S. Cerevisiae* (7). Whereas the use of light stimuli has been used to control gene expression in yeasts (5, 8), to regulate intracellular signalling dynamics in mammalian cells (6) and to drive protein levels by using light-switchable two-component systems in bacteria (9).

Thus, the state of the art of the actuation strategies implemented to deal with the control of intra-cellular dynamics, can be divided into two major strands: a) microfluidics-based actuation and b) optogenetics-based actuation.

3.2.1 Microfluidics - based actuation

Microfluidics is a discipline born by applying the design principles of microelectronic circuits to the design of fluid and droplet dispenser systems at the microliter scale. Channels and valves used to deliver fluids are designed and implemented exploiting the same concepts that are behind the realisation of electronic boards and circuits.

The first microfluidic product commercialized in early $90s$ of the XX Century on a large scale, was the inkjet printer head today installed in all the inkjet printers (41).

Nowadays, microfluidic devices are being massively used in chemistry and biology since they allow to tightly regulate the concentrations and the administration of chemicals compounds to cells (as for the ink on paper sheets), thus giving the opportunity to maintain cells and biological samples in their ideal physiological conditions. Moreover the application of microfluidics to biological research is advantageous to perform high-throughput experiments, thus having a huge amount of data to be collected and analysed for several purposes, with the significative advantage of using very small volumes of reagents. The analysis and the regulation of intracellular dynamics can be achieved by means of microfluidics devices keeping cells in an insulated environment and, providing external stimuli administering the fluids fed into these chips.

Danino and colleagues (4) designed a novel microfluidic device meant to investigate synchronization dynamics in a population of *E. Coli* integrating a synthetic oscillator. The design of the synthetic circuit they built is based on the use of orthogonal biological parts implementing quorum sensing functions in *Vibrio fischeri* and *Bacillus Thurigensis*. The transcription of *luxI*, *aiiA* and *yemGFP* genes is driven by three copies of the *luxI* promoter. LuxI produces a small molecule acyl-homoserine lactone (AHL) that can diffuse across the cell membrane into surrounding cells thus activating the *luxI* promoter; aiiA, negatively regulates the same promoter catalysing the degradation of AHL (Figure 3.3 a). The topology of the network, comprising a delayed negative feedback loop, leads to oscillations, as long as the diffusion of the AHL is effective; this diffusion, and thus the coupling among cells, is affected by cell density and proximity. To overcome this issue and to precisely control cell growth, Danino and colleagues used a custom microfluidic device housing a main channel feeding nutrients to a rectangular trapping chamber (Figure 3.3 b). Size and shape of the chamber, as well as of the channel, guaranteed to achieve the ideal cell density and AHL diffusion to

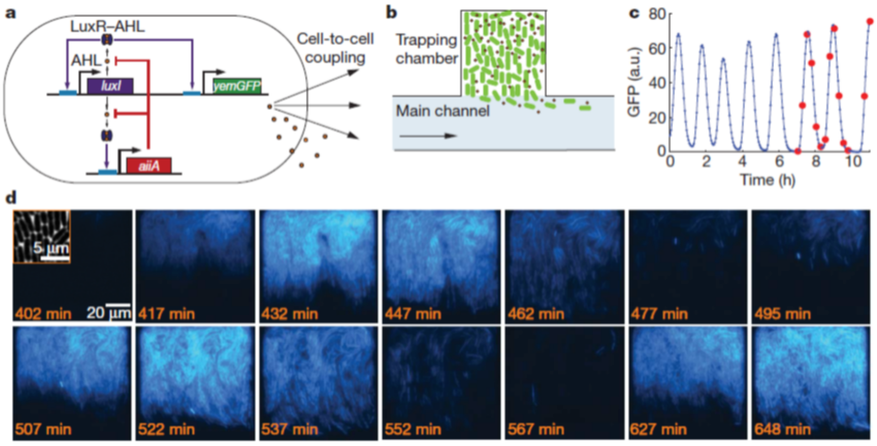


Figure 3.3: Control of synchronized genetic clocks*(a) Synthetic circuit topology.LuxI produces the AHL molecule that, diffusing to neighboring, activates the transcription from the luxI promoter; aiiA catalyse the degradation of AHL thus exerting a negative regulation of the same promoter (b) Scheme of the microfluidic device adopted to maintain cells at a constant density (c) Quantified fluorescence of the reporter during oscillations, red dots refer to the images reported in d). (d) Images taken demonstrate the synchronization of oscillations among the entire bacterial population.From (4)*

observe synchronised oscillations of the fluorescence produced by the fluorescent reporter (Figure 3.3 c-d). Authors demonstrated, by performing several experiments and producing a quantitative mathematical model, that the period and the amplitude of synchronised oscillations can be modulated controlling flow rate in the main channel; the higher is the flow velocity the higher are the period and the amplitude.

The modulation of osmotic stress to yeast cells loaded in a microfluidic chip was exploited by Uhlendorf and colleagues to implement a real-time controller of gene expression from an endogenous osmostress inducible promoter (7). Yeast response to an osmotic shock is mediated by the high osmolarity glycerol (HOG) signaling cascade ((10)). Among all the genes up-regulated in response to a hyper-osmotic stress, authors focused their attention on the *STL1* gene; they decided to integrate a fluorescent reporter under the control of its own promoter and defined as control objective the regulation of this fluorescence to fixed (set-point control) and time varying amounts (signal tracking control). To accomplish this task they assembled an experimental platform comprising a H-shaped microfluidic device to grow cells and to image them using a fluorescence microscope and a system of pumps, connected to microfluidic channels, to change growing medium osmolarity. Uhlendorf *et al.* employed a control law based on a mathematical model of the system being controlled: at each control instant the controller, on the basis of the actual state of the system, simulates, over a certain prediction horizon, the response of the mathematical model to several control inputs and chooses the input signal that, over the same horizon, minimises the error between the model output and the control reference. This iterative control strategy, can be implemented with several simulation and minimisation algorithms and it is called Model Predictive Control (MPC) (42). To implement MPC, they used a simplified two state nonlinear model of the *STL1* activation upon an osmotic shock to cells, and a state estimator (Kalman Filter (43)) to retrieve initial conditions for the simulation of the model starting from the measurement of the fluorescence.

Authors demonstrated that this control strategy was effective in obtaining fixed and time varying amount of fluorescence over thousand minutes, controlling the entire cell population as well as single cells (the controlled variable was the fluorescence expressed by cells) .

3.2.2 Optogenetics - based actuation

Control of gene expression from the *GAL1* promoter in yeast cells, has been achieved by Milias-Argeitis *et al.* using light stimuli to provide inputs to cells(5). To implement this optogenetics control strategy, they built a yeast strain integrating the light responsive Phy/PIF module (plant photoreceptor chromoprotein PhyB and Phytochrome interacting factor) from *Arabidopsis Thaliana*, in particular they fused the photosensory domain of PhyB to *GAL4* binding domain and PIF3 to *GAL4* activation domain; a yellow fluorescent reporter (YFP) driven by the *GAL1* promoter, that contains Gal4 binding sites, was used as a read out of system dynamics. The stimulation with red (650nm) and far-red (730 nm) pulses of light allowed to switch on and off, respectively, the YFP fluorescence (Figure 3.5 A).

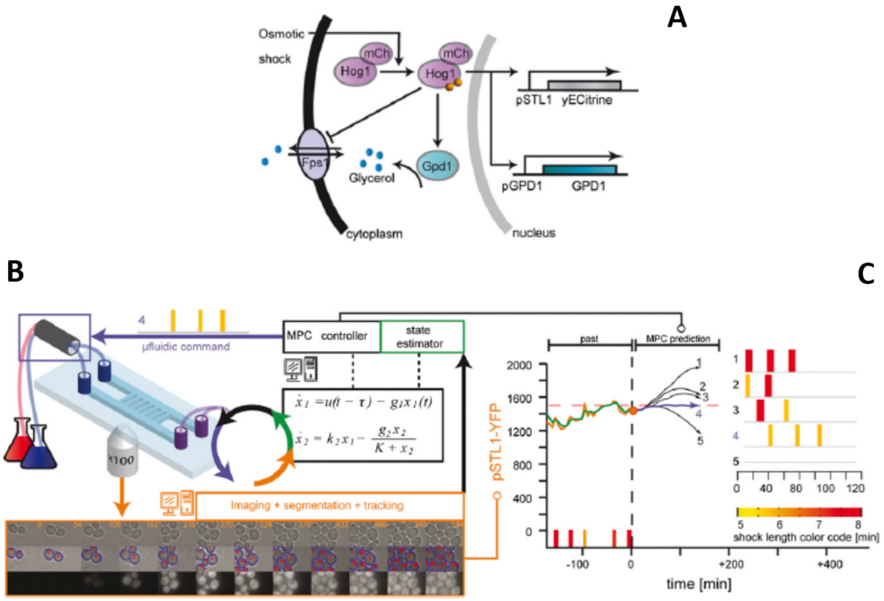


Figure 3.4: Control of gene expression from an endogenous osmostress inducible promoter*(A) HOG signaling cascade functioning and activation of the STL1 promoter upon an osmotic shock. (B) Experimental platform designed to accomplish the control task. Cells grow in a microfluidic device mounted under an inverted fluorescence microscope. Controlled pumps are connected to the chip to control the osmolarity of cell growing medium. (C) Model Predictive Control (MPC) functioning: a mathematical model of the system is simulated in response to several input. The control input applied to the system is the one that minimises the distance of the model output from the control reference. Adapted from (7)*

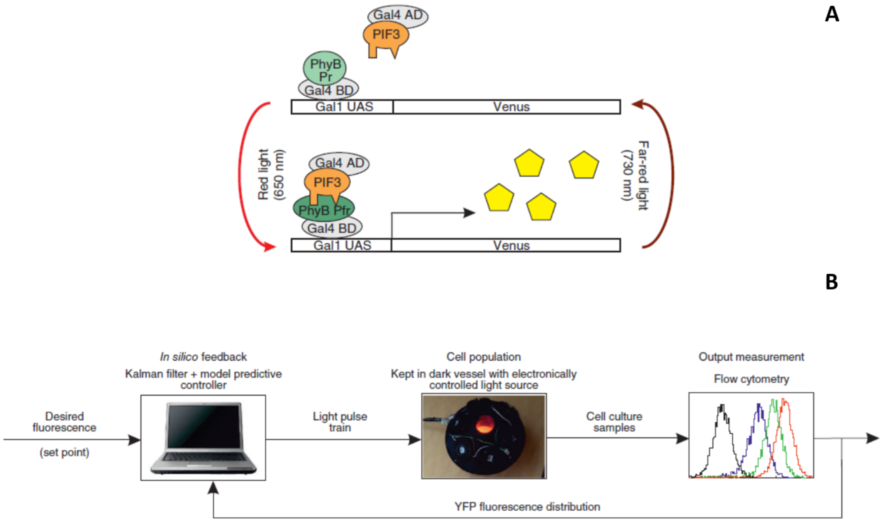


Figure 3.5: Control of gene expression from the *GAL1* promoter via optogenetics*(A) Phy is fused to Gal4 binding domain and PIF3 to Gal4 activation domain, thus upon red light stimulation and the formation of the Phy/PIF complex the transcription of the Gal4-dependent genes is activated; the complex can be divided with far-red light input to cells. (B) feedback control strategy implemented for set point control task: the average of the YFP fluorescence distribution is fed to a MPC regulator coupled with a state estimator; the controller computes the right sequence of light pulses to keep system output equal to desired value. Adapted from (5)*

The control objective was a set point regulation of the average of YFP fluorescence distribution over the entire cell population measured via flow cytometry. They firstly characterised the dynamics of this light-switchable system by providing a series of alternated pulses of red and far-red light to produce input-output data to derive a mathematical model of the process. Then they used this model to implement a MPC regulator that, they demonstrated to be suitable to accomplish the desired control objective (Figure 3.5 B).

The same light inducible Phy/PIF has been used by Toettcher and colleagues to control membrane recruitment of protein of interest in mammalian cells (6) (Figure 3.6 A). They devised a PI control strategy to administrate different ratios of red and far-red light, to specific regions of cells, to obtain user defined mem-

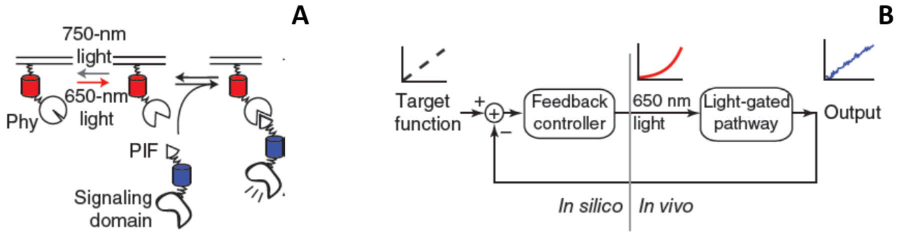


Figure 3.6: Light controlled protein membrane recruitment*(A) Light-gated system activated with red light and de-activated with far-red light. (B) Negative feedback control scheme implemented. The controller has been implemented as a Proportional Integral controller that calculates the light intensity to control the fluorescence on the cell membrane. Adapted from (6)*

brane concentration of a fluorescent tagged PIF protein measured by fluorescence microscopy. Their strategy was effective in performing both set point and signal tracking control (Figure 3.6 B).

More recently Olson *et al.* have designed an instrument, they called Light Tube Array (LTA), allowing to stimulate 64 wells, containing standard test tubes, at the same time with blue, green, red and far-red LEDs (9). They adopted this device to characterise the dynamics of two different engineered light switchable two-component systems in bacterial cells providing for them a quantitative mathematical description; then, the inferred models, have been used to compute a sequence of light pulses in order to control, in open loop, the fluorescence of a reporter regulated by these light sensors.

The two synthetic circuits studied in (9) rely on the presence of a light-switchable sensor histidine kinase containing an N-terminal phytochrome-family photosensory domain and, a C-terminal bifunctional kinase-phosphatase signaling domain. The first circuit is activated when stimulated with a green light and deactivated with red light stimulation, thus its activity is regulated by modulating the intensity of a green LED while a red LED is maintained at its maximum intensity; the second engineered system is dark-activated and red-deactivated so, its dynamics are controlled by the intensity of the red LED (Figure 3.7 A).

The authors were able to estimate the parameters of two nonlinear models describing their dynamics by performing step-response experiments of the two systems.

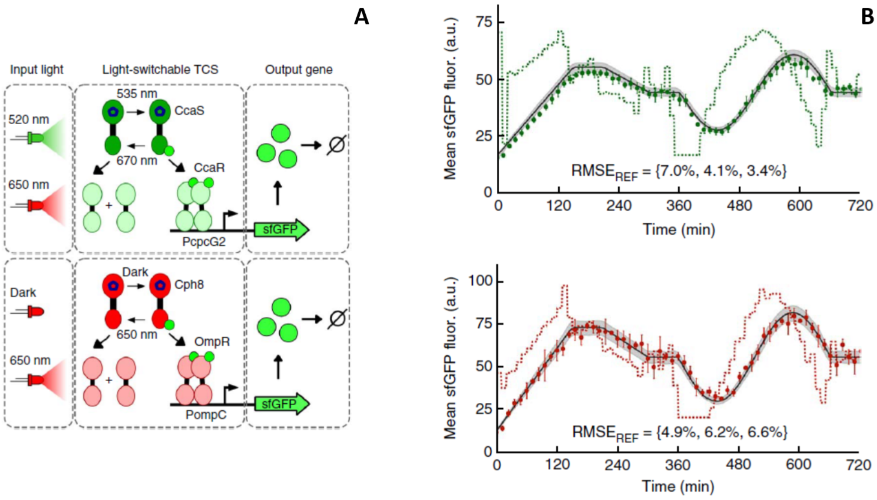


Figure 3.7: Characterisation and control of light-switchable two-component systems in bacteris*(A) Two-component systems analysed in (9): the first is green-activated and red-deactivated, its activity is regulated modulating the intensity of the green LED while the red is kept at its maximum intensity value; the second circuit is dark activated and red deactivated, thereby its activity is controlled by the intensity of the red LED. (B) Results of the open loop control experiments performed on the two-component systems (green activated, upper panel and dark activated lower panel): the black solid line is the reference signal, dashed line is the intensity of the LED used to elicit system dynamics, dots are the average of the fluorescence measured via flow cytometry. Adapted from (9)*

The dynamical models thus identified, were used to run an algorithm that iteratively simulated the response of the models optimising, at the same time, the input until the deviation from a desired time profile for the gene expression was considerably ”‘small”’. Olson and colleagues repeated this optimisation procedure over several different reference signals and applied the calculated input to living cells achieving the control task (Figure 3.7 B). Their control scheme, being an open loop regulation, relies totally on the accuracy of the dynamical models derived.

Melendez *et al.* have implemented a feedback control scheme to control the concentration of a fluorescent reporter by means of light stimuli in yeast cells (8). The novelty of their approach resides in the designed culturing apparatus that comprises a chemostat in which cells grow and replicate and, a microfluidic device in which yeasts, sampled from the chemostat at regular time intervals, are automatically loaded to be imaged for fluorescence quantification (Figure 3.8 A). The biological system controlled resembles the one analysed and regulated in (5): the Cry2 and Cib1 protein from *Arabidopsis Thaliana* are fused, respectively, to the *GAL4* binding and *GAL4* activation domains so that the transcription of all Gal4 dependent genes becomes blue-light inducible. To monitor the dynamics of this system they used a yellow fluorescent reporter driven by the *GAL1* promoter. To achieve set point regulation they used a simple ON-OFF control strategy implemented on a control board that automatically decides, on the basis of the control error, whether to turn ON or OFF the blue LED. The experiments carried out confirm the effectiveness of the control strategy in closed loop and, highlight that the devised technology can be exploited for further future analysis of biological networks where steady state reproducible growth conditions has to be maintained.

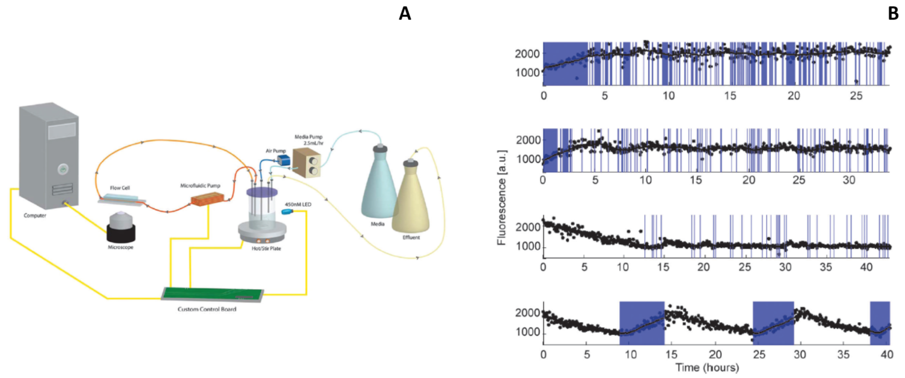


Figure 3.8: Characterisation and control of light-switchable two-component systems in bacteris*(A) At the core of the experimental platform designed by Melendez and colleagues there is a custom chemostat in which cells grow and replicate; by the mean of a microfluidic pump at regular time intervals cells from the culturing device are sampled and loaded in a microfluidic device to be imaged for fluorescence quantification. On the basis of the fluorescence intensity and of the control reference, an ON-OFF control strategy decides whether to turn on or off the blue LED to reach and maintain the set point. (B) Feedback set-point control results and (lower panel) oscillations induction in fluorescent protein concentration. Adapted from (8)*

4

A platform for controlling yeast cells

In this Chapter I will present the experimental platform designed for the external control of a population of yeast cells. The design was inspired by the requirements that the entire setup had to fulfil: *a)* to guarantee all the physiological conditions for the cells, *b)* to administer external input to properly elicit cellular dynamics and *c)* to monitor in real-time the desired output.

To fit all these needs I contributed to the design of an experimental platform based on a microfluidic device, a time-lapse microscopy apparatus and, a set of automated syringes all controlled by a computer (Figure 4.1). Microfluidics allows to grow cells and to precisely change their environmental conditions in real-time; moreover the cells in the device can be imaged with the microscope at high sampling rate, in order to evaluate the effects of the input provided to the system. This is achieved by measuring over time the fluorescence of a reporter used to track the output of the phenomenon of interest. The number of measured outputs relies on the number of different colours that can be tracked at the same time by fluorescent microscopy (up to 4 can be easily quantified).

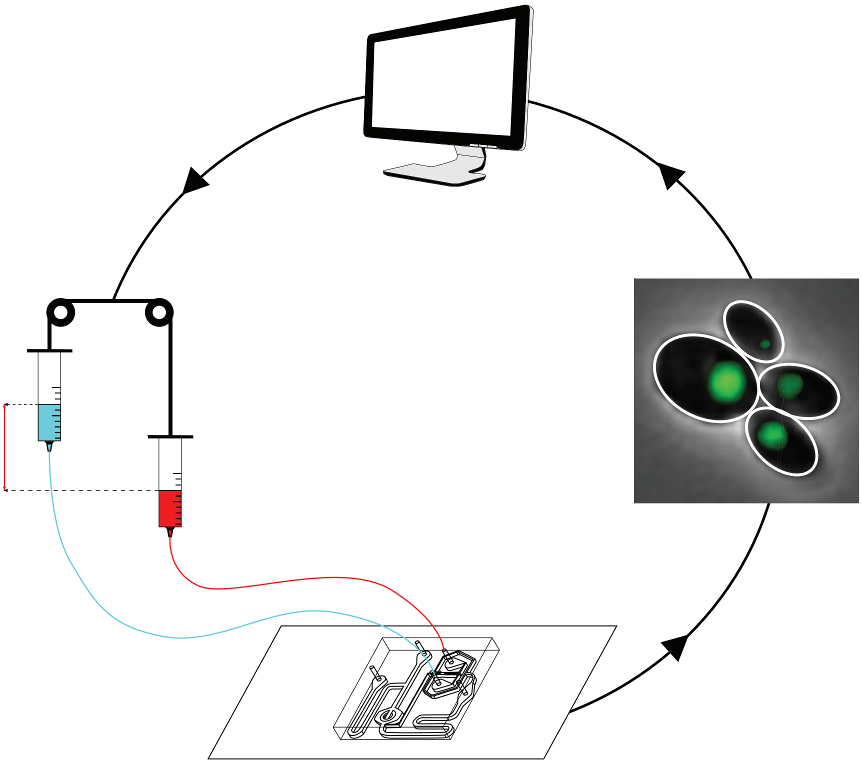


Figure 4.1: Technological platform. *A system of automated syringes, controlled by a computer, is used to administer different inputs to cells that are loaded in a microfluidic device ensuring their survival for thousands of minutes. The number of measured outputs relies on the number of different colours that can be tracked at the same time by fluorescent microscopy.*

4.1 Design and implementation

4.1.1 Microfluidics

Microfluidic devices allow to isolate cells from external disturbances as well as to continuously refresh or change their growing medium in order to avoid the depletion of nutrients due to cell consumption; moreover in order to obtain more productive experiments (optimization of reagents amount, consumables and time), and especially to capture single cell behaviour using high resolution microscopy the use of microfluidics device becomes necessary. These devices are, essentially, chips where fluid dynamics at the microliter scale are exploited. Given the channel dimensions and the fluid properties, it is possible to determine the flow regime in the channels. In microchemostat devices, considering the Reynolds number, for an aqueous fluid the flow is exclusively laminar; these flows contain highly predictable, parallel flow streams resulting in fairly easy to model profiles.

The principle is to have an area where the cells are forced to be in (cell trap), together with a series of channel used to provide multiple compounds to the trap in order to regulate cells environment or, to collect cells and fluids wastes.

The device I chose for this study is the MFD0005a device (Figure 4.2), designed by the lab of Jeff Hasty (UCSD) (44), housing a micro-chamber (height: $3.5\mu m$) which "traps" yeast cells, that can only grow in a mono-layer, thus allowing an easier automated image analysis. We produced replicas of the device designed by Ferry and colleagues (44) thanks to the master-mold they kindly provided us as a blueprint, following the protocol described in §8.

During an experiment, the fluids used as "inputs" enter from ports 1 and 2 arriving at the dial - a - wave (DAW) junction (Figure 4.2 B); this junction has two inlets and three outlets; the ratio of the inputs from port 1 and 2 leaving the junction to the cell chamber is determined by modulating the difference in hydrostatic pressures at the two inlets. Excess fluid is diverted through a shunt network to port 3, which is a waste port. Fluid leaving the central fork of the junction for the cell chamber travels through a long channel where it is mixed into a uniform concentration by staggered herringbone mixers (SHM) (Figure 4.2 C); these are designed to induce a corkscrew effect in the fluid stream and increase

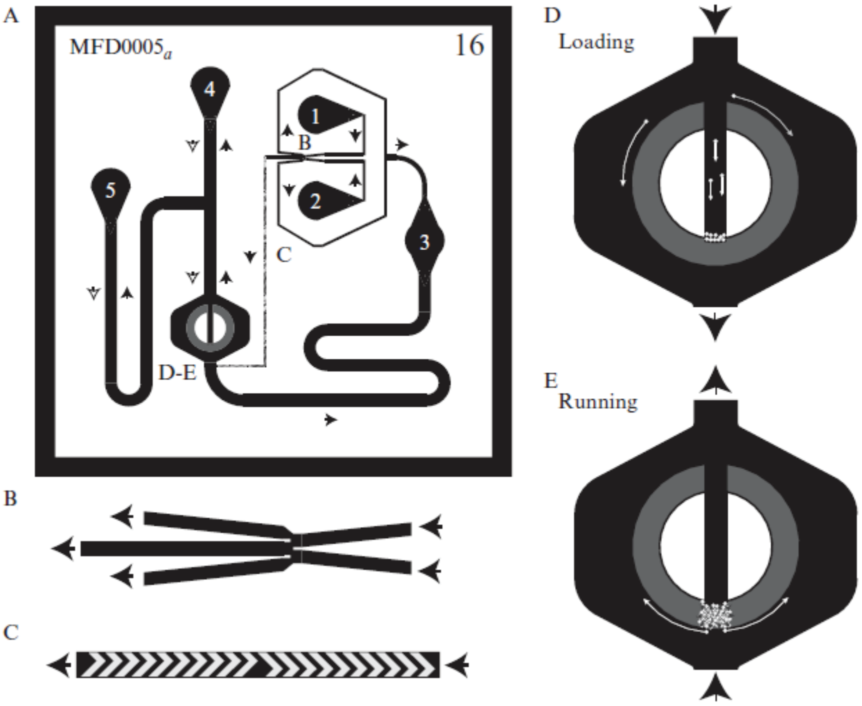


Figure 4.2: Microfluidic device MFD0005a. *Overview of the MFD005a device architecture. (B) DAW junction. (C) Staggered herringbone mixers. (D) Cells trap, loading. (E) Cells trap, running an experiment. Adapted from (44)*

the surface area available for mixing (45). After mixing, fluid from port 1 and 2 enters the cell chamber and proceeds to the outlet ports 4 and 5. Fluid also enters a diversion channel and exits at port 3. By controlling the height of port 3 relative to ports 4 and 5, it is possible to set the ratio of fluid passing through the chamber versus exiting through the diversion channel. The modulation of this ratio allows to control flow velocity across cell chamber. For further details see §8.1 and refer to (44).

4.2 Actuation System

As mentioned above DAW junction (Figure 4.2 B) works by changing the relative pressures at DAW ports, while keeping the total pressure the same; thus the input ratio to cells is the result of the pressure difference at ports 1 and 2 (Figure 4.2 A). The actuation aim is to establish this difference in order to appropriately modulate, according to the control, the inputs concentration in the fluid reaching the cell trap. Physically this can be achieved by changing the hydrostatic pressure of the syringes linked to the two inlet ports. To accomplish this task, I designed and built two vertically mounted linear actuators; using this system it is possible to change the height of liquid-filled syringes that feed into the DAW junction. The actuation system comprises two linear guides; every linear actuator is designed to move independently from the other; the motion is realised through a stepper motor, while the transmission by using a timing belt and two pulleys. The transmission gear adopted is an ideal solution, with no need of high torques, to guarantee good performance in terms of actuation speed (46). Moreover the use of a stepper motor is the simplest and cheapest way to achieve this result; this motor is controlled through an appropriate excitation sequence of the stator circuit (using specific electronic drivers), each rotor position corresponds to a step (particular excitation status of stator circuit), hence the name stepper motor (46) Thereby given the steps number, it is possible to know rotor angular position and the number of complete revolutions made by the motor, thus there is no need of an angular position measure because it is intrinsically known. Further details regarding the sizing and the specifications of the actuation system are reported in §8.2.

4.3 Microscopy and Image Analysis

To monitor cellular processes dynamics, as well as, to check for the right administration of external inputs to trapped cells, I have taken advantage of an inverted fluorescence microscope (Nikon Eclipse Ti) equipped with an automated and programmable stage, an incubator to guarantee fixed temperature and gasses to cell environment and a high sensitivity Electron Multiplying CCD (EMCCD) Camera (Andor iXON Ultra897). The microscope and the camera can be programmed to acquire, at regular time intervals, images from a fixed area of the cell trap as well as from different points of the microfluidic device. Moreover at each time point the camera takes two types of images: *(a)* a bright field image (phase

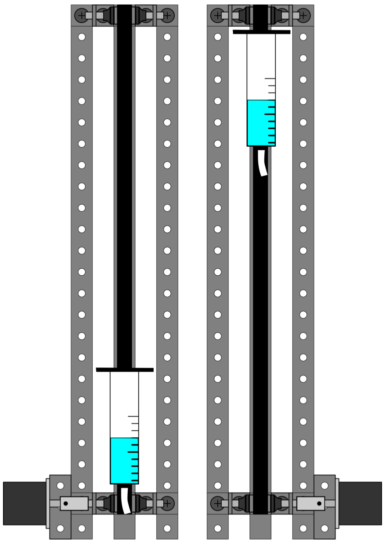


Figure 4.3: Linear actuators designed. *Linear actuators, representative picture.*

contrast) and *(b)* fluorescence images (with the appropriate filters). Additional informations on the microscope setup for real time image acquisition are available in §8.3.

Once cells have been imaged, image analysis methods can be applied to estimate their fluorescence; to this end, I adapted a custom image processing algorithm previously developed in the Laboratory where this study was completed (47). The algorithm devised is able to locate cells within each Phase Contrast image thus identifying all the pixels belonging to cells. This information is used to calculate the fluorescence expressed, for each fluorescent reporter, by the entire population as well as each single cell. An extended explanation of the methods employed to carry out this analysis are reported in §8.4.

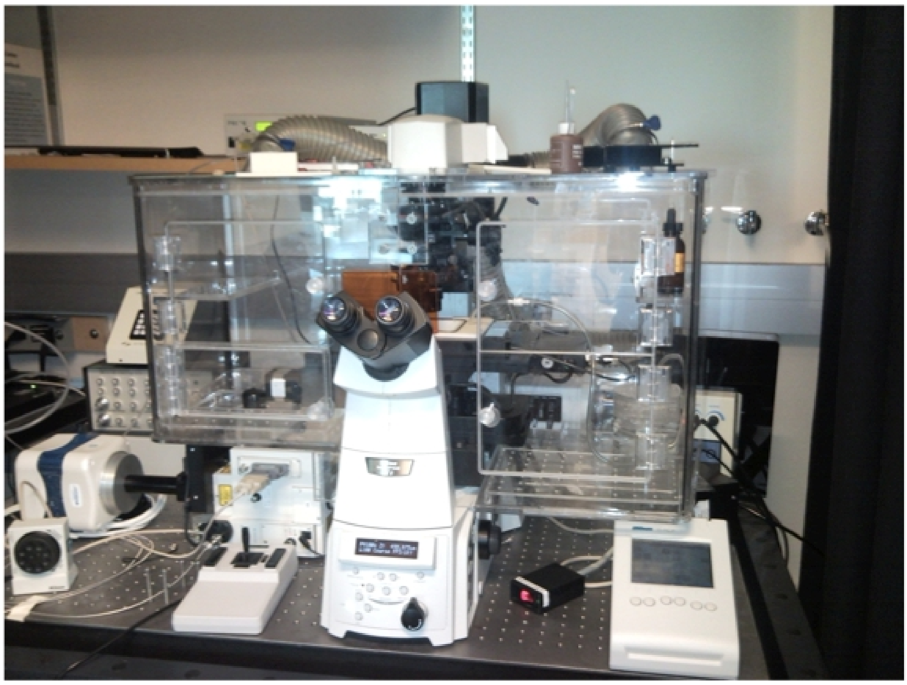


Figure 4.4: Nikon Eclipse TI fluorescence microscope. *The inverted fluorescence microscope (Nikon Eclipse TI) used in this study, equipped with a EMCCD high sensitivity camera and an incubator to control gasses and temperature of cell environment.*

5

Identification and modelling of transcriptional processes in living cells

In this Chapter I discuss the use of the experimental platform presented in Chapter 4, to identify and compare different linear and non-linear models for the *GAL1* promoter in yeast cells driving expression of a green fluorescent protein (Gfp) fused to the *GAL1* gene (48) (see §5.1 §5.2). I show that the experimental set up I have implemented allows to infer quantitative dynamical models of transcriptional processes from measured input and output data, with or without any *a priori* knowledge of the underlying characteristics of the system modelled. The possibility of acquiring dense time series data from biological processes, is instrumental to derive mathematical description not only to predict the behaviours of the system in response to various stimuli but, even to design feedback control strategies meant to steer the same process to a desired trend over the time. Part of this work has been published in (49)

Moreover here I provide details (§5.3 and §5.4) regarding the topology and the mathematical model of two synthetic gene regulatory networks stably integrated in the host cell genome, one called *IRMA* and embedded in yeast cells (11) and, the other, an inducible positive feedback loop, integrated in Chinese Hamster Ovary (CHO) cell line (25).

5.1 Modelling *GAL1* promoter dynamics in yeast S. Cerevisiae, black box identification

The identification of the parameters of mathematical models able to capture the dynamics of the *GAL1* promoter has been carried out without assuming any *a priori* knowledge of the underlying chemical and physical processes occurring inside the cells. Input and output data have been used to fit different model structures thus considering the system as a ”black box” (1).

5.1.1 Biological system

The biological system that I used to test and evaluate the potential of the experimental set-up developed is a strain of yeast cells (yGIL337, Gal1-GFP::KanMX,GAL10-mCherry::NatMX) constructed by Lang et al(48). In these cells the Green Fluorescent Protein (Gfp) is fused to the Gal1 protein and expressed from the *GAL1* promoter and a variant of the red fluorescent protein (mCherry) is fused to the Gal10 protein and expressed from the *GAL10* promoter (48) (Figure 5.1 Panel A).

The Gal1 protein is one of the enzymes needed by yeast for Galactose utilisation, thus the the activity of the *GAL1* promoter is related to the presence in the cells' environment of a sugar, Galactose, which is sensed by the cells as a 'switch on' signal for the expression of the *GAL1* gene. On the contrary, the presence of another sugar, Glucose, represses the production of Gal1 protein, because Glucose is the preferred carbon source requiring much less energy to be metabolised (50). Thus, cells will first consume all the available Glucose and then switch to utilise Galactose, if any is available in the medium. Because of this, the input provided to the yeast cells can either be Glucose (which switched off Gfp production) or Galactose (which switched on Gfp production), but not a combination of the two sugars because yeast cells will not respond to Galactose if Glucose is present.

Moreover, a recent dynamical model of the Galactose regulation system in yeast has revealed that the Galactose system works as a low-pass filter, thus dampening the effect of switches between Galactose and Glucose (50). Hence the frequency of the input signal has to be low enough for the system to respond.

The strategy I have followed was to dynamically modulate the presence of two sugars in the medium in which the cells are grown, as input to the system, and to follow the dynamics of the *GAL1* promoter in response to such an input by measuring Gfp fluorescence, which is considered as the output of the system (Figure 5.1 Panel B).

5.1.2 Experimental results

Yeasts have been imaged for up to 16 hours. During this interval Galactose and Glucose were alternatively provided to the yeast chamber. The frequency was chosen according to the previous literature (50), specifically cells were provided with a Galactose enriched medium for 180 min and then with Glucose for the following 180 min and so on until the end of the experiment, as depicted in Figure5.2, Lower Panel. The concentration of Galactose was tracked with a red fluorescent dye (Sulforhodamine B), so that it was possible to obtain a time profile of the actual input provided to the cells by measuring the fluorescence of the medium in the red spectrum (Figure 5.2, Lower Panel.). The average Gfp fluorescence of the cells' population, in the green spectrum, has been instead taken as the system output (Figure 5.2, Upper Panel).

The representative dataset shown in Figure 5.2 is consistent with the expected behaviour of the promoter under investigation. The presence of Galactose induces the expression of Gfp from the *GAL1* promoter, by activating the Gal4 transcription factor, whereas Glucose represses the activity of the *GAL1* promoter and hence of the Gfp production. Therefore, the Gfp fluorescence level is expected to increase or to stay at a high steady state value when the cells are fed with Galactose. It should instead decrease, or stay at a low steady state, when Glucose is provided. At the beginning of the experiment, cells exhibit a high fluorescence level, since cells were taken from a Galactose overnight culture. Gfp level was expected to remain almost constant in the first 180 minutes of the experiment (Figure 5.2). However, as can be seen in Figure 5.2, its value was observed to slightly decrease, probably because of the stress during the loading into the microfluidic device. In the rest of the experiment, the activity of the *GAL1* promoter in response to the other two pulses of Galactose is consistent

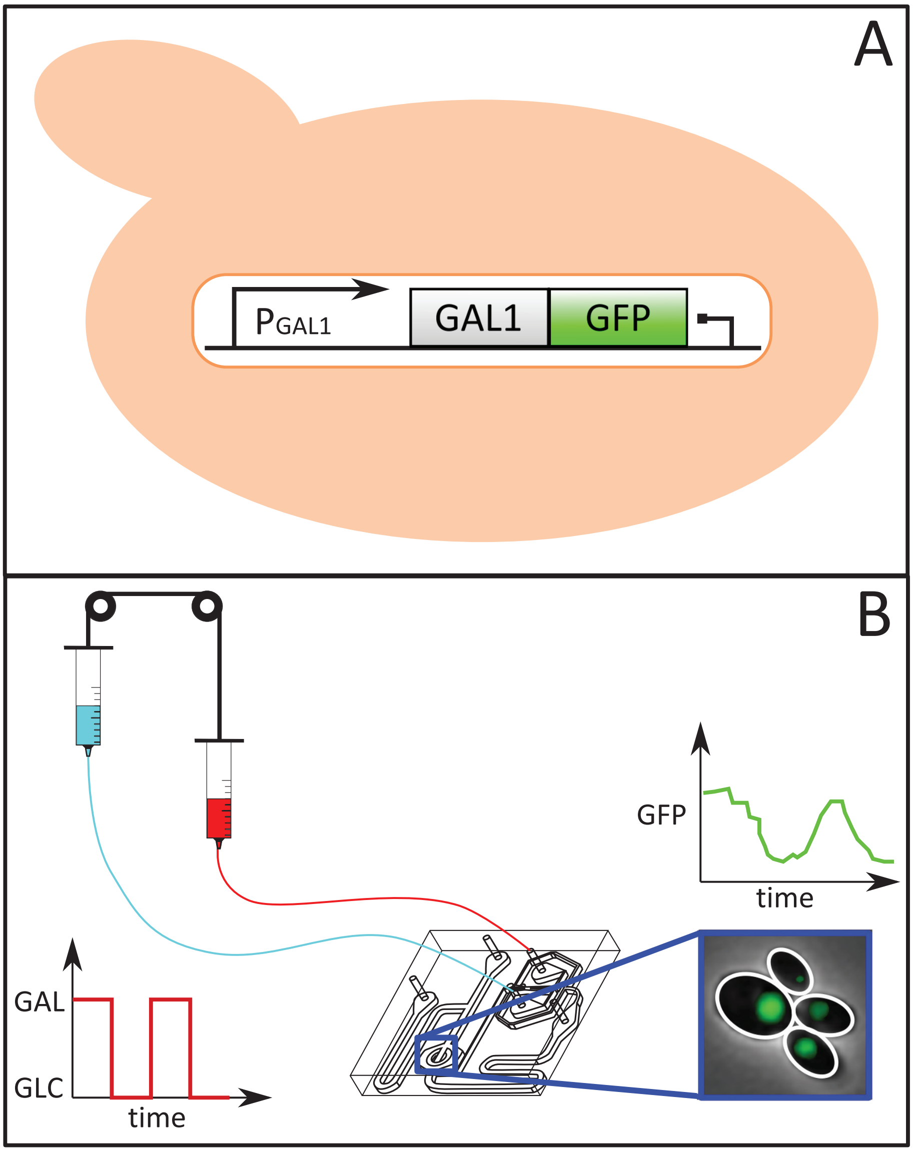


Figure 5.1: Biological system and technological platform. Panel A: *the system under investigation is a strain of yeast cells in which the Green Fluorescent Protein (Gfp) is fused to Gal1 and expressed from the GAL1 endogenous promoter* . Panel B: *A system of automated syringes, controlled by a computer, is used to administer two different sugars to the cells; yeasts are loaded in a chamber of a microfluidic device ensuring their survival for thousands of minutes. Cells fluorescence is quantified to follow the dynamics of the GAL1 promoter in response to sugar stimuli.*

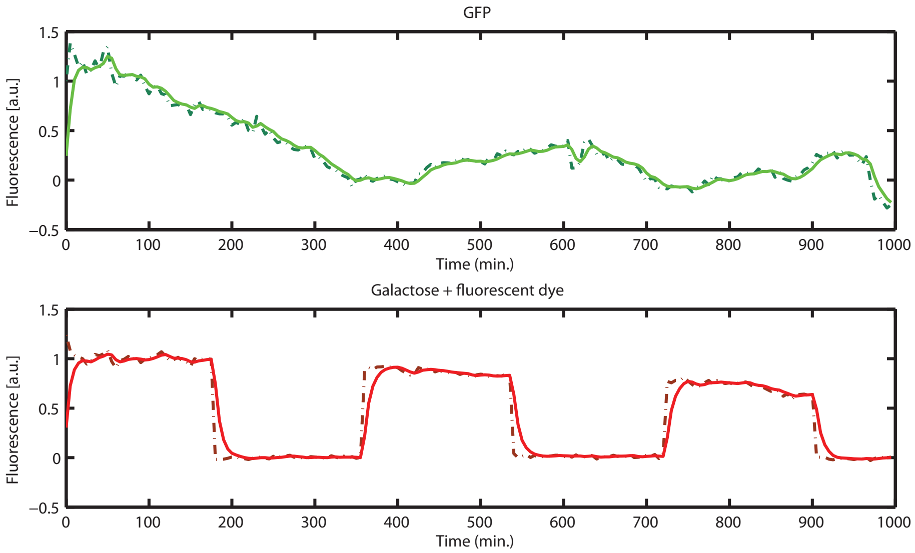


Figure 5.2: Experimental data. Upper Panel: *Cell population average fluorescence (dashed line) measured during the experiment, the same signal but filtered (solid line) with a low pass filter in order to reduce the noise of the measured data; this signal is considered as the output of the system.*. Lower Panel: *Fluorescence of the dye added to the Galactose (dashed line) measured during the experiment; a high level corresponds to Galactose, whereas a low level to Glucose. This signal has been filtered (solid line) with the same low pass filter used for the output, this time profile is considered as the input of the system*

with the expected behaviour, as revealed by the Gfp fluorescence level, (Figure 5.2).

5.1.3 Candidate models

I have used the experimental platform to test and compare the following common identification methods (a description of the metrics used to assess the effectiveness of the identification is given in §5.1.4):

1. **ARX models**

   Auto Regressive exogenous models (ARX) (1) with a delay from input to output of the form:

$$y(t)+a_1y(t-1)+...+a_{n_a}y(t-n_a) = b_1u(t-1-n_k)+...+b_1u(t-n_b-n_k) \quad (5.1)$$

the output $y(t)$ represents the measured fluorescence level at the current time t and it is assumed to be proportional to the sum of its n_a past values, a sort of memory of the system, and to the sum of n_b past values of the input (i.e. Galactose/Glucose). In addition the input is assumed not to act instantaneously on the output but after a delay of $n_k + 1$ samples.

The estimation of the coefficients $a_1, a_2, ..., a_{n_a}$ and $b_1, b_2, ..., b_{n_b}$ was carried out via the *Prediction Error Minimisation* (PEM) criterion (1). The model structure, namely the vector $[n_a, n_b, n_k]$, was chosen to minimise the *Akaike's Information Criterion* (AIC) value as described in §5.1.4 and (1).

2. **First order transfer function with delays**
First-order transfer function with a time delay of the form:

$$G(s) = \frac{K_p}{1 + sT_p} e^{-T_d s}. \quad (5.2)$$

The transfer function is just a different mathematical representation, using the Laplacian operator, of the first-order linear ordinary differential equation (ODE) reported below:

$$\dot{y}(t) = K_p u(t - T_d) - \frac{1}{T_p} y(t) \quad (5.3)$$

This equation represents the rate of change of the Gfp fluorescence level ($\dot{y}(t)$) as a function of a production term proportional to the input ($K_p u(t-T_d)$, i.e. Galactose/Glucose)), which is assumed to act on the output only after a delay equal to T_d. A linear degradation term for the reporter protein ($\frac{1}{T_p} y(t)$) is also present.

The parameters I estimated, via the Prediction Error Minimisation (PEM) criterion (1), are the transfer function gain K_p, the time constant T_p and the delay T_d.

3. **State space models**

   Higher-order (linear time-invariant) state space model, which assumes that more than one differential equation is needed to correctly model the promoter dynamics. These extra equations can either represent physical quantities (i.e. mRNA, protein) or abstract quantities useful to model the system.

   The generic state-space model of interest can be written as:

   $$\begin{aligned} \dot{x} &= Ax + Bu, \quad x(0) = x_0 \\ y &= Cx \end{aligned} \tag{5.4}$$

   where x is the state vector of dimension n, A is a n-by-n matrix, B is a column vector of dimension n, C is a row vector of dimension n and x_0 is the vector of initial conditions.

   For this class of models, two different algorithms were considered, prediction error minimisation (PEM applied to state space models) and N4SID (1). The system order n, the coefficients of the matrices A, B and C, and the vector of initial conditions x_0 were estimated from the experimental data.

4. **Nonlinear model**

   I have considered, further to the previous black box models, the use of a nonlinear model of the form:

   $$\dot{y}(t) = \alpha + v\frac{u(t-\tau)^H}{K^H + u(t-\tau)^H} - Dy(t) \tag{5.5}$$

   this nonlinear differential equation models the rate of change of of the Gfp fluorescence level ($\dot{y}(t)$ in response to Galactose stimulation (the $u(t-\tau)$, delayed of τ minutes) as a non linear function of a production term (the Hill function (51)) and of a linear degradation term ($Dy(t)$). In the Hill function, K represents the value of u needed to achieve half of the maximal production rate v, and the H exponent, *Hill coefficient*, governs the steepness of this function (the higher the value, the more similar the function is to a step function). The α parameter represents the promoter "leakiness", i.e. the production rate in the absence of Galactose (i.e. u=0).

I have added an explicit delay τ to the input, as in the case of the ARX model and the Transfer Function, since from the experimental results (see Figure 5.2) the response of the system, namely changes in Gfp values, in response to switches between Galactose and Glucose and vice-versa, appears to be delayed.

I applied the Simulated Annealing algorithm (52) to estimate the parameters α,v,H,K,D,τ and the initial condition (IC) for the state variable x starting from an initial guess of all of them. The estimation has been carried out by minimising the following objective function:

$$\mathrm{J} = \frac{\sqrt{\sum_{i=1}^{N}\left(\widehat{y_i} - y_i\right)^2}}{\sqrt{\sum_{i=1}^{N}\left(y_i - \overline{y}\right)^2}} \tag{5.6}$$

where for the i-th datapoint, $\widehat{y_i}$ is the model output in response to the measured input that leads to the real system output y_i and, $\overline{y}$ is the average of y.

It is worth pointing out that a systematic experimental comparison of the identification approaches described above when applied to *in vivo* biological systems has not been carried out in the existing literature.

5.1.4 Metrics and Validation

To assess the performance of each identification scenario and to carry out a comparison among different methods, I have used the following metrics to evaluate their predictive ability.

1. **Akaike's Final Prediction Error and Information Criterion**
   Akaike's Final Prediction Error (FPE) or Aikake's Information Criterion (AIC) (1) can be used to evaluate the quality of a given model by testing how it captures the system response to a known input signal. The metrics can be computed as follows:

$$\mathrm{FPE} := \left(\frac{1+\frac{m}{N}}{1-\frac{m}{N}}\right)\frac{1}{N}\sum_{i=1}^{N}\epsilon^{2}\left(i,\theta_{N}\right), \tag{5.7}$$

and

$$\mathrm{AIC} := log\left(\frac{1}{N}\sum_{i=1}^{N}\epsilon^{2}\left(i,\theta_{N}\right)\right)+\frac{m}{N}, \tag{5.8}$$

where θ_N is the vector of estimated parameters, m is the number of estimated parameters, N is the dimension of the estimation dataset and $\epsilon\left(i,\theta_N\right)$ are the prediction errors. Both FPE and AIC take their smallest values when the model is the most accurate.

2. **Fitting percentage**
This index provides a measure of the percentage of the output variation that is reproduced by the model and is given by the following formula (1):

$$\mathrm{FIT} := 100\left(1-\frac{\sqrt{\sum_{i=1}^{N}\left(\widehat{y}_i-y_i\right)^2}}{\sqrt{\sum_{i=1}^{N}\left(y_i-\overline{y}\right)^2}}\right) \tag{5.9}$$

where for the i-th datapoint, $\widehat{y}_i$ is the model output in response to the measured input that leads to the real system output y_i and, $\overline{y}$ is the average of y. This index can effectively be used also for cross-validation purposes by evaluating the model ability to capture data that are different from those used for the identification of its parameters.

5.1.5 Identification strategies

I have tested the suitability of models identified, with each of the strategies here considered, in two sets of different scenarios. In the first set, both identification and validation of the models were performed on the same dataset (scenarios I and II). In the second set, the predictive ability of each of the models is evaluated by using a dataset different from the one used for identification (scenarios III and IV).

- **Scenario I**
  The parameters of each class of models were estimated on the dataset depicted in figure 5.2 that consist of experimental measurements for both the input (red dye fluorescence) and output (Gfp fluorescence) filtered via a low pass filter to reduce measurement noise. The same dataset was used for validation.

- **Scenario II**
  The data used to identify the mathematical models are shown in figure 5.3; differently from Scenario I, here the input is the ideal concentration of Galactose (i.e. a square waveform) and not the estimated concentration as measured by the red dye fluorescence; the output is the same as in Scenario I, i.e. the filtered measured green fluorescence of the cells. The models obtained were then validated on the same dataset used for identification.

- **Scenario III**
  In this scenario, I have used only the first 700 minutes of the data in Scenario I (figure 5.2) to estimate the parameters of the mathematical models described in §5.1.3, validation was performed on the entire dataset.

- **Scenario IV**
  In this scenario, I have used only the first 700 minutes of the data in Scenario II (figure 5.3) to estimate the parameters of the mathematical models. As in the case of scenario III, validation was performed on the entire dataset.

5.1.6 Results

The results of the identification for the different model structures across the four scenarios are described below. The value of the indices given in §5.1.4 are summarised in Table 5.1.

The order of the ARX models used to capture the system behaviour over scenarios I-IV obtained by the System Identification algorithm are the following:

- Scenario I: $[n_a, n_b, n_k] = [5, 5, 10]$ [Figure 5.4(A)]

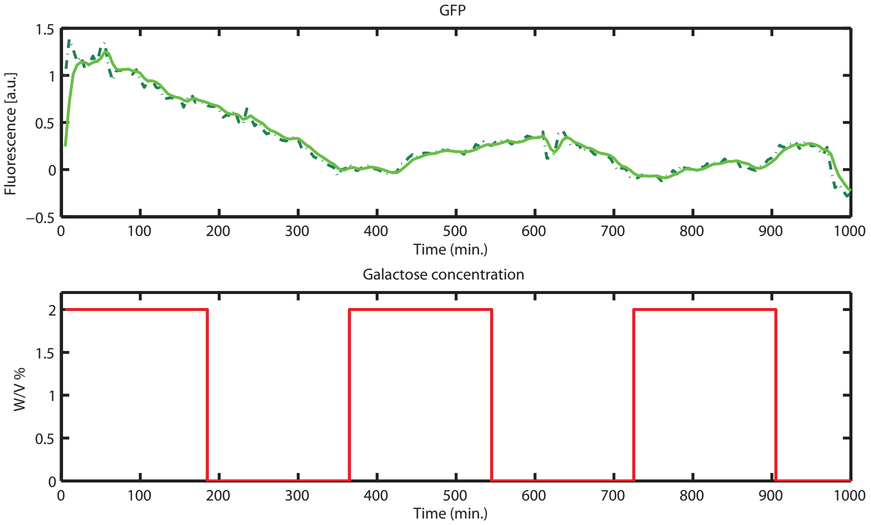


Figure 5.3: Experimental data and ideal input. Panel A: *Cell population average fluorescence (dashed line) measured during the experiment, the same signal filtered (solid line) with a low pass filter in order to reduce the noise of the measured data; this signal is considered as the output of the system.* Panel B: *The concentration of Galactose in the medium (solid line) provided to the system. A square wave of Galactose is administrated to cells to stimulate the activity of the GAL1 promoter*

- Scenario II: $[n_a, n_b, n_k] = [5, 5, 12]$, [Figure 5.5(A)]
- Scenario III: $[n_a, n_b, n_k] = [5, 3, 14]$, [Figure 5.6(A)]
- Scenario IV:$[n_a, n_b, n_k] = [4, 5, 12]$, [Figure 5.7(A)]

For the delayed transfer function, we obtained the following parameters:

- Scenario I: $[K_p, T_p, T_d] = [0.36, 94.47, 57.82]$ [Figure5.4 (B)]
- Scenario II: $[K_p, T_p, T_d] = [0.13, 106.28, 61.10]$ [Figure5.5 (B)]
- Scenario III: $[K_p, T_p, T_d] = [0.43, 77.10, 68.07]$ [Figure 5.6(B)]
- Scenario IV:$[K_p, T_p, T_d] = [0.19, 87.10, 66.20]$ [Figure 5.7(B)]

When state space models are used, we obtained the following order (the full matrices were n:

- Scenarios I & II: $n = 4$ with both N4SID and PEM [Figures 5.4(C) and 5.5(C)]
- Scenarios III & IV: $n = 5$ with both N4SID and PEM [Figures 5.6(C) and 5.7(C)]

Finally for the non linear model we found the following values for the parameters of equation (5.5) for each scenario:

- Scenario I: $[\alpha, v, H, K, D, IC, \tau] =$
  $[0.0000, 0.0055, 2.3678, 1.8792, 0.0048, 1.2719, 54.2024]$ [Figure5.4 (D)]
- Scenario II: $[\alpha, v, H, K, D, IC, \tau] =$
  $[0.0000, 0.0138, 1.7930, 2.5890, 0.0075, 1.4449, 66.4625]$ [Figure5.5 (D)]
- Scenario III: $[\alpha, v, H, K, D, IC, \tau] =$
  $[0.0014, 1.2429, 3.4970, 3.8305, 0.0126, 1.2381, 70.9909]$ [Figure 5.6(D)]
- Scenario IV:$[\alpha, v, H, K, D, IC, \tau] =$
  $[0.0018, 0.0801, 2.2255, 3.3595, 0.0101, 1.9086, 72.1324]$ [Figure 5.7(D)]

	Scenario I	**Scenario II**
FPE	$arx = 2.06 \cdot 10^{-4}$ $tf = 0.0086$ $ss - n4sid = 0.0084$ $ss - pem = 0.0072$ $nl = 0.0223$	$arx = 1.94 \cdot 10^{-4}$ $tf = 0.0097$ $ss - n4sid = 0.0107$ $ss - pem = 0.0100$ $nl = 0.0184$
AIC	$arx = -8.4810$ $tf = -4.7606$ $ss - n4sid = -4.7767$ $ss - pem = -4.9300$ $nl = -3.8392$	$arx = -8.5428$ $tf = -4.6394$ $ss - n4sid = -4.5340$ $ss - pem = -4.5963$ $nl = -4.0300$
FIT (%)	$arx = 73.78$ $tf = 74.88$ $ss - n4sid = 76.11$ $ss - pem = 77.68$ $nl = 58.90$	$arx = 70.71$ $tf = 73.39$ $ss - n4sid = 73.02$ $ss - pem = 73.03$ $nl = 62.64$
	Scenario III	**Scenario IV**
FPE	$arx = 1.67 \cdot 10^{-4}$ $tf = 0.0036$ $ss - n4sid = 0.0056$ $ss - pem = 0.0049$ $nl = 0.0381$	$arx = 1.71 \cdot 10^{-4}$ $tf = 0.0038$ $ss - n4sid = 0.0140$ $ss - pem = 0.0116$ $nl = 0.0259$
AIC	$arx = -8.6894$ $tf = -5.6185$ $ss - n4sid = -5.1581$ $ss - pem = -5.2934$ $nl = -3.3182$	$arx = -8.6681$ $tf = -5.5659$ $ss - n4sid = -4.2470$ $ss - pem = -4.4376$ $nl = -3.7044$
FIT (%)	$arx = 71.71$ $tf = 73.59$ $ss - n4sid = 75.12$ $ss - pem = 74.55$ $nl = 55.72$	$arx = 66.48$ $tf = 69.92$ $ss - n4sid = 68.16$ $ss - pem = 69.96$ $nl = 63.50$

Table 5.1: Values of the indices defined in §5.1.4 across different scenarios. *arx*: ARX model, *tf*: delayed transfer function, *ss*: state space models, *nl*: nonlinear model

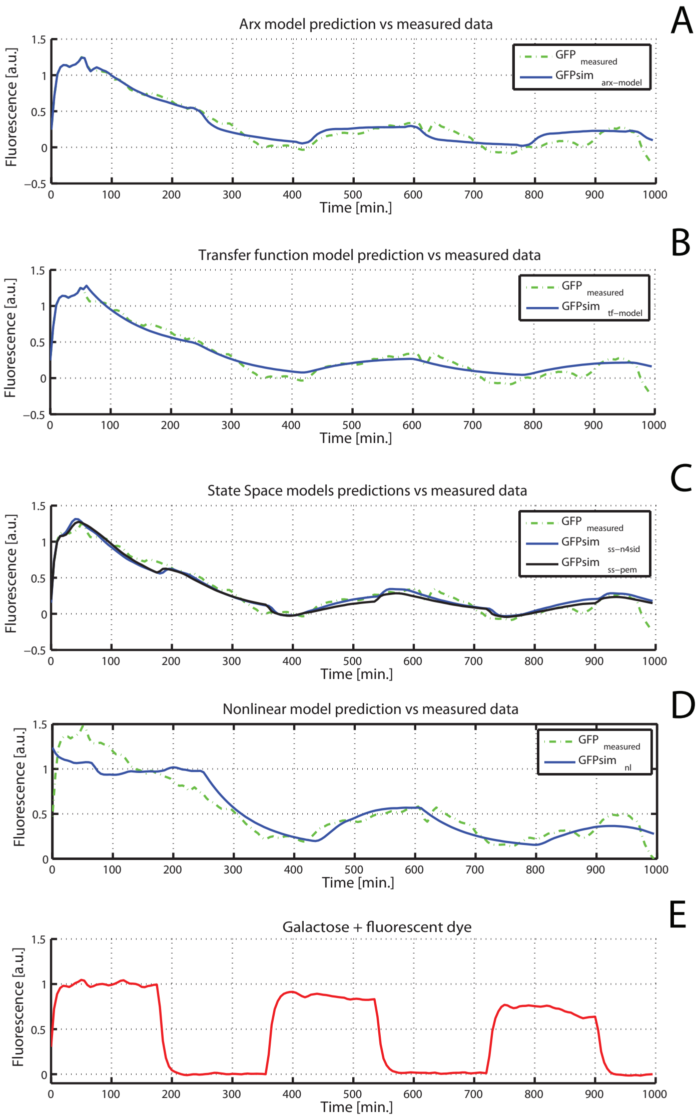


Figure 5.4: Scenario I - Fitting. Panel A: *The solid line represents the output of the Arx model in response to the input used for the identification, the dashed line is the filtered average cell fluorescence.* Panel B: *The solid line represents the output of the transfer function model in response to the input used for the identification, the dashed line is the the filtered average cell fluorescence.* Panel C: *Gray and black lines are respectively the outputs of the state space models identified with N4SID and PEM algorithms, the dashed line is the the filtered average cell fluorescence.* Panel D: *The solid line represents the output of the nonlinear model in response to the input used for the identification whereas, the dashed line is the the filtered average cell fluorescence.* Panel E: *The filtered fluorescence of the dye is used both to identify and to validate the models obtained.*

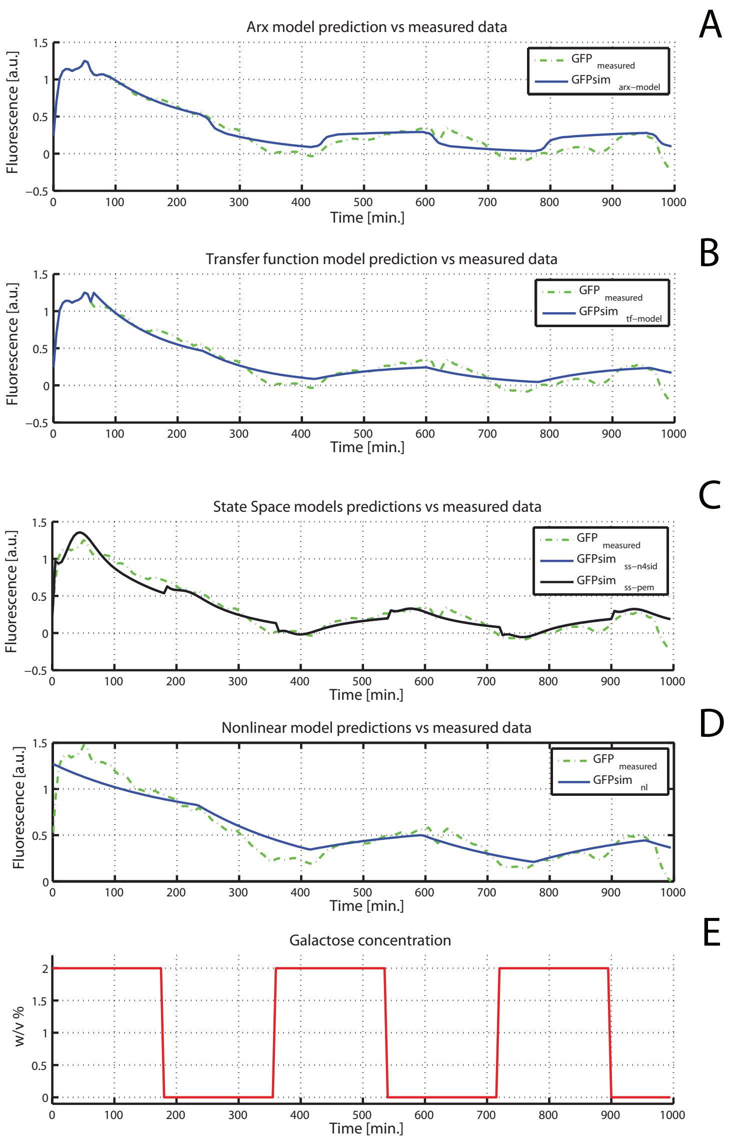


Figure 5.5: Scenario II - Fitting. Panel A: *The solid line represents the output of the Arx model in response to the input used for the identification, the dashed line is the filtered average cell fluorescence.* Panel B: *The solid line represents the output of the transfer function model in response to the input used for the identification, the dashed line is the the filtered average cell fluorescence.* Panel C: *Gray and black lines are respectively the outputs of the state space models identified with N4SID and PEM algorithms, the dashed line is the the filtered average cell fluorescence.* Panel D: *The solid line represents the output of the nonlinear model in response to the input used for the identification whereas, the dashed line is the the filtered average cell fluorescence.* Panel E: *Galactose concentration is used both to identify and to validate the models obtained.*

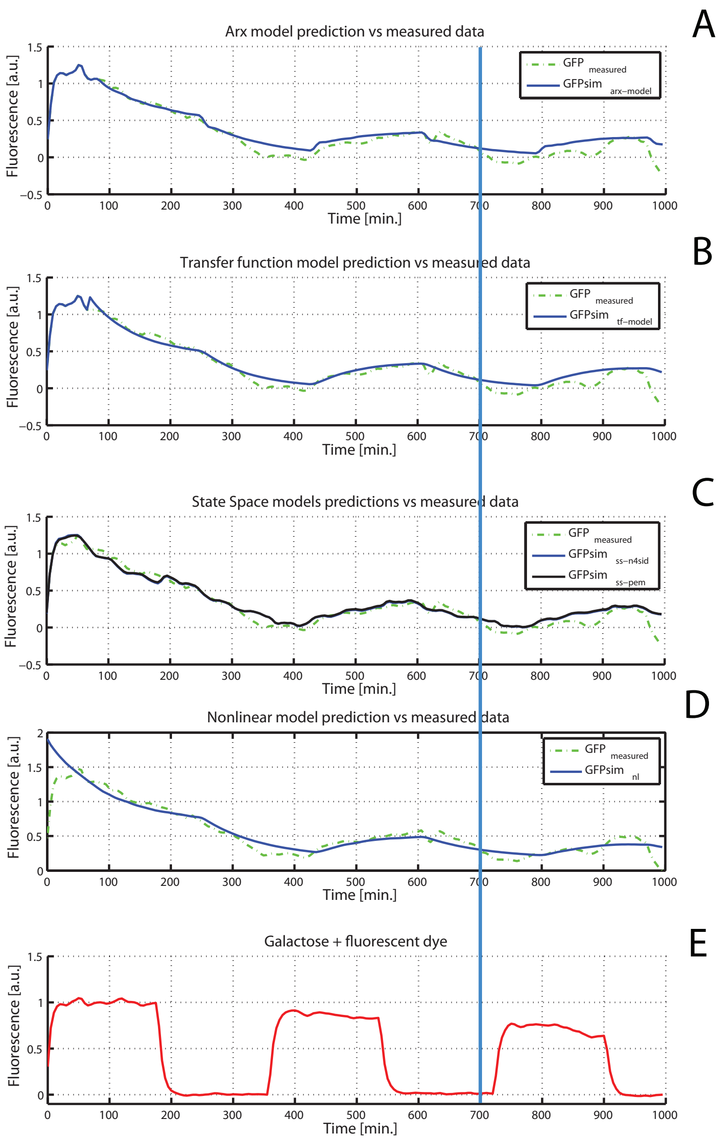


Figure 5.6: Scenario III - Fitting. Panel A: *The solid line represents the output of the Arx model in response to the input used for the identification, the dashed line is the filtered average cell fluorescence.* Panel B: *The solid line represents the output of the transfer function model in response to the input used for the identification, the dashed line is the the filtered average cell fluorescence.* Panel C: *Gray and black lines are respectively the outputs of the state space models identified with N4SID and PEM algorithms, the dashed line is the the filtered average cell fluorescence.* Panel D: *The solid line represents the output of the nonlinear model in response to the input used for the identification whereas, the dashed line is the the filtered average cell fluorescence.* Panel E: *The first 700 minutes of the filtered fluorescence of the dye were used to identify the parameters of the mathematical models; the validation was performed by using the entire signal.*

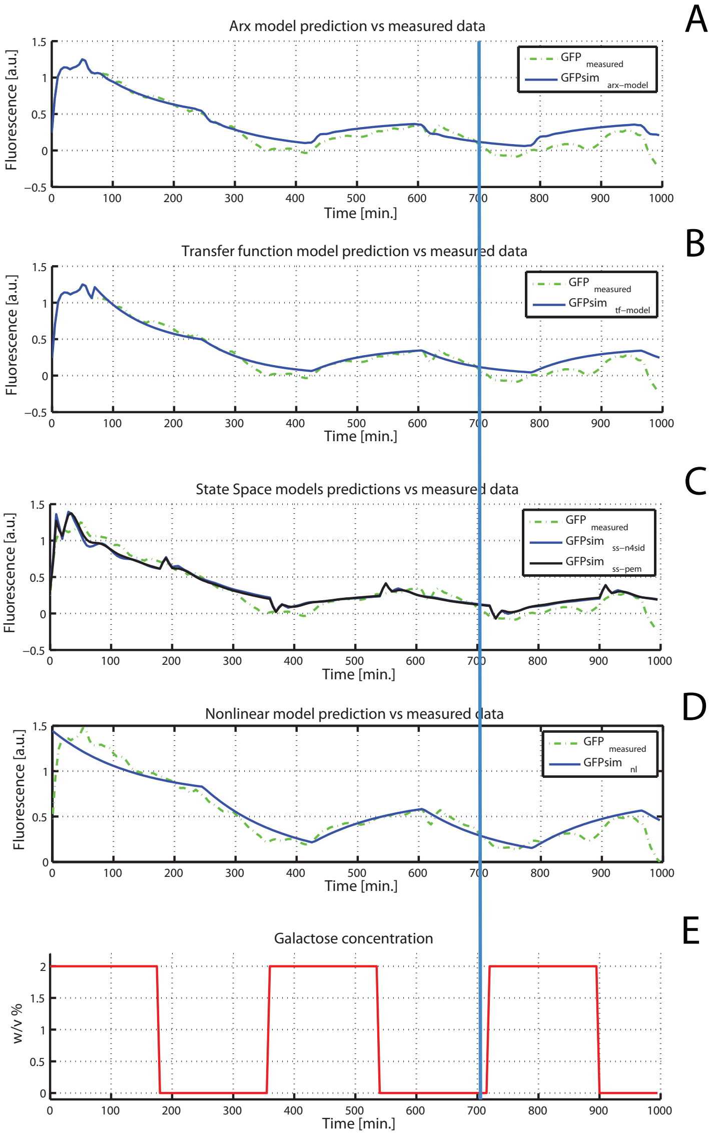


Figure 5.7: Scenario IV - Fitting. Panel A: *The solid line represents the output of the Arx model in response to the input used for the identification, the dashed line is the filtered average cell fluorescence.* Panel B: *The solid line represents the output of the transfer function model in response to the input used for the identification, the dashed line is the the filtered average cell fluorescence.* Panel C: *Gray and black lines are respectively the outputs of the state space models identified with N4SID and PEM algorithms, the dashed line is the the filtered average cell fluorescence.* Panel D: *The solid line represents the output of the nonlinear model in response to the input used for the identification whereas, the dashed line is the the filtered average cell fluorescence.* Panel E: *The first 700 minutes of the Galactose concentration were used to identify the parameters of the mathematical models;the validation was performed by using the entire signal.*

Both the ARX model and the Transfer Function explicitly include a parameter to account for a delayed response of the promoter to a change in Galactose concentration. In the ARX model, the delay is captured by the parameter n_k which was estimated, in the different scenarios, to range from 10 to 14 samples; since each sample is measured at 5 min time intervals, the delay can be estimated to be between 50 min and 70 min. Interestingly, the delay (T_d) estimated by the transfer function is in the same range, varying from 57.82 min to 68.07 min across the four scenarios. The state space model does not include an explicit time delay, although it is possible to add one, and hence it needs a relatively high number of states (4 or 5) to correctly capture the observed dynamics. For the nonlinear model, the estimated time delay ranges from 54.02 min to 72.13 min, in agreement with the linear models.

Biologically, the presence of the delay may be explained by several biological processes, such as the time needed to activate the Gal4 transcription factor following Galactose induction, or for the transcription initiation complex to form. However, since the *GAL1* promoter has been very well studied in the literature, it is known that both the Gal4 activation and initiation of transcription from the *GAL1* promoter are quite fast, in the order of minutes (53). On the contrary, the estimated delay from the system identification procedure is in the order of an hour, therefore the most likely explanation is the time required by the fluorescence reporter protein (Gfp) to fold and mature, as well as, its high half-life known to be in the order of hours (54).

These observations raise an obvious but important consideration: the reporter protein half-life, i.e. the protein stability, must be commensurate to the dynamics of the promoter to be modelled. Hence, it is important to have at least a rough estimate of the promoter dynamics, otherwise we may filter out fast dynamics due to the reporter protein (54). In theory a very unstable protein, with a very short half-life, should be the best choice to avoid filtering out fast promoter dynamics, however unstable proteins have a weak fluorescent signal, and hence they increase measurement errors, therefore in practice a balance between half-life and fluorescent intensity must be found to perform successful experiments (54).

Regarding the performance of the different models on the various scenarios, I can draw two main conclusions by inspecting Table 5.1: (1) the FIT index (as defined in §5.1.4), which is the only one out of the three indices independent of the number m of model parameters, is consistently smaller (i.e. worse) in scenario III and IV when compared to scenario I and II. This is to be expected since in scenarios III and IV, the models are identified using a smaller number of samples. Moreover, the prediction error is estimated on samples not used for the identification. The FPE and AIC indices are not so easy to interpret since in each scenario a different model order (i.e. the number m) is chosen (at least for the ARX and the state-space models) and these indices depend also on the model order; (2) by comparing scenarios I and III with scenarios II and IV, I have noticed that using the "ideal" input signal (scenario II and IV) decreased the performance of all the models, but for the ARX. This reduction may indicate that the fluctuations present in the red fluorescent dye measurement, which is proportional to the Galactose concentration, are not due to measurement errors but probably contain some relevant information.

The nonlinear model is seemingly the worst performer, however this has to be attributed to the heuristic algorithm (i.e. simulated annealing) that I have used to estimate model parameters, rather than to the model structure. Indeed, unlike linear models, where it is possible to exactly compute the optimal solution which minimises the cost function, in the case of nonlinear models a heuristic approach must be employed. The parameters of the heuristic method have to be carefully chosen for best performance (i.e. the starting temperature, the cooling function, etc.).

5.2 Modelling GAL1 promoter dynamics in yeast S. Cerevisiae, grey box identification

In §5.1.6 I have analysed the results achieved in identifying different models structure (§5.1.3) starting from data (input and output) directly measured via fluorescence microscopy by applying the so called "black box" identification approach (1). Although all the models obtained were able to describe *GAL1* promoter driven transcription, the high number of parameters (i.e. in ARX and state space models), or the presence of delays as well as nonlinearities in the parameters (i.

e. in transfer function and nonlinear models) make these models difficult to be analysed either to elucidate the intrinsic characteristics of the biological system under exam, or with the end of designing a control law to regulate the dynamics of the *GAL1* promoter.

For this reason, once verified that the experimental and the computational methods above described were effective, I have repeated the identification process; now with the end of estimating parameters of a mathematical model derived as a trade off in between the affinity to the *a priori* knowledge of the biological process and, the simplicity of its structure.

5.2.1 Data set

I have used for this purpose the data depicted in Figure 5.3, where the input for the system is a square wave of Galactose and Glucose and, the output considered is the measure of the average intensity of the fluorescence. Moreover I have produced another dataset by performing a longer experiment (2180 minutes), in which after keeping cells for the first 180 minutes in Galactose, a random sequence of pulses in between Galactose and Glucose has been provided to cells to elicit *GAL1* promoter activity and for the identification I have considered this signal as the input (Figure 5.8, Lower Panel) and the measure of the average fluorescence intensity expressed by the cells as the output (Figure 5.8, Upper Panel).

I have not used any filter on the output data, since the order and the structure of the model considered were fixed, and thus there was any risk of an increase of the model order due to noise over estimation (1).

5.2.2 Candidate model, identification strategy and validation methodology

To fit the behaviour of the system, I chose a model comprising two continuous linear differential equations: the first describing the dynamics of the mRNA produced upon the transcription (Equation 5.10), and the latter modelling the dynamics of the fluorescent reporter (Equation 5.11).

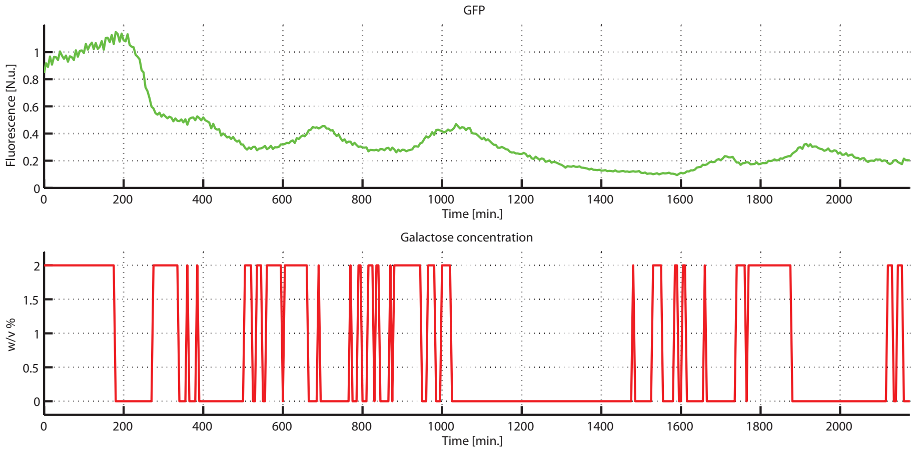


Figure 5.8: Experimental data. Upper Panel: *Cell population average fluorescence (green line) measured during the experiment.* Lower Panel: *Input signal, a high level corresponds to Galactose, whereas a low level to Glucose, after the first* 180 *minutes in which cells are fed with Galactose, the input is calculated as a random sequence of pulses in between Galactose and Glucose.*

$$\frac{dx_1}{dt} = -d_1x_1 + bu \tag{5.10}$$

$$\frac{dx_2}{dt} = v_2x_1 - d_2x_2 \tag{5.11}$$

In the Equation 5.10, u is the only external input to the model and it is assumed to be equal to 2 when cells are fed with Galactose, whereas, when Glucose is provided to yeasts, it is assumed to be equal to 0 (these values are related to the concentration of Galactose added to the growing medium, §8.6.1); moreover d_1 is a degradation coefficient for the mRNA and b, the coefficient of the input u, is its production rate. In the Equation 5.11, d_2 is the degradation rate of the fluorescent reporter and, v_2 is the production rate of the Gfp.

I have considered four different scenarios for the estimation and the validation of the parameters of Equations 5.10 and 5.11:

- **Scenario I**

  The parameters of the model were estimated on the dataset depicted in Figure 5.3 .The same dataset was used for the validation

- **Scenario II**
  The parameters of of the model were estimated on the dataset depicted in Figure 5.3 .The cross-validation was completed on the data depicted in Figure 5.8

- **Scenario III**
  The parameters of the model were estimated on the dataset depicted in Figure 5.8 . The same data were used for the validation process.

- **Scenario IV**
  The parameters of of the model were estimated on the dataset depicted in Figure 5.8 .The cross-validation was completed on the data depicted in Figure 5.3

Thus at the end of the identification process I obtained two different linear models inferred using experimental data represented in Figures 5.3 (Scenarios I and II) and 5.8 (Scenarios III and IV); parameters, and the initial conditions $x_1(0), x_2(0)$, have been estimated with the PEM method (1). Models performances in reproducing the corresponding data used for the identification and the other data set available (cross-validation) have been assessed by calculating the indices reported and explained in §5.1.4

5.2.3 Results

The results achieved in the estimation of model parameters across the different scenarios are described below, the validation results are summarised in Table 5.2:

- *Model 1* - Scenarios I and II: $[d_1, v_2, d_2, b, x_1(0), x_2(0)] =$ $[0.0047, 0.0078, 0.0124, 0.0035, 1.0617, 1.2211]$, (Figure 5.9, Panels A and B)

- *Model 2* - Scenarios III and IV: $[d_1, v_2, d_2, b, x_1(0), x_2(0)] =$ $[0.0063, 0.0274, 0.0166, 0.0018, 1.0343, 1.0424]$, (Figure 5.9, Panels C and D)

	Scenario I	**Scenario II**	**Scenario III**	**Scenario IV**
FPE	0.0043	0.0148	0.0032	0.0124
AIC	-5.3900	-4.2131	-5.7500	-4.3901
FIT (%)	67.18	40.88	78.84	58.26

Table 5.2: Values of the indices defined in §5.1.4 across different scenarios.

The two models built with the grey box identification approach, although without any term accounting for an explicit delay between input and output, are able to predict the experimental data across all the identification and validation scenarios (Figure 5.9 A-D).

Moreover the two input signals (Figure 5.3 and 5.8, lower panels), here used for the identification, have different frequencies (number of switches per time unit), the second higher than the first; this results in different dynamical properties for the two models obtained.

To explore these differences it is worth to recapitulate and use some basic concepts of System Theory: a) the *step response* of a system is the time behaviour of the output when the input changes from 0 to 1 in a very short time (*step input*), b) the *time constant*, indicated with τ is the parameter characterising the *step response* of a linear system; the smaller is the *time constant*, the faster is the response to input variations (37). In the case of state space linear systems, with order greater than one, the time constant τ can be calculated as the inverse of the smaller eigenvalue associated to the system itself; in this case, due to model structure where:

$$\dot{x(t)} = Ax(t) + Bu(t), \quad x(0) = x_0 \tag{5.12}$$

with the matrix A that is lower diagonal, given by:

$$A = \begin{pmatrix} -d_1 & 0 \\ v_2 & -d_2 \end{pmatrix} \tag{5.13}$$

it is possible to demonstrate that the eigenvalues are equal to the elements on the diagonal, and for both the models the smaller eigenvalue is $-d_1$ (mRNA degradation rate). Thus *model 1* has a *time constant* $\tau_1 = 212\,$minutes and *model*

2 has $\tau_2 = 159$ minutes. It is not surprising that the *model 2*, identified using the input signal with the highest frequency, is the faster. This explains why, in the cross validations scenarios II and IV, the second model performs better than the first; *model 2*, being inferred starting from the high frequency input, is capable of responding promptly even to a slower signal, conversely *model 1*, for fast stimuli, behaves as a filter not reproducing properly system behaviour in response to high frequency signals. Furthermore, as it is possible to appreciate from Figure 5.8, the input signal (lower panel) and cell fluorescence (upper panel) are highly correlated, thus the most reliable model describing this system behaviour is *model 2*.

Interestingly the results obtained point out, for both the models, that dynamics of the system, in terms of responsiveness to input variations, are governed by the degradation of the mRNA and not by that of the fluorescent reporter (mRNA slower than the protein). This is due to the reduced (only two equations) linear structure of the model that, despite its simplicity is capable to capture in satisfactory way *GAL1* promoter dynamics.

Comparing the performance indices (Table 5.2) with those calculated for the black box models and the nonlinear model (Table 5.1), it is possible to notice that the models here identified, and validated in Scenarios I and III, offer similar performances to, state space models identified as black boxes. This is remarkable since state space black box models were much more complex (higher order). A simpler model offers the intrinsic advantage to be easily manipulated, not only to predict system response to several input but, more interestingly, to devise and test feedback control strategies to steer process output towards desired trends over the time.

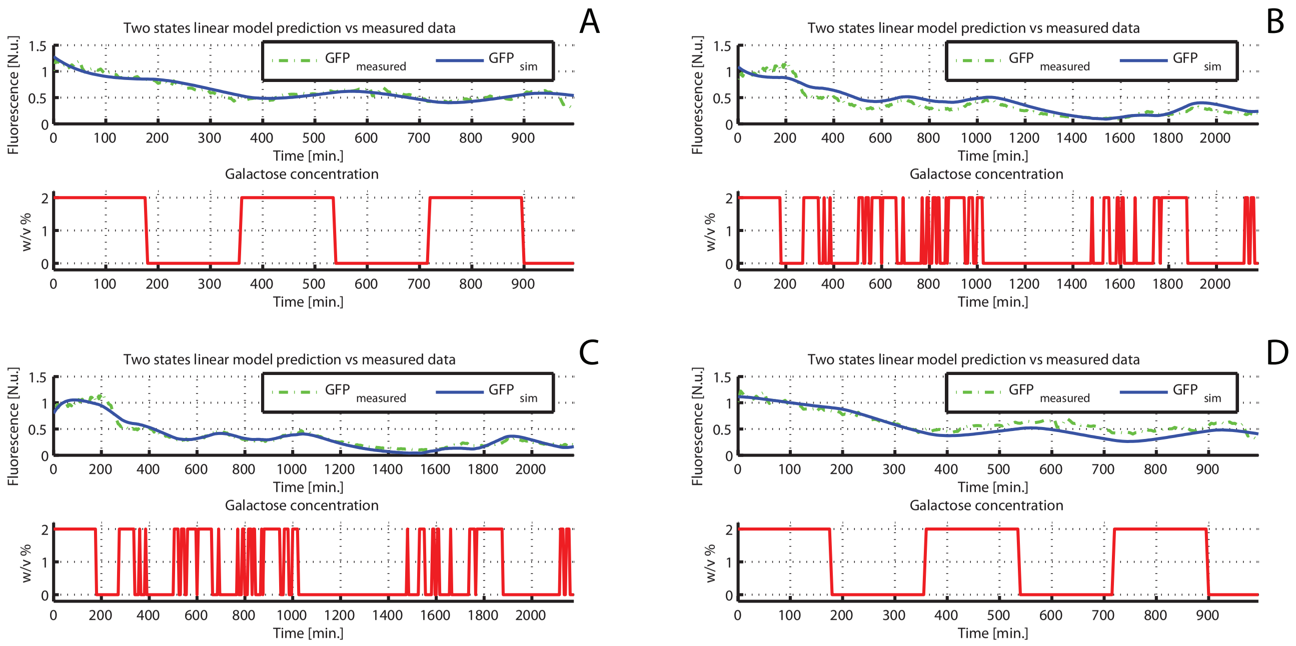


Figure 5.9: Identification and validation results of two state space linear models. Panel A: *The solid blu line represents the output of the inferred model in response to the input (red square wave) used for the identification, the dashed green line is the measured cell fluorescence.* Panel B: *The model inferred with experimental data of Panel A, has been validated in reproducing the measured cell fluorescence (dashed green line) obtained in response to the selected input (red square wave); the result of model simulation, carried out with the same input, is represented by the solid blue line* Panel C: *The solid blu line represents the output of the inferred model in response to the input (red square wave) used for the identification, the dashed green line is the measured cell fluorescence..* Panel D: *he model inferred with experimental data of Panel C, has been validated in reproducing the measured cell fluorescence (dashed green line) obtained in response to the selected input (red square wave); the result of model simulation, carried out with the same input, is represented by the solid blue line*

5.3 IRMA: a complex synthetic network embedded in S. Cerevisiae

IRMA (In-vivo Reverse engineering Method Assessment) was developed as a testbed synthetic network in yeast for the design and validation of reverse engineering and modelling approaches (11). It consists of 5 genes regulating each other via positive and negative feedback loops, and represents one of the most complex synthetic networks built so far (55). The Cbf1-Gfp fusion protein is expressed from the *HO* promoter controlled by two transcription factors: a cell cycle-independent Swi5p mutant (swi5AAA) and Ash1p. The network comprises a transcriptional positive feedback loop from *CBF1* back to itself, via *GAL4* and *SWI5*; and a transcriptional negative feedback loop via *ASH1*. A further regulation is present between *GAL80*, *GAL4* and *SWI5*, whose expression is driven by the *GAL10* promoter, bound by GAL4p. The network can be "switched on" by administering Galactose (GAL) in the medium, which allows *SWI5* to be transcribed by the *GAL10* promoter, or "switched off" by Glucose.

Of note, *CBF1-GFP* expression is delayed with respect to the other genes (11). This delay is due to the sequential recruitment of chromatin-modifying complexes at the *HO* promoter, which follow binding of Swi5p and other transcription factors (56), and it is estimated in the range of 100 min (11).

Galactose and Glucose can be used to control the network's dynamics, which, in turn, can be tracked by estimating the fluorescence level of Cbf1-Gfp, one of IRMA's proteins. Interestingly, IRMA dynamical properties are commonly observed in endogenous gene regulatory networks and pathways. IRMA contains two of the most common regulatory motifs found in eukaryotic cells, i.e. positive and negative transcriptional feedbacks loops (57). Moreover, a protein-protein regulatory interaction is also present, which is much faster than transcriptional regulatory interactions, thus adding concurrent dynamics at different time-scales typical of endogenous regulatory networks.

To capture the dynamics of the network a hybrid model (Figure 5.11) approximating the dynamics in Glucose (F_1) and Galactose (F_2), has been readapted from (11). Both the vector fields F_1 and F_2 share the same model structure as well as most of the parameters ($\hat{v}_3$, $\hat{k}_4$ and $\hat{\gamma}$ need a specific argumentation) as reported below:

$$\frac{dx_1}{dt} = \alpha_1 + v_1 \left(\frac{x_3^{h_1}(t-\tau)}{(k_1^{h_1} + x_3^{h_1}(t-\tau)) \cdot \left(1 + \frac{x_5{}^{h_2}}{k_2^{h_2}}\right)} \right) - d_1 x_1 \tag{5.14}$$

$$\frac{dx_2}{dt} = \alpha_2 + v_2 \left(\frac{x_1^{h_3}}{k_3^{h_3} + x_1^{h_3}} \right) - d_2 x_2 \tag{5.15}$$

$$\frac{dx_3}{dt} = \alpha_3 + \widehat{v_3} \left(\frac{x_2^{h_4}}{\widehat{k_4}^{h_4} + x_2^{h_4}(1 + \frac{x_4^4}{\widehat{\gamma}^4})} \right) - d_3 x_3 \tag{5.16}$$

$$\frac{dx_4}{dt} = \alpha_4 + v_4 \left(\frac{x_3^{h_5}}{k_5^{h_5} + x_3^{h_5}} \right) - d_4 x_4 \tag{5.17}$$

$$\frac{dx_5}{dt} = \alpha_5 + v_5 \left(\frac{x_3^{h_6}}{k_6^{h_6} + x_3^{h_6}} \right) - d_5 x_5 \tag{5.18}$$

where $x_1 = [CBF1GFP], x_2 = [GAL4], x_3 = [SWI5], x_4 = [GAL80], x_5 = [ASH1]$ are the system states. Hill functions have been used to model transcription rates from promoters; the multiple regulation on *CBF1* is modelled by the product of two Hill functions (AND regulation). A time delay τ is present in the equation for x_1 modelling the transcription of *CBF1*, which is affected by a 100 minute-long time delay due to the sequential recruitment of chromatin-modifying complexes to the *HO* promoter (which follows binding of *SWI5* and other transcription factors) (56). A list of all model parameters can be found in Supplementary Table S1 in (11).

Note that the model is hybrid as parameters $\hat{v}_3, \hat{k}_4$ and $\hat{\gamma}$ switch between two different sets of values depending on the carbon source (Galactose or Glucose).

To assess the predictability of the mathematical model derived from (11) I have performed several "switch off" experiments on cells integrating IRMA network. I have loaded into a microfluidic chip (see Chapter 8 for the complete procedure) IRMA cells coming from a Galactose culture; thus these cells were expected to express the Cbf1-Gfp protein at its high steady state. I have fed these yeasts for 180 minutes with Galactose (ON signal, 1 for the mathematical model) and for 420 minutes with Glucose (OFF signal, 0 for the mathematical

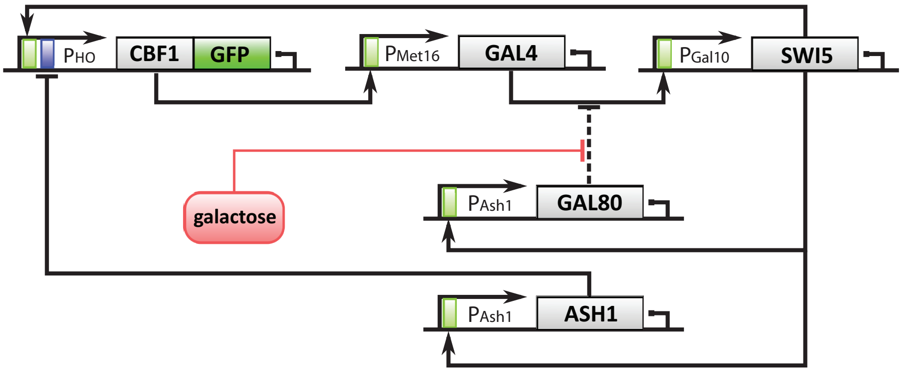


Figure 5.10: IRMA synthetic network topology. IRMA is composed of 5 genes encoding for transcription factors modulating the expression of each other. Both the transcription factors in the network and the promoters driving their expression are shown (adapted from (11)). Solid lines model transcriptional interactions, while dashed lines are meant to represent protein-protein interactions.

model) and after fluorescence quantification (as reported in §8.3) I have compared the experimental *in vivo* results with data obtained *in silico* by simulating the network's mathematical model. As it is possible to appreciate from Figure 5.12, the model is able to capture well both the timescales and the dynamical range of variation of fluorescence during network's switch off.

5.4 An inducible Positive Feedback Loop stably integrated in mammalian cell line

As mentioned in §2.1, the Positive Feedback Loop (PFL), conversely with respect to the NFL, slows down response times and increases cell to cell variability; Siciliano and colleagues in (25) has provided the experimental proof of this assumptions by stably integrating, and modelling, an inducible synthetic PFL circuit in Chinese Hamster Ovary (CHO) cells. They also built another network, by using the same biological parts of the PFL but lacking the positive feedback, that they called NOPFL and used it to perform a comparison between the two circuits, to better elucidate the intrinsic properties of the positive feedback loop.

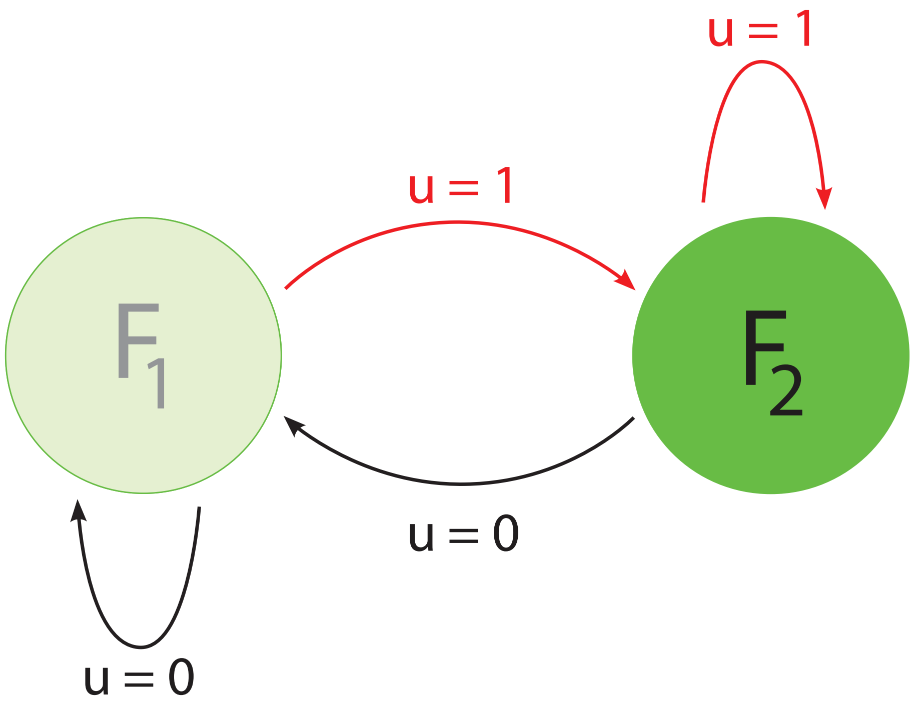


Figure 5.11: IRMA hybrid model. A hybrid model featuring two distinct vector fields (F_1 and F_2) has been derived from the model presented in (11). As long as Glucose is administered ($u = 0$) F_1 is activated, while the system switches to F_2 as soon as Galactose is added to the medium to reflect the inner dynamics of the synthetic circuit.

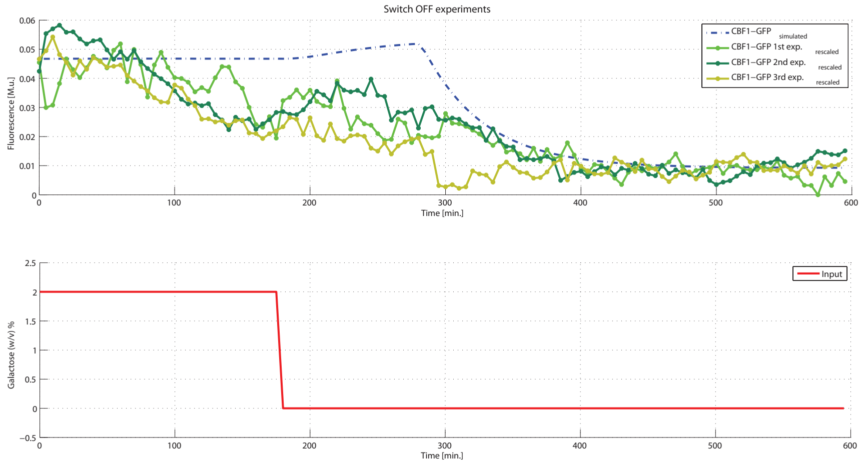


Figure 5.12: IRMA switch off experiment. Top panel: the green signals represent the measured fluorescence during *in-vivo* switch - off experiments, the blue signal is the result of *in-silico* switch off experiment using the dynamical model of IRMA (all the experimental signals are rescaled to the model range). Bottom panel: the input used to perform the experiment; cells have been fed for 180 minutes with Galactose (ON signal, 1 for the mathematical model) and for 420 minutes with Glucose (OFF signal, 0 for the mathematical model).

The PFL has been implemented, achieving a complete control of its behaviour, by using well known and characterised regulators of gene expression. The authors (25) exploited the properties of the Tet regulatory system: the expression of the Tetracycline-controlled transactivator tTA is self-controlled by a *CMV-TET* promoter, responsive to the tTA itself unless the Tetracyline (or Doxycycline) is added to the medium in which cells are grown. To follow the dynamics of this circuit Siciliano and colleagues placed a fluorescent reporter, the destabilised yellow-green variant of the enhanced green fluorescent protein (d2EYFP), under the control of the same promoter (Figure 5.13, Panel A). To devise the NOPFL circuit, as depicted in Figure 5.13 Panel B, Siciliano *et al.* constructed a cassette containing the same *CMV-TET* promoter upstream of the d2EYFP reporter. In this case they placed the tTA protein under the control of a costitutive promoter, thus breaking the feedback loop.

To assess the dynamics of the two circuits in order to derive a mathematical model for both of them, Siciliano and colleagues treated PFL and NOPFL cells with different amounts of Doxycycline in order to ”switch off” the two circuits, by preventing the tTA protein from binding the *CMV-TET* promoter. They imaged cells for more than 40 hours and quantified the fluorescence intensity of the whole cell populations.

The experimental data thus generated were used to infer models of the PFL and NOPFL networks using ODEs. For each of the species (mRNAs and correspondent protein concentrations), authors wrote an equation expressing the change in concentration of the species in a given time interval, as the result od a production term and a degradation term. The resulting model for the PFL is:

$$\frac{dx_1}{dt} = v_1\left(\alpha_1 + (1-\alpha_1)\frac{\left(\frac{\theta^{h_2}}{\theta^{h_2}+D^{h_2}}x_2\right)^{h_1}}{K_1^{h_1} + \left(\frac{\theta^{h_2}}{\theta^{h_2}+D^{h_2}}x_2\right)^{h_1}}\right) - d_1x_1, \tag{5.19}$$

$$\frac{dx_2}{dt} = v_2x_1 - d_2x_2, \tag{5.20}$$

$$\frac{dx_3}{dt} = v_2x_1 - (d_3+K_f)x_3, \tag{5.21}$$

$$\frac{dx_4}{dt} = K_fx_3 - d_3x_4. \tag{5.22}$$

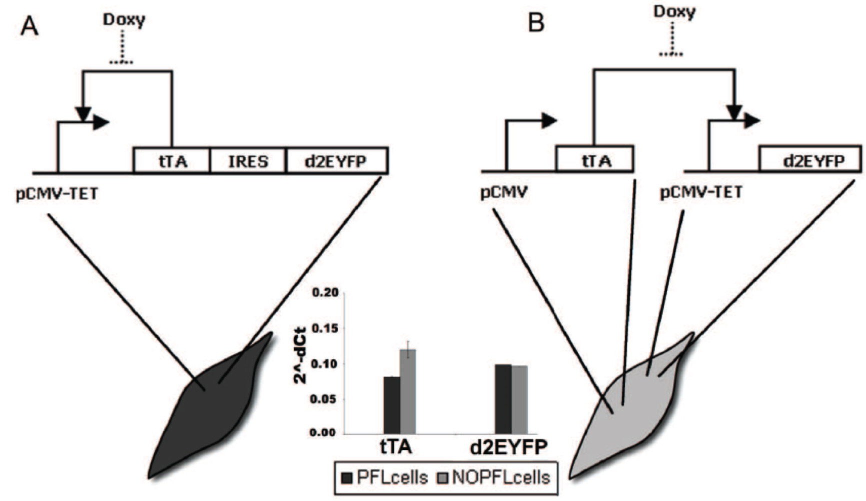


Figure 5.13: PFL and NOPFL topologies. Panel A, PFL: *the promoter CMV-TET consists of seven direct repeats of a 42-bp sequence containing the tet operator sequences (tetO), located just upstream of the minimal CMV promoter (PminCMV). The Tetracycline-controlled transactivator tTA derives from the addition of the VP16 activation domain to the transcriptional repressor TetR. The d2EYFP is the destabilised yellow-green variant of enhanced green fluorescent protein.* Panel B, NOPFL: *the CMV promoter drives the expression of the tTA, which in turns drives the transcription of the d2EYFP from the CMV-TET promoter. (Inset) RealTime PCR performed on DNA extracted from PFL and NOPLF cells shows that the DNA levels of tTA and d2EYFP are comparable among the two clonal cell populations.* From (25)

.

where x_1 is the tTA-IRES-d2EYFP mRNA concentration, x_2 is the tTA protein concentration, x_3 is the unfolded d2EYFP protein concentration and x_4 is the folded d2EYFP protein concentration. The concentrations of tTA and d2EYFP proteins depend on the same mRNA, hence on the same variable x_1. The NOPFL model is very similar to the model of the PFL, except for the fact that here x_1 represents only d2EYFP mRNA and the tTA protein (here constitutively expressed from the *CMVTET* promoter) is assumed equal to a constant value $\bar{x}_2$. The equations thus become:

$$\begin{aligned}
\frac{dx_1}{dt} &= v_1\left(\alpha_1 + (1-\alpha_1)\frac{\left(\frac{\theta^{h_2}}{\theta^{h_2}+D^{h_2}}\bar{x}_2\right)^{h_1}}{K_1^{h_1} + \left(\frac{\theta^{h_2}}{\theta^{h_2}+D^{h_2}}\bar{x}_2\right)^{h_1}}\right) - d_1x_1, && (5.23)\\
\frac{dx_3}{dt} &= v_2x_1 - (d_3+K_f)x_3, && (5.24)\\
\frac{dx_4}{dt} &= K_fx_3 - d_3x_4. && (5.25)
\end{aligned}$$

They estimated 12 parameters, 11 of which were common to both the PFL and NOPLF models; all parameters with the estimation procedure description are extensively discussed in (25).

By looking at the two model equations it is possible to observe that the NOPFL is a linear, time-invariant system, that according to the theory of linear dynamical systems has a dynamic behaviour controlled by the smallest among the three different degradation parameters (d_1, d_3, $(d_3 + K_f)$); this means that any variation in the concentration of Doxycycline would affect only the steady state value reached at the end of the switch off, but not the speed at which that steady state is reached (Figure 5.14). Conversely, for the PFL , as confirmed by the experimental data (Figure 5.14), the concentration of Doxycycline plays a decisive role in determining switch off temporal dynamics In both cases, there is a transcription factor, responsive to an external signal, that activates some target genes; but only in one case, it activates itself as well. A linear response to the inducer might be useful to the cell in all cases when the response activated is transient, maybe an adaptation to a stress stimulus, or to nutrients. But there

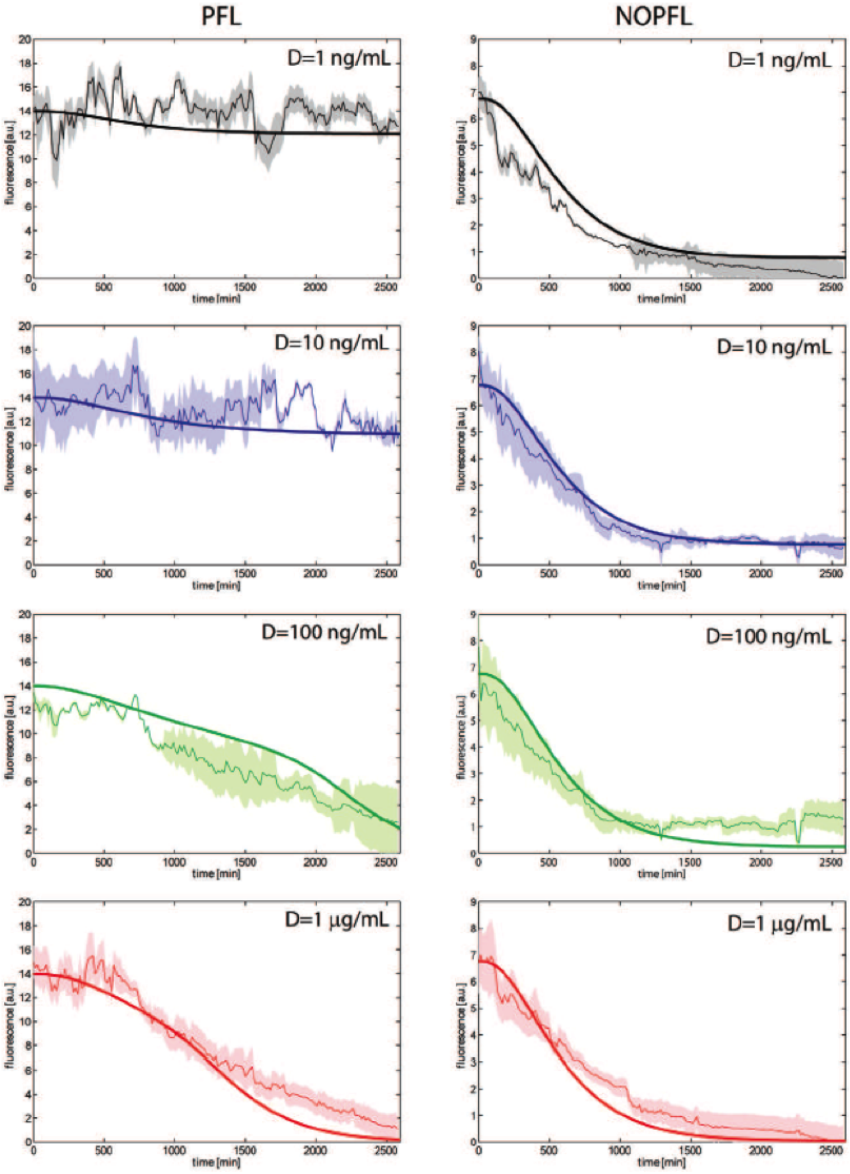


Figure 5.14: PFL and NOPFL experimental and simulated switch off time course. *Experimental data (thin lines) and model simulations (thick lines) were reported for the PFL (left) and NOPFL (right) cells. Shaded areas represent standard deviations from replicate experiments.* From (25)

.

are circumstances when the downstream effect of the transcription factor has a fundamental effect on the life of the cell, for example if it triggers irreversible events such as differentiation. In these cases, the cell might gain more by waiting than by responding quickly to any signal; if, and only if, the signal is prolonged in time, then the cell will respond to it. Positive feedbacks on transcription factors might indeed have evolved to such purpose. This behaviour has been described in (57) as "persistence detection", thus as a way to distinguish between persistent and transient stimuli in cell signalling. A parallel with control engineering, where positive feedbacks are used to generate memory circuits, can be done in this case: these systems, i.e. "switches", are able to exist stably in two different states (ON or OFF), without inadvertently being altered by transient perturbations

6

In vivo feedback control of endogenous and synthetic circuits in yeast

In this Chapter I describe the results achieved when controlling the level of expression of a reporter protein fused to the Gal1p protein from the endogenous *GAL1* promoter (Figure 6.1 a) and in the complex synthetic network IRMA (Figure 6.1 b). Furthermore, I have used these promising results as a starting point to improve the control law adopted to accomplish the regulation task and, to test and compare other feedback control strategies by assessing their performances in regulating expression level from the *GAL1* promoter. I have carried out *in-silico* (numerical simulations) and *in-vivo* experiments to validate the implementation of those strategies and, to investigate which of them was able to guarantee the best result (§6.2) Part of this work has been published in (58)

6.1 *In-vivo* Proportional Integral (PI) control of *GAL1* promoter and IRMA network

Topologies and dynamics of the *GAL1* promoter and IRMA network constructs, have been deeply described in §5.1 and §5.3. Here I propose an analysis of the two systems from a Control Engineering perspective, specifying which is the control

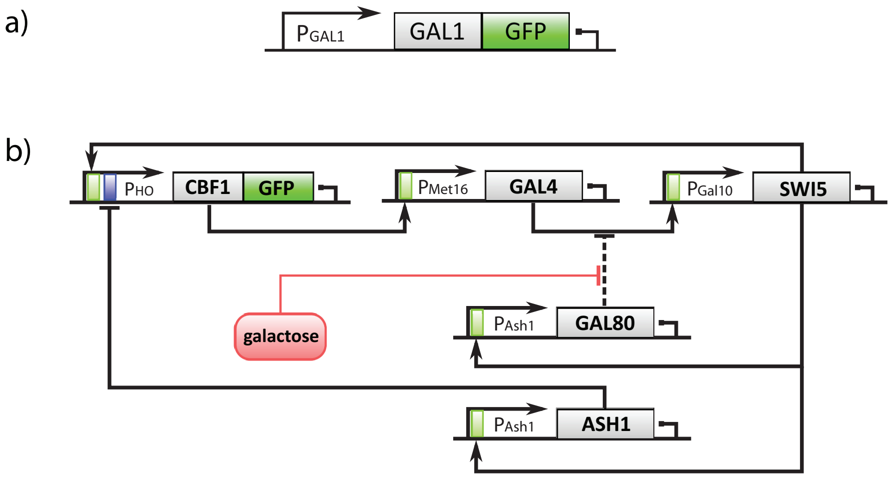


Figure 6.1: Biological systems. (a): *The Gfp protein was integrated downstream of the endogenous GAL1 promoter (yeast strain courtesy of Prof. Botstein lab). Described in §5.1* (b): *IRMA is composed of 5 genes encoding for transcription factors modulating the expression of each other. Both the transcription factors in the network and the promoters driving their expression are shown; solid lines model transcriptional interactions, while dashed lines are meant to represent protein-protein interactions. Described in §5.3*

objective and the requirements to accomplish to achieve it.

GAL1 promoter: As already mentioned in §5.1, the *GAL1* promoter drives the expression of the Gal1-Gfp fusion protein in yeast S. Cerevisiae (Figure 6.1 a). It can be viewed as a single input-single output (SISO) dynamical system. The input $u(t)$ describes the presence of Galactose or Glucose in the growth medium. The output $y(t)$ is the measured average level of fluorescence of the Gal1-Gfp protein in the cell population. Cells can respond either to Galactose or Glucose, but not to an intermediate concentration of them. This is due to the fact that cells can consume Glucose at a lower energetical cost (50), thus as soon as Glucose is administered to the cells these stop responding Galactose, even if it is still present in the medium. Thereby the control input (interpreted as Galactose concentration of 2 w/v% in the total volume of fluid reaching cells) is restricted to be either ON (Galactose) or OFF (Glucose).

IRMA network: IRMA can be modelled as an input-output system where the input u models the presence/absence of Galactose and the output y is the concentration of one of its genes, namely *Cbf1* (x_1, Equations 5.14 - 5.18). Note that the input acts nonlinearly on the dynamics of the network as the presence of Galactose changes the values of all the Galactose-dependent parameters (namely $\hat{v}_3, \hat{k}_4$ and $\hat{\gamma}$, Equation 5.16). As in the case of the *GAL1* promoter cells do not sense intermediate concentrations of the two sugars, therefore, the control input is restricted to be either ON (Galactose) or OFF (Glucose). The system output $y = x_1$ cannot be measured directly as a concentration. Instead, the cells were engineered so that CBF1p is fused with a GFP, the green fluorescent protein (11). In this way, higher concentrations of Cbf1p are associated to higher levels of fluorescence. From a control perspective, the gene network model is, therefore, a highly nonlinear, hybrid, time-delayed dynamical system of the form:

$$\dot{x} = \begin{cases} F_1(x, x(t-\tau), \mu), & \text{if } u = OFF, \\ F_2(x, x(t-\tau), \hat{\mu}), & \text{if } u = ON \end{cases}$$

where $x = [x_1 \ x_2 \ x_3 \ x_4 \ x_5]^T$, μ is the vector of parameter values in Glucose ($u = OFF$) and $\hat{\mu}$ is the vector of parameter values in Galactose ($u = ON$).

Hybrid systems are often used to model gene networks (e.g. see (59, 60, 61)), where it is quite common to observe threshold dependent and switch-like activation or inhibition functions governing the dynamics of protein-protein or protein-gene interactions.

6.1.1 Control objective and controller design

The control objective for both systems was a set-point regulation, where the cell populations were required to express, over several generations, a constant amount of fluorescence (control reference $r(t)$).

When dealing with living cells, one of the major issues is represented by the uncertainty affecting transcriptional and translational processes, introducing a remarkable cell-to-cell variability in mRNA and protein production. One of the way to account for this problem is to consider as the system output the average fluorescence intensity expressed by all cells, thus dampening the effects due to noisy measurements.

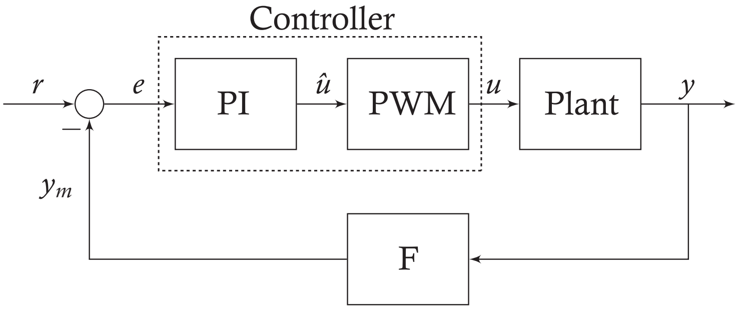


Figure 6.2: PI-PWM feedback control scheme. The controller consists of a Proportional-Integral (PI) block followed by a Pulse Width Modulation (PWM) block encoding of the control input $\hat{u}$. The PWM transforms the continuous control action $\hat{u}$ into a train of rectangular pulses u, which represents either Galactose (high) or Glucose (low); to overcome drawbacks introduced by the saturating effect added by the PWM an anti-windup compensation scheme (as the one reported in Figure 2.7) is added. The alternating series of Glucose and Galactose pulses is applied to the cell population to be controlled (*Plant*), whose output y (the controlled variable) can be filtered (y_m) by a low-pass filter (F) before being fed back to the controller, to dampen the effect of noisy measurements. The difference between y_m and its desired reference level r, namely the error e, is used by the PI controller to compute the control input to be supplied to the system to minimise the error signal e.

Furthermore, from the analysis presented in the previous paragraphs, it is possible to desume which are the main constraints that the control law has to fit to be implemented on these two biological circuits:

- It is possible to either feed only Galactose or only Glucose to cells.
- The output can be measured only via the Gfp fluorescence intensity.
- The control action should be robust to parameter variation, biological noise and external disturbances.

Moreover, in this initial stage the control law should be as simple as possible, in order to not add complexity to the control scheme and further pitfalls to performance evaluation.

To this end, I designed a control algorithm based on a Proportional-Integral (PI) regulator, whose output $\hat{u}(t)$ is a function of the control error $e(t)$ (the mismatch between the desired and current output of the system $e(t) = r(t) - y(t)$) defined as:

$$\hat{u}(t) = K_p \cdot e(t) + K_I \cdot \int_0^t e(\tau) d\tau \tag{6.1}$$

Where the parameters K_p and K_i have to be tuned to optimise controller yield.

The constraint on the control input (Galactose ON, Glucose OFF) allows an analogy with the problems faced in the design of feedback control strategies for power electronic circuits (62). Here, switches and SCRs (silicon controlled rectifiers) can only be turned on or off, some output is typically measured or estimated and, particularly in industrial applications, compensating noise and external disturbances is of utmost importance. The simplest and most widely used control technique in this context is to use the PI regulator coupled to a PWM (Pulse Width Modulation) control strategy (63). This is also the strategy I adopted to control the cell population. In the simplest feedback implementation of the PI-PWM, a sawtooth signal is compared with $\hat{u}(t)$ (Figure 6.2) in order to modulate the width of a rectangular pulse train, which is then used as control input (see Figure 6.3 and (64) for further details). Namely, let

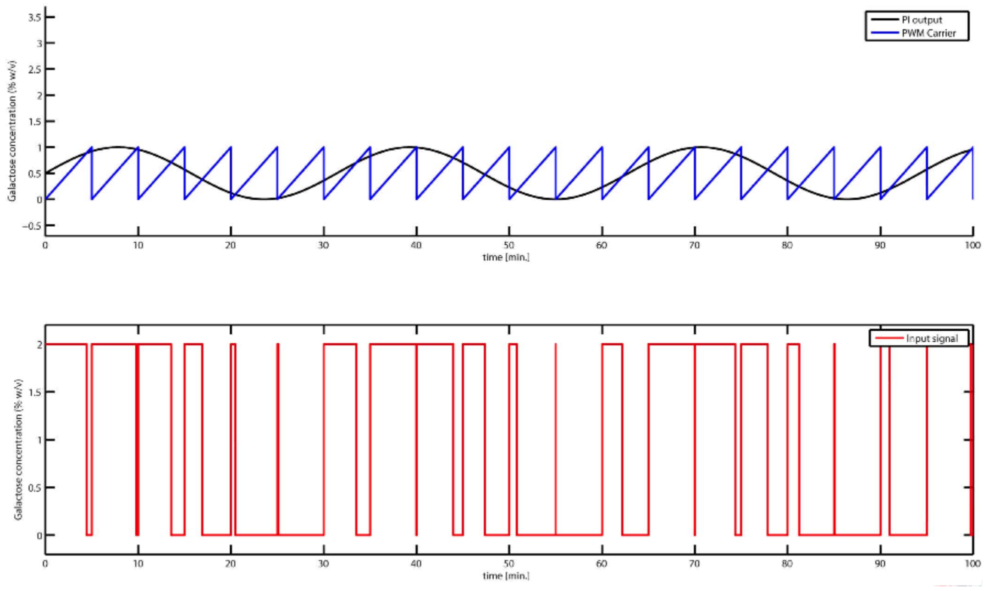


Figure 6.3: Pulse With Modulation (PWM) strategy . *A continuous signal (black line) is compared to a sawtooth modulation waveform (blue line), whose frequency is chosen appropriately. When the value of the continuous signal is greater or equal than the modulating waveform, the output of the PWM is in the ON state, otherwise it is OFF (red line).*

$$\eta(t) = \alpha + \beta(t \bmod T) \tag{6.2}$$

be the sawtooth signal; then

$$u(t) = \begin{cases} OFF, & \text{if } \eta(t) - \hat{u}(t) > 0, \\ ON, & \text{otherwise} \end{cases}$$

I have tuned all the parameters of the PI-PWM (K_p, K_i, α, β and T) for the *GAL1* promoter and IRMA control separately.

As concerns the period T of the sawtooth signal used by the PWM strategy (Equation 6.2), for both the systems to be controlled, I have fixed it as $T = 5\ min$; this time interval is equal to the image acquisition time lapse, that I chose as an ideal trade off to avoid phototoxicity effects to cells and, at the same time, to guarantee a sufficient measure resolution to follow fluorescence variation dynamics (44).

The presence of the PWM introduces a saturation on the PI output $\hat{u}$. As already described in §2.3, the presence of a saturation downstream to the PI regulator can lead to a break of the feedback loop thus affecting controller performances. To overcome this issue I have implemented the anti-windup scheme described in §2.3 (Figure 2.7), I have chosen the feedback gain $K_t = 1$ so that the integrator is reset instantaneously once the signal $\hat{u}(t)$ saturates.

GAL1 promoter: I chose the sawtooth wave parameters as follows: $\alpha = 0$, $\beta = 2$ and $T = 5min$. The gains of the PI controller, namely $K_p = 6$ and $K_i = 0.3$, were tuned by applying the time domain Ziegler-Nichols' tuning method (37) to the linear transfer function describing the *GAL1* promoter previously derived in §5.1 (the best performing model, whose parameters have been estimated in Scenario I):

$$G(s) = \frac{K_p}{1 + sT_p} e^{-T_d s}. \tag{6.3}$$

with parameters $K_p = 0.36$, $T_d = 57.82$ and $T_p = 94.47$.

IRMA network: As in the case of the *GAL1* promoter, I used the PI-PWM control strategy for the set-point control task. The sawtooth wave parameters for the PWM were set to $\alpha = 0$, $\beta = 10E - 5$ and $T = 5\ min$. A Proportional-Integral controller takes e in input and computes the control signal $\hat{u}(t)$ with the gains $K_p = 175.6$ and $K_I = 2.11$. These gains were found, as previously described for the *GAL1* promoter, by applying the time domain Ziegler-Nichols' tuning method (37) to an approximation of the system in the form of a linear transfer function derived evaluating its step response:

$$G_{approx}(s) = \mu \frac{e^{-ds}}{1 + \Theta s} \tag{6.4}$$

The parameters of the transfer function in eq. (6.4) were found to be $\mu = 0.0467$, $d = 146.85$ and $\Theta = 667.62$.

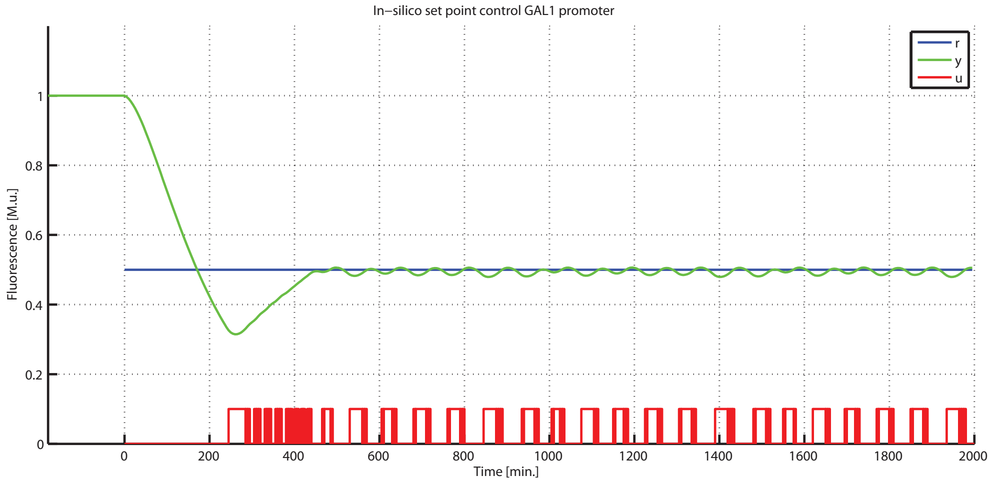


Figure 6.4: In-silico PI-PWM set point control of *GAL1* promoter. *The PI-PWM control algorithm is applied to control the dynamical model of the GAL1 promoter to a constant reference signal (r in blue). The set point is equal to* 50% *of the maximum value for the simulated Cbf1 time evolution evaluated until* $t = 0$min. *The control input, computed after time* 0*, is shown in red (u high level: Galactose; low level: Glucose). The simulation, as explained in the text, was performed by controlling the dynamical model 2, inferred in* §*5.2.*

6.1.2 *In-silico* validation

***GAL1* promoter:** To validate the PI-PWM control strategy *in-silico*, as a proxy for the system behaviour, I have used a mathematical model of the *GAL1* promoter activity; specifically I have applied the designed feedback control to the two variable state - space linear model inferred in ”Scenario III” in §5.2, that I have named *model 2*. Thus the control objective was to regulate the output y of *model 2*, to reach and maintain a given reference signal r.

As shown in Figure 6.4, the controller was able to control the output of the model to the desired value; the output y oscillates around the reference r, and this is mainly due to the switching control input produced by the PWM.

IRMA network: As in the case of the *GAL1* promoter, for the IRMA network, I have assessed the performances of the feedback control law by using the mathematical model derived in (11) and described in 5.3. The delay term present

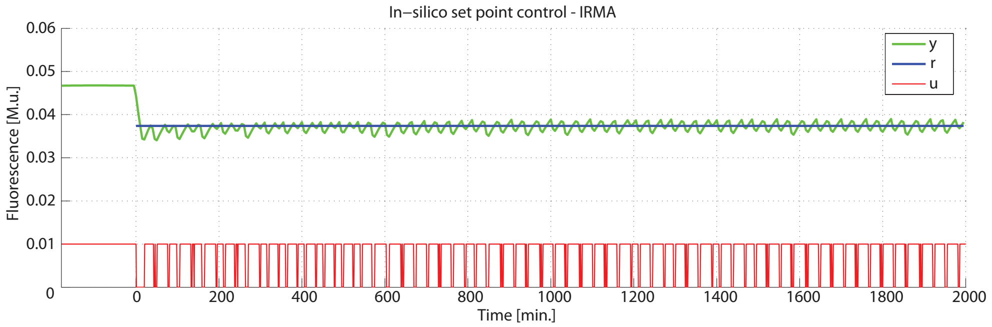


Figure 6.5: In-silico PI-PWM set point control of IRMA. *The PI-PWM control algorithm is applied to control the dynamical model of IRMA to a constant reference signal (rin blue). The set point is equal to* 75% *of the maximum value for the simulated Cbf1 time evolution evaluated until* $t = 0\text{min}$*. The control input, computed after time* 0*, is shown in red (u high level: Galactose; low level: Glucose). The simulation, as explained in the text, was performed by controlling the dynamical model without delay.*

in this mathematical model (Equation 5.14), models the time required for the activation of the *HO* promoter driving expression of the *CBF1* gene in the network, that was quantified to be equal to 100 min (11). The quick indirect activation via Galactose and Glucose switches, could prevent the promoter to be completely silenced via chromatin remodelling, thus considerably reducing the transcriptional delay. For this reason, to simulate the control feedback, I have removed the delay from the model. As shown in Figure 6.5, the PI-PWM control strategy is effective in achieving the control objective (keeping the output of the model close to the set point).

6.1.3 *In-vivo* experiments

Once I assessed the controllers *in-silico*, I substituted, in the feedback loop, the model of the two biological circuits with the real cells growing in the microfluidic device.

Controller implementation: To perform *in - vivo* control experiments, I integrated the devised control law with the experimental platform. At the beginning of each experiment cells are loaded in the microfluidic device, and are fed for 180 min with Galactose. This was done for two main reasons: a) to allow them to adapt to the microfluidic environment and b) to calculate the high steady - state of fluorescence since the reference signal is expressed as a percentage of this value; I called this part of the experiments *calibration phase*. (further details in §8).

To integrate the PI-PWM based control strategy with the experimental platform, I implemented it as a Finite State Automaton (FSA) in MATLAB programming environment. The FSA works as follows: after the calibration phase, at each control step (k) an image is acquired by the microscope, and the normalised fluorescence signal is computed thanks to the image processing algorithm described in §8. The fluorescence signal $y(k)$ is compared against the reference signal $r(k)$, to obtain the error $e(k)$. The control input $u(k)$ is then computed using the discrete-time implementation of the PI controller discussed in (63).The control input $u(k)$ is used to determine the duration of the pulse of Glucose or Galactose by means of the PWM strategy. The duration of each pulse corresponds to the time interval during which the syringe loaded with Galactose remains higher than the one containing Glucose (or vice-versa). At the next instant ($k + 1$) a new image is acquired and the feedback computation takes place. The error $e(k + 1)$ is available for a new control iteration and each step is repeated again. A pseudocode implementation of the FSA is reported Algorithm 1; where the first *while* construct accounts for the calibration phase and, the second, for the control loop implementation.

GAL1 promoter: The control experiment consisted in a set-point control task, i.e. forcing yeast cells to reach and maintain a constant level of fluorescence equal to 50% of their maximum fluorescence level when grown in Galactose-rich medium.

As shown in Figure 6.6 the control action works effectively in keeping the output, namely the measured fluorescence, close to the desired set-point for 2000 min. Remarkably the results achieved *in-vivo* among all the replicates (Figure 6.6 A-D) are consistent with those of the *in-silico* control (Figure 6.4), the amplitude and the period of the oscillations arount the set point, predicted by numerical

Algorithm 1 FSA pseudocode implementation

```
$k = 0; t = 0; T = 5$
while $t \leq 180$ do
    acquire image;
    process image;
    wait T minutes;
    $t = t + T$
end while
calculate $e(0)$;
$k = 1$
while $k \leq dim(r)$ do
    $u(k) = PI/PWM(e[0, k-1])$
    move syringes;
    wait T minutes;
    acquire image;
    $y(k) =$ process image;
    $e(k) = r(k) - y(k)$;
    $k++$;
end while
```

simulation, are reproduced in the control of the real process. This confirms, once again, that *model 2*, inferred in §5.2 and used to test *in-silico* the PI-PWM control strategy, predicts accurately system dynamics and that the control algorithm integrated with the experimental platform works as desired.

Furthermore taking advantage of the image processing algorithm implemented (§8), I calculated the number of cells within each frame and the fluorescence expressed by each single cell for each of the experiments of Figure 6.6. With these data I calculated the standard deviation σ of the fluorescence for each frame of each experiment and, the coefficient of variation $CV = \frac{\sigma}{\mu}$, (where μ is the average of the fluorescence), that measures the "relative variability" of the fluorescence of each cell compared to the controlled variable (average fluorescence of the whole population). Despite the increasing number of cells and the cell-to-cell variability intrinsic to gene expression, the control error remained bounded, and the CV, for the whole experiment, did not change considerably, and its level is well in the expected range for living cells (65). Hence the population of cells is entrained by the control signal which keeps them from deviating from the reference signal(Figure 6.7 and (58) Supplementary Informations).

To further assess the effectiveness of the feedback control strategy, I compared the feedback control results (Figure 6.6) with two different types of "negative control" experiments: (1) the dataset reported in Figure 5.8 described in §5.2 and used for model inference (after the calibration phase, cells were fed with a random sequence of switches in between Galctose and Glucose); (2) yeast cells fed for 2000 min only with Galactose (sustained "ON" input).

The results of the negative control experiments are shown in Figure 6.8. It can be appreciated that, as expected, when cells were kept in constant Galactose (Figure 6.8) the measured GFP fluctuated and diverged from the initial value; whereas, when a random input was applied (Figure 6.8) the output did not reach the desired value, thus confirming that the the control input calculated via negative feedback is essential to accomplish the control task.

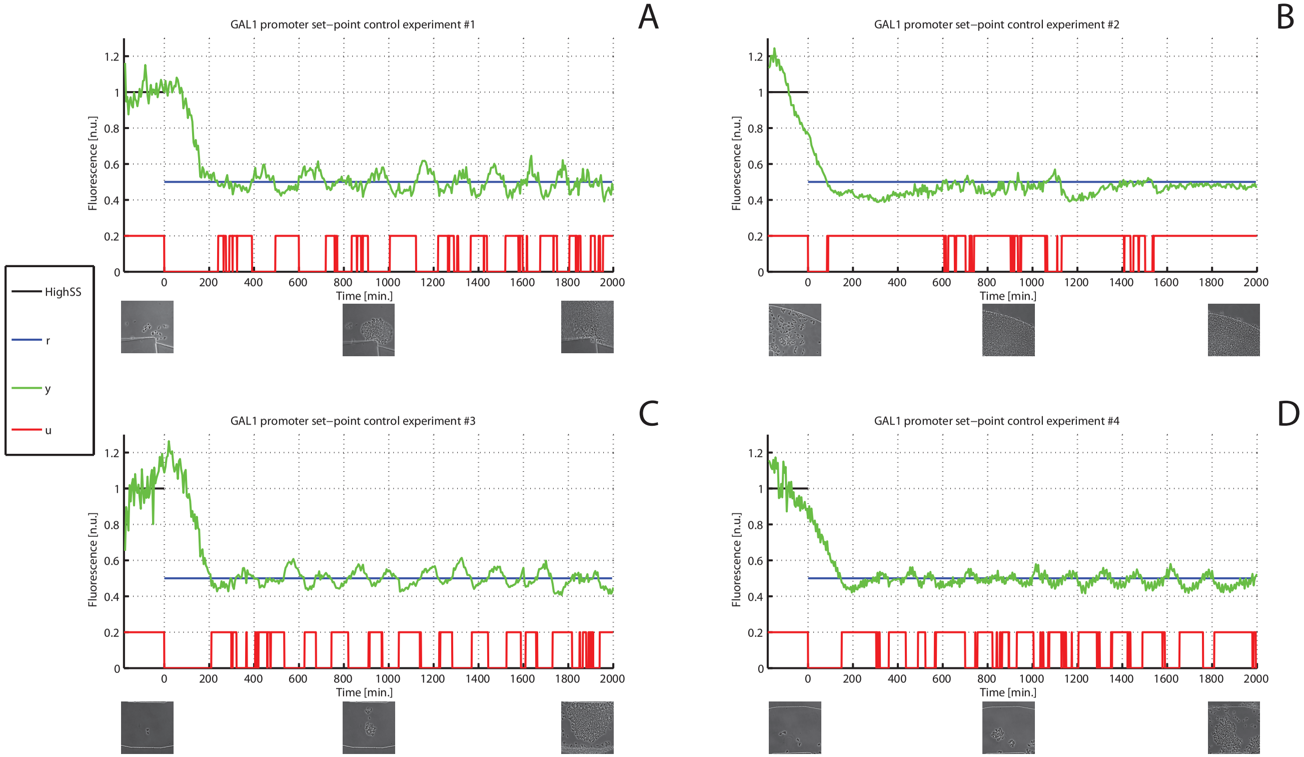


Figure 6.6: In-vivo set point control experiments on the GAL1 promoter. (A-D) *Four in-vivo set point control experiments were performed on the GAL1 promoter. The desired (r in blue) and experimentally quantified GFP fluorescence (y in green) in the cell population are shown for the whole duration of the experiments; the control action starts at time* $t = 140$ min *and lasts for* 2000 min. *The fluctuations in fluorescence during the* 180 min *calibration phase are due to stress response after loading cells in the microfludics device. The input signal u, computed in real-time by the control algorithm, is shown in red: a high signal corresponds to Galactose-rich growth medium, a low signal to Glucose growth medium.* (Insets) *Images taken during the experiments show the growing yeast populations at the beginning, at the half and at the end of each experiment.*

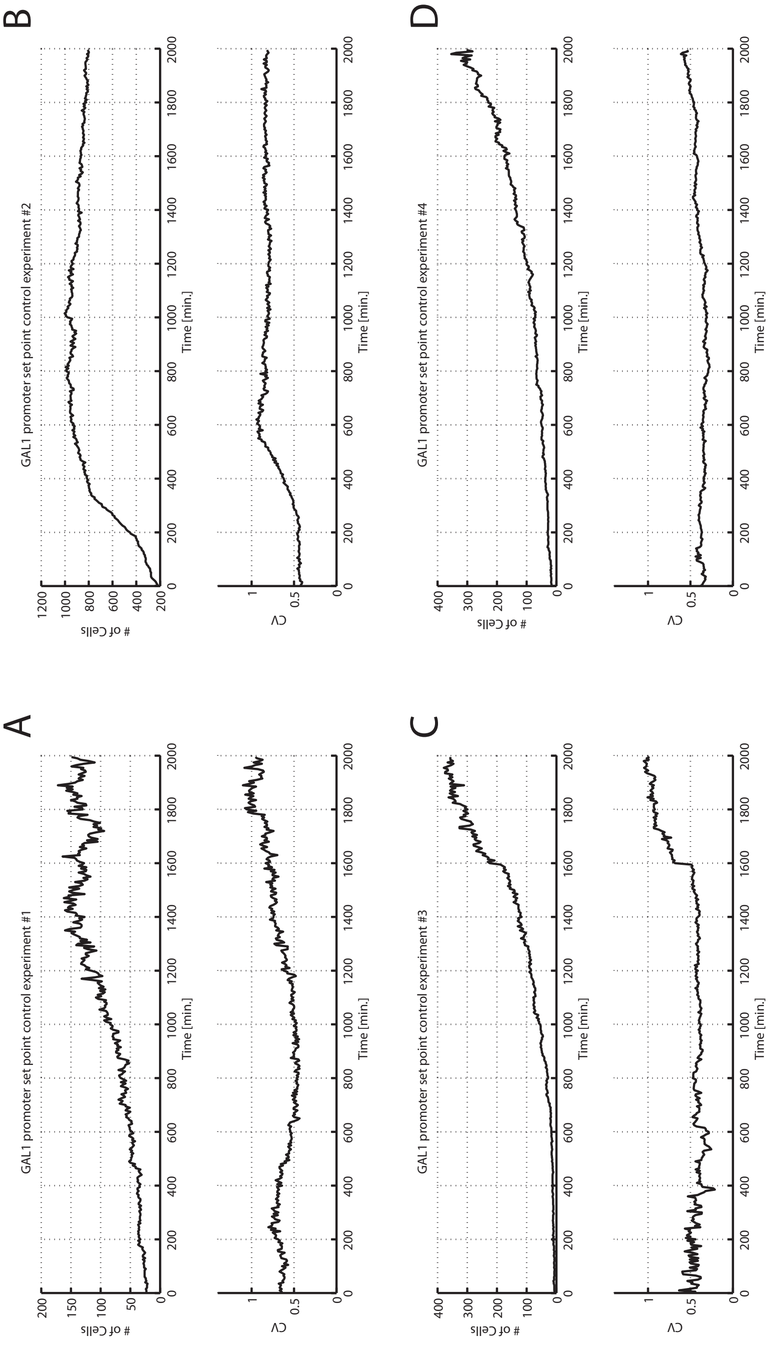


Figure 6.7: In - vivo set point control experiments *GAL1* promoter - Cell count and coefficient of variation. (A-D) For each of the experiments of Figure 6.6 the number of cells (top) and the coefficient of variation (bottom) are shown.

IRMA network: The control objective was the same as for the *GAL1* promoter previously described, i.e. controlling the level of expression of the reporter protein (Cbf1-Gfp). However, in the IRMA network, unlike the *GAL1* promoter system, the *CBF1-GFP* gene is not under the direct control of the inducer molecule (i.e. Glucose or Galactose). Indeed, as shown in Figure 5.10 b, Galactose activates Gal4p, which then drives the expression of *Swi5p* that ultimately binds the *Gal10* promoter driving Cbf1p-GFP expression. This adds a considerable delay in the Cbf1-Gfp activation following Galactose treatment(11).

I performed a set-point control experiment in the IRMA network, where the cell population was required to reach and maintain a fluorescence level equal to 75% of its maximum value in Galactose over a time interval of 2000 min (Figures 6.9 and 6.10). As in the case of the Gal1 promoter, the experiment started with a short calibration phase of 180 min in Galactose to estimate the maximum Cbf1-Gfp fluorescence level produced by the cell population.

As shown in Figure 6.9, the desired fluorescence level was successfully achieved and maintained for over 24 hours, the control error did not diverge and remained bounded around zero. The cell-to-cell variability, estimated using the CV, did not change appreciably throughout the experiment, and was found in the expected range (65), despite the increase in the number of cells (estimated from 25 to 120 cells; Figure 6.11).

As expected, however, due to the more complex network, the fluctuations around the set-point are more evident. In this case, I also performed an additional statistical analysis to test the control action performance in regulating the protein expression level to the desired set-point. Indeed, due to cell-to-cell variability, the fluorescence level in the cell population varies among the cells. Referring to Figure 6.10, I considered two classes of events: (NC) the fluorescence measured in single cells during the first 180 minutes of experiment, when No Control input is applied; (C) the fluorescence measured in single cells after the first 180 minutes of experiment, when the Control action has began. I then compared the control error in class (NC) (dashed black line in Figure 6.10) to the control error in class (C) (solid black line in Figure 6.10) using a one-tail t-test to check if we it was possible to reject the null hypothesis $H_0 = e_{NC} \leq e_C$, where e represents the control error. I obtained a significant $p - value$ of 1.75E-11, that demonstrates

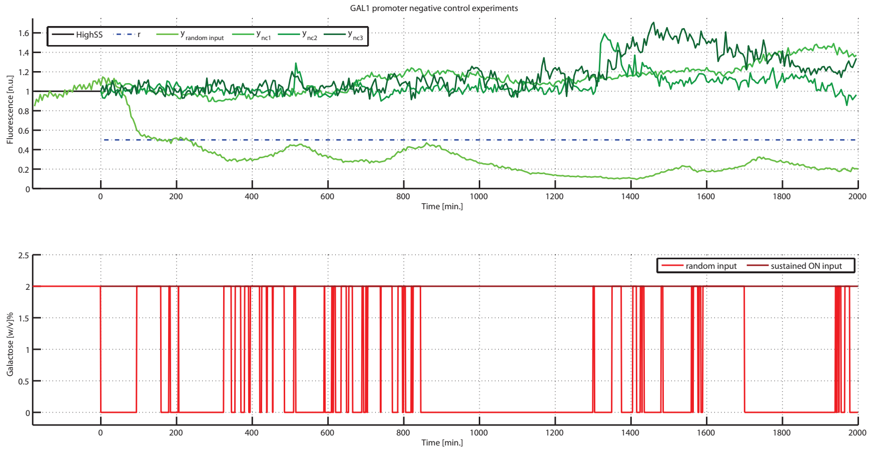


Figure 6.8: In-vivo negative control experiments on the GAL1 promoter. (Top panel) *the three green signals (y_{nc1}, y_{nc2} and y_{nc3}) represent the measured GFP fluorescence in the cell population for a constant concentration of Galactose. The desired (r in blue) and experimentally quantified GFP fluorescence ($y_{randominput}$ in light green) for the whole duration of the random input negative control experiments are also shown.* (Bottom panel) *the dark red line represents the constant concentration of Galactose (2%) provided to cells corresponding to the experiments y_{nc1}, y_{nc2} and y_{nc3}; the light red series of pulses, corresponding to the experiment $y_{random-input}$.*

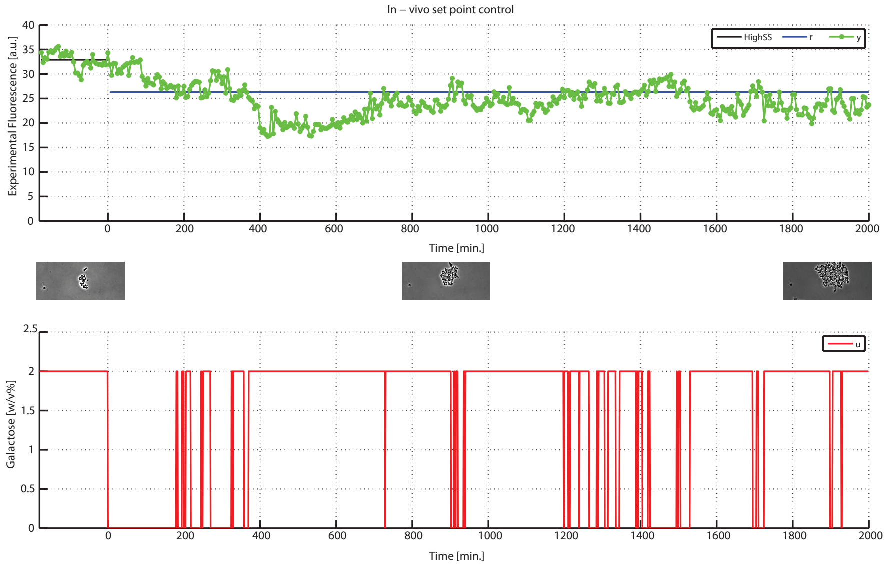


Figure 6.9: In-vivo set point control experiment on IRMA. Top panel: *the desired (r in blue) and experimentally quantified GFP (y in green) are shown for the whole duration of the experiment; the control action starts at time* $t = 0$ min *and lasts for* 2000 min. Bottom panel: *the input signal u computed by the control algorithm is shown in red.* (Insets) *Images taken during the experiment show the growing yeast population at the beginning, at the half and at the end of the experiment.*

that despite the cell-to-cell variability (see standard deviation bounds in Figure 6.10) the control action is really effective.

For comparison in Figure 6.12, I also reported an experiment without control input, showing that without active control, but only with a sustained ON (Galactose) input, protein expression fluctuates during the course of the experiment.

Discussion The control quality obtained by the control scheme is remarkably good in the case of the *GAL1* endogenous promoter, but it may seem unsatisfying in the case of the IRMA network when compared to classic control engineering approaches applied to engineering systems and devices. This is the first attempt to control gene expression in a complex network using feedback control in a noisy biological system. Indeed, the presence of cell-to-cell variability is one of the key

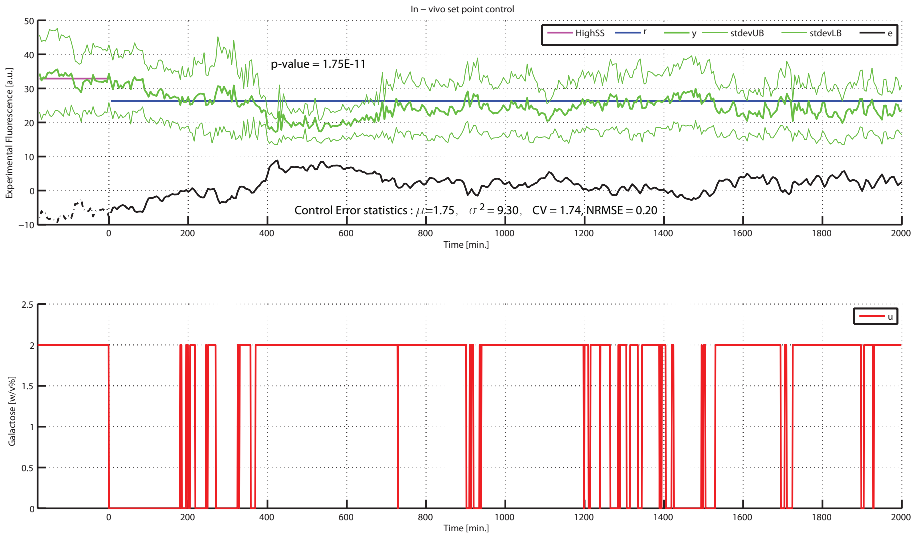


Figure 6.10: In - vivo set point control experiment for the IRMA network - fluorescence standard deviation. Top panel: *by using the off-line analysis described in the text it is possible to calculate the standard deviation of the fluorescence for each frame acquired during the control. The desired amount of protein (r in blue), the quantified GFP (y green line), the standard deviation's upper and lower bounds (thin green lines) and the control error e in black are shown; mean μ, variance σ and coefficient of variation CV of the control error are also shown; the p-value was computed as described in the text.* Bottom panel: *the input signal u computed by the control algorithm is shown in red.*

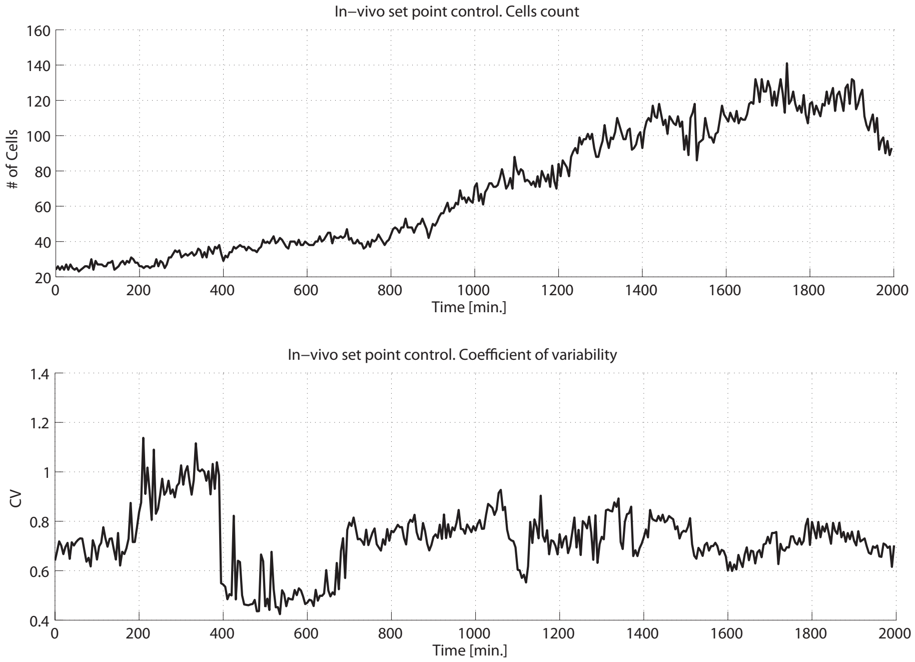


Figure 6.11: In - vivo signal tracking control experiment for the IRMA network - Cell count and coefficient of variation. *For the experiment of Figure 6.10, the number of cell (top panel) and the coefficient of variation (bottom panel) are shown.*

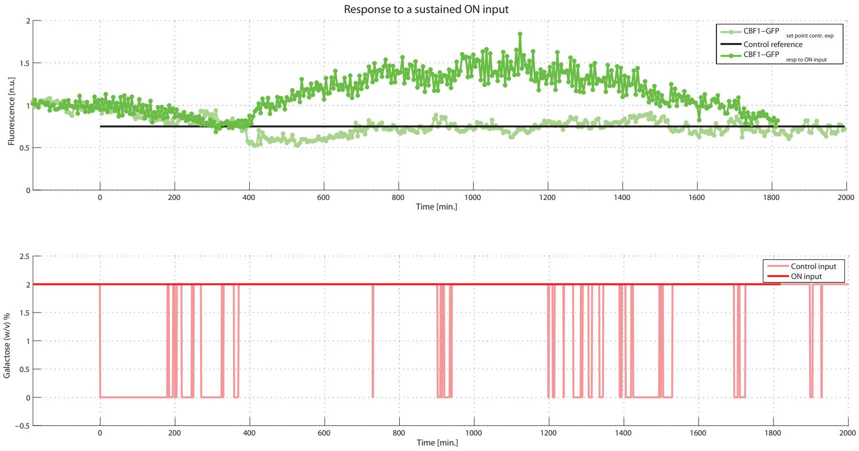


Figure 6.12: Response to a sustained Galactose input for the IRMA network. *Green line: fluorescence measured when the cells are treated with Galactose for the whole experiment; light green line: fluorescence measured during the in-vivo set point control experiment (Figure 6.9); black line: the control reference of the set-point control experiment (Figure 6.9); red line: the sustained Galactose input provided to the cells population; light red: the input calculated automatically by the control algorithm and used to regulate the production of GFP to the desired level in in-vivo set point control experiment (Figure 6.9).*

obstacles when implementing control strategies for living systems. This is why, as already mentioned, I aimed at controlling the average fluorescence level of the cell population, which is shown to converge towards the desired value. Moreover, the control scheme keeps biological noise from increasing and at a physiological level as estimated by the CV.

The microfluidics-based control strategy I developed enables control experiments using small volumes of reagents with minimal perturbations to the cells. It can be easily implemented with limited costs to fine tune the expression of a protein of interest from an endogenous promoter with minimal intervention (i.e. introduction of a fluorescent reporter gene).

6.2 *In-vivo* comparative analysis of feedback control strategies for gene expression regulation

The experimental results described up to know convincingly demonstrate that the expression of a protein can be controlled in vivo in real-time, using an inducer molecule acting directly or indirectly on protein expression, by applying principles drawn from classical control theory. In general the regulation of every gene product can be achieved as long as inducer molecules and fluorescent reporters are available. Anyway, as can be appreciated from Figures 6.6 although the control error is bounded, the controlled variable oscillates. This is mainly due to the kind of control input (switching signal) used to steer system response; the type of input is constrained, as already argumented, by how cells metabolise Galactose and Glucose, but other methods to calculate the duration of input pulses could be investigated. Moreover, in relation to the control objective, it is well known from Control Theory that even though PI regulators are well suited for steady-state (set point) control problems, they cannot guarantee good performances in signal tracking regulation (37); thus changing the reference signal could affect control outcome.

The natural question arising from the above considerations, is to understand whether it is possible to further improve control performances and if the control strategy adopted is really the most suitable. To address these question I have decided to use the transcription driven by the *GAL1* promoter as a testbed to compare and assess the performances of three different control strategies: a) an improved version of the PI regulator already used, b) the Model Predictive Control (MPC) feedback law (already applied to biological systems (5, 7)) and c) the Zero Average Dynamics control devised and used for the regulation of power converters (66, 67).

Prior of providing a detailed description of the control strategies, and of their implementations here compared, it is worth to recapitulate the constraints on the control input, and to provide details of all tools used to apply these regulators to the chosen testbed.

6.2.1 Control objective and implementation tools

Control objective I have applied the three regulation strategies either to set-point and signal-tracking control tasks:

1. **set-point control, reference signal:** the set-point is calculated as the 50% of the average of the fluorescence expressed by the cells during the calibration phase (§8). The reference has a duration of 1000 minutes
2. **signal-tracking control, reference signal:** the reference is a step-like signal in which each step has respectively a value equal to the 75% , the 50% and the 25% of the average of the fluorescence expressed by the cells during the calibration phase (§8). Each step has a duration of 500 minutes.

Comparison metrics To assess and compare control performances from the three control algorithms used, I have used metrics based on the analysis of the control error e. These metrics in general are adopted to optimise the tuning of of PI and PID regulator gains on the basis of the control outcome(37), so they provide a valid measure of control quality.

The *Integral Square Error* (ISE), defined as:

$$ISE = \int_0^t e(\tau)^2 d\tau \tag{6.5}$$

integrates the square of the error over the time. ISE penalises large errors more than smaller ones.
The *Integral Absolute Error* (IAE), given by:

$$IAE = \int_0^t |e(\tau)| d\tau \tag{6.6}$$

integrates the absolute error of the control over time; a weighted version of the IAE is the *Integral Time Absolute Error* (ITAE) calculated as:

$$ITAE = \int_0^t \tau |e(\tau)| d\tau \tag{6.7}$$

that integrates the absolute error multiplied by the time. It penalises more persisting errors than those at the start of the response.

Control input: During each sampling period T the control input u can assume only two values so that:

$$u(t) = \begin{cases} u_{MAX} = ON & kT \leq t < (k + d_k)T \\ u_{MIN} = OFF & (k + d_k)T \leq t < (k+1)T \end{cases} \tag{6.8}$$

referring to Figure 6.13, controllers have to calculate the duration of Galactose pulses (ON value), as a left-sided PWM, i.e. the ON pulse starts at the beginning of each period T. The length of the pulse t_{ON} is defined by the duty-cycle $d_k = \frac{t_{ON}}{T}$ with $d_k \in [0, 1]$.

Mathematical model: Both MPC and ZAD strategies rely on a mathematical model of the process being controlled to calculate input to exert the regulation. As it will be described, these two algorithms allow to directly calculate the duty cycle d_k of the input square wave at each sampling time. To speed up the computation process I decided to use a discretised version of the *model 2* inferred in §5.2, assuming that the input is piece-wise constant during the sampling period T (zero-order hold method described in (68)), thus obtaining:

$$x_{k+1} = Ax_k + Bu_k, \quad y_k = Cx_k. \quad x(0) = x_0 \tag{6.9}$$

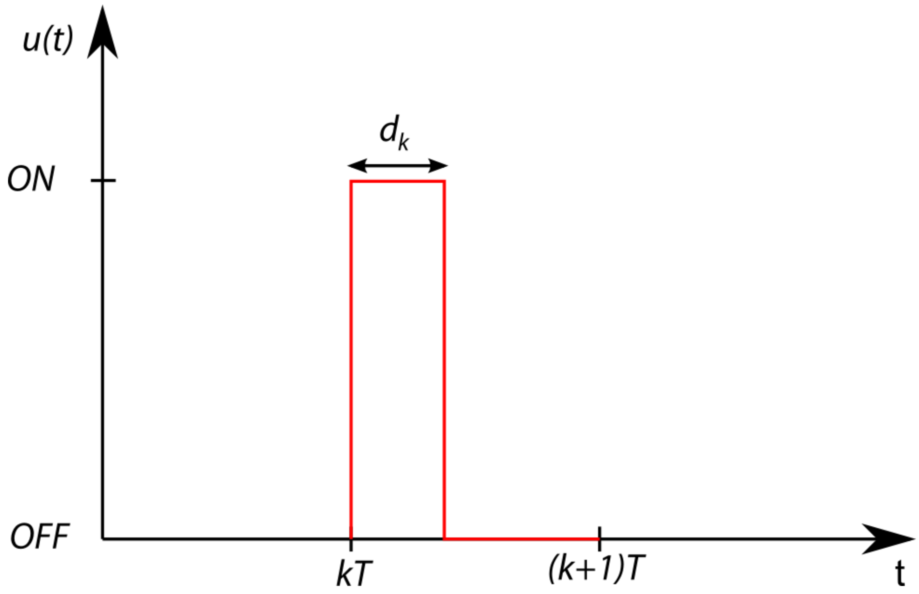


Figure 6.13: Input signal. *During each sampling time T the input signal is a pulse whose duration t_{ON} is defined by the duty-cycle $d_k = \frac{t_{ON}}{T}$ with $d_k \in [0, 1]$. Since the pulse starts at the beginning of the period this modulation is called left-sided PWM*

where $x_k = \binom{x_1(kT)}{x_2(kT)}$ is the system state, $u_k = u(kT)$ is the input and $y_k = y(kT)$ is the output, with k being a natural number ($k \in [1, 2, \ldots]$)

where the matrix A is:

$$A = \begin{pmatrix} a_{1,1} & 0 \\ a_{2,1} & a_{2,2} \end{pmatrix} \tag{6.10}$$

B:

$$B = \begin{pmatrix} b_1 \\ 0 \end{pmatrix} \tag{6.11}$$

and C:

$$C = \begin{pmatrix} 0 & c_2 \end{pmatrix} \tag{6.12}$$

thus the model has preserved its structure, with negative elements on the main diagonal, $a_{1,1}$ and $a_{2,2}$, greater than -1 (system asymptotically stable (37)). The control input, as in the continuous time case, affects only the first system state via the coefficient b_1, and it can assume the values $u_{MAX} = 2$ and $u_{MIN} = 0$.

State estimator During *in-vivo* experiments it is only possible to measure directly system output (y_k in the model) and not the states x_{k_i}. These are needed by MPC and ZAD algorithms to calculate the control input. Thus to estimate at each sampling time their values from measured data, I have implemented a classical Kalman filter as described in (43). The Kalman filter algorithms operates recursively on streams of noisy measured data to produce a statistically optimal estimate of the underlying system state.

6.2.2 PI regulator

The PI control strategy was re-designed to dampen output oscillations around the set-point. To this end, I tuned new parameters for the regulator and I devised a new modulation procedure to convert the analog signal generated by the PI to the digital accepted by the system.

I re-calculated proportional and integral gains, K_p and K_i, with the Ziegler-Nichols' method (37) already adopted, the difference in this case is that I used the results of the step response evaluated for *model 2*, inferred in §5.2. Thus I have set $K_p = 13.49$ and $K_i = 0.17$.

I removed the PWM block downstream of the PI, thus the duty cycle of the input d_k is calculated as:

$$d_k = \frac{\hat{u}}{u_{MAX}}. \tag{6.13}$$

where $\hat{u}$ is the output of the PI regulator saturated between $u_{MIN} = 0$ and $u_{MAX} = 2$. Even in this case I have used the anti-wind up scheme of Figure 2.7, with $K_t = 1$.

6.2.3 Model Predictive Control

This feedback control strategy, at each sampling time kT, uses a mathematical model of the process being controlled, to calculate the control input; the algorithm simulates, over a defined prediction horizon and starting from the actual system state at kT, the open loop response of the dynamical model to several inputs and, chooses among them the one that minimises a cost function measuring the distance between the model output and the control objective (42). In absence of external disturbances and other sources of uncertainty, the optimal input found could in principle be applied to the physical process over the entire prediction horizon. However in order to make control action robust, the feedback loop is closed by applying the calculated input only up to the next sampling time $(k+1)\,T$ when the entire procedure is repeated.

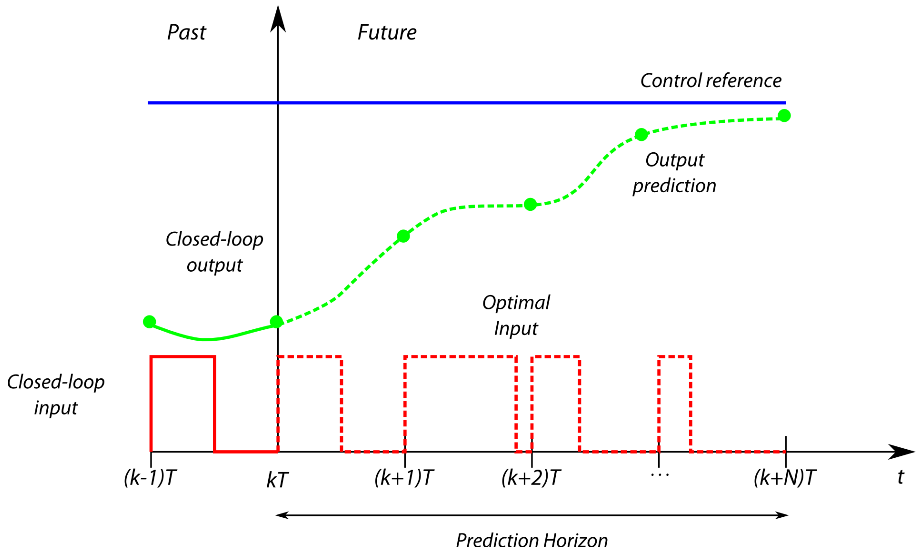


Figure 6.14: Model Predictive Control principle. *On the basis of the actual system output (i.e. estimated state), at each sampling interval, the controller simulates, over a defined prediction horizon, the open loop response of the mathematical model to several input signals. The regulator feeds the physical system with the first pulse of the optimal input, i.e the one that in simulation minimises an objective function measuring the distance of the model output from the control reference. This procedure is iteratively repeated at each sampling time.*

Among the several cost functions that can be used for the calculus of the optimal input (42), I have chosen the sum of the squared error (SSE), defined as:

$$SSE \triangleq \sum_{i=k+1}^{k+N} \big(N+1+k-i\big)\,\epsilon_i^2 \quad = \sum_{i=k+1}^{k+N} \big(N+1+k-i\big)\,\big(y_i\,-\,r_i\big)^2 \quad (6.14)$$

where the integer N defines the length of the prediction horizon in terms of sampling intervals (set as $N = 12$). I have added the weighting factor $(N + 1 + k - i)$ to weight more the errors at the beginning of the prediction horizon than those at the end; this can guarantee more prompt corrections of output deviations from the reference. The optimisation is carried out by adopting the Matlab implementation of the Genetic Algorithm described in (69). The result of the optimisation is an array of N optimal duty cycles $d_{k_i}, i \in [1, N]$; as previously explained, only the first element of this array is used to decide with which sugar (Galactose or Glucose, ON or OFF) and for how long, during the sampling period T, cells have to be fed.

6.2.4 Zero Average Dynamics Control

The ZAD control strategy allows to directly calculate the duty cycle d_k of a switching signal (66). It relies on a modification of the Sliding Mode Control, where the control objective consists in attracting system states to slide over a fixed sliding surface, i.e the reference, defined as $s(x) = 0$(70). In the ZAD control the sliding condition has to be fulfilled only on average over each sampling period, so that:

$$\mathbb{E}_T\big[s\big(x(t)\big)\big] \;=\; \frac{1}{T}\int_{kT}^{(k+1)\,T} s\big(x(t)\big)\,dt \;=\; 0 \qquad (6.15)$$

I considered the following sliding surface, namely the reference, to control *GAL1* promoter dynamics:

$$s\big(x(t)\big) \;=\; \big(x_2(t) - x_{2_{ref}}(t)\big) + \big(\dot{x}_2(t) - \dot{x}_{2_{ref}}(t)\big) \qquad (6.16)$$

where x_2 is the state variable describing the dynamics of the fluorescent reporter; note that the second term of $s\big(x(t)\big)$ is equal to 0 in the case of set-point regulation.

The solution of (6.15) can be computationally expensive and very slow, thus to overcome this issue I considered the piecewise-linear approximation of the sliding surface $s\big(x(t)\big)$ proposed in (67), that in the case of left-sided PWM control inputs (Figure 6.13) becomes:

$$s\big(x(t)\big) \; = \begin{cases} s_k \; + \; (t - kT) \; \dot{s}_k^{\,\mathrm{on}} & kT \; \leq \; t \; < \; (k + d_k)\,T \\ s_k \; + \; d_k \; T \; \dot{s}_k^{\,\mathrm{on}} \; + \; (t - (k + d_k)\,T) \; \dot{s}_k^{\,\mathrm{off}} & (k + d_k)\,T \; \leq \; t \; < \; (k+1)\,T \end{cases} \tag{6.17}$$

where s_k, $\dot{s}_k^{\,\mathrm{on}}$, and $\dot{s}_k^{\,\mathrm{off}}$ are:

$$\begin{aligned} s_k \; &= \; s(x_k) \\ \dot{s}_k^{\,\mathrm{on}} \; &= \; \dot{s}(x_k)\Big|_{u\,=\,2} \\ \dot{s}_k^{\,\mathrm{off}} \; &= \; \dot{s}(x_k)\Big|_{u\,=\,0} \end{aligned} \tag{6.18}$$

Substituting the piecewise-linear approximation (6.17) into (6.15):

$$\begin{aligned} \mathbb{E}_T\big[s\big(x(t)\big)\big] \; =& \frac{1}{T}\int_{kT}^{(k+d_k)\,T} \big[s_k \; + \; (t - kT) \; \dot{s}_k^{\,\mathrm{on}}\big] \, dt \\ &+ \frac{1}{T}\int_{(k+d_k)\,T}^{(k+1)\,T} \big[s_k \; + \; d_k \; T \; \dot{s}_k^{\,\mathrm{on}} \; + \; (t - (k + d_k)\,T) \; \dot{s}_k^{\,\mathrm{off}}\big] \, dt \end{aligned} \tag{6.19}$$

solving the integral (6.19):

$$\mathbb{E}_T\big[s\big(x(t)\big)\big] \; = \; 0 \; \Longrightarrow \frac{1}{2}\, d_k^2 \, T\,(\dot{s}_k^{\,\mathrm{off}} - \dot{s}_k^{\,\mathrm{on}}) - d_k \, T\,(\dot{s}_k^{\,\mathrm{off}} - \dot{s}_k^{\,\mathrm{on}}) + s_k + \frac{1}{2}\, T\, \dot{s}_k^{\,\mathrm{off}} \; = \; 0 \tag{6.20}$$

The duty cycle d_k can be calculated by solving the second order equation 6.20, thus finding:

$$d_k \; = \; \frac{-T\,(\dot{s}_k^{\,\mathrm{on}} - \dot{s}_k^{\,\mathrm{off}}) \; \pm \; \sqrt{T\,(\dot{s}_k^{\,\mathrm{on}} - \dot{s}_k^{\,\mathrm{off}})\,(2\, s_k + T\, \dot{s}_k^{\,\mathrm{on}})}}{-T\,(\dot{s}_k^{\,\mathrm{on}} - \dot{s}_k^{\,\mathrm{off}})} \tag{6.21}$$

Moreover, considering that:

$$\dot{s}_k^{\,\mathrm{off}} - \dot{s}_k^{\,\mathrm{on}} \; = \; -2\, b_1 \, a_{2,1} \; < \; 0 \quad \Longrightarrow \quad \dot{s}_k^{\,\mathrm{on}} - \dot{s}_k^{\,\mathrm{off}} \; > \; 0 \tag{6.22}$$

the solutions of (6.15) are:

$$d_k = 1 \mp \sqrt{\frac{2\, s_k + T\, \dot{s}_k^{\text{on}}}{T\, (\dot{s}_k^{\text{on}} - \dot{s}_k^{\text{off}})}} \tag{6.23}$$

Since the duty cycle assumes values only in $\big[0, 1\big]$, the only admissible solution is:

$$d_k = 1 - \sqrt{\frac{2\, s_k + T\, \dot{s}_k^{\text{on}}}{T\, (\dot{s}_k^{\text{on}} - \dot{s}_k^{\text{off}})}} \tag{6.24}$$

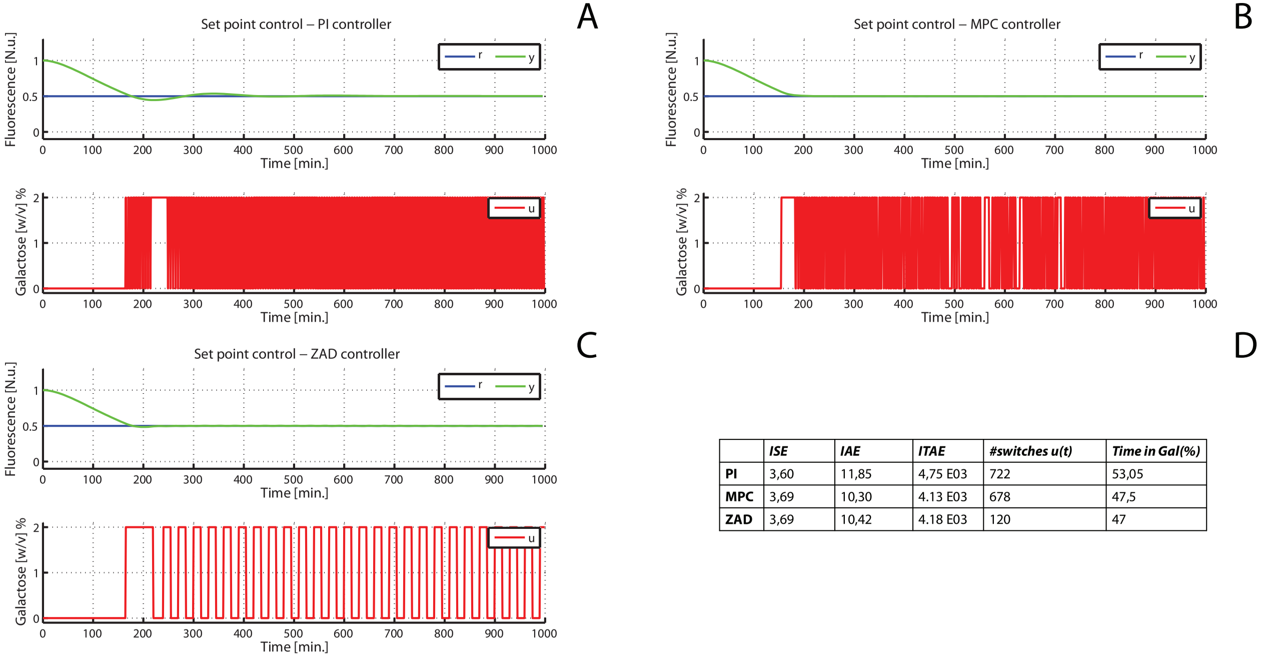
Furthermore, to avoid saturation, it has to be:

$$0 \leq \frac{2\, s_k + T\, \dot{s}_k^{\text{on}}}{T\, \big(\dot{s}_k^{\text{on}} - \dot{s}_k^{\text{off}}\big)} \leq 1 \tag{6.25}$$

6.2.5 *In-silico* validation

I tested *in-silico* the control strategies described above, by using as a proxy for the system behaviour *model 2* inferred in §5.2. The controllers were simulated to perform both set-point and signal-tracking regulation of the model output.

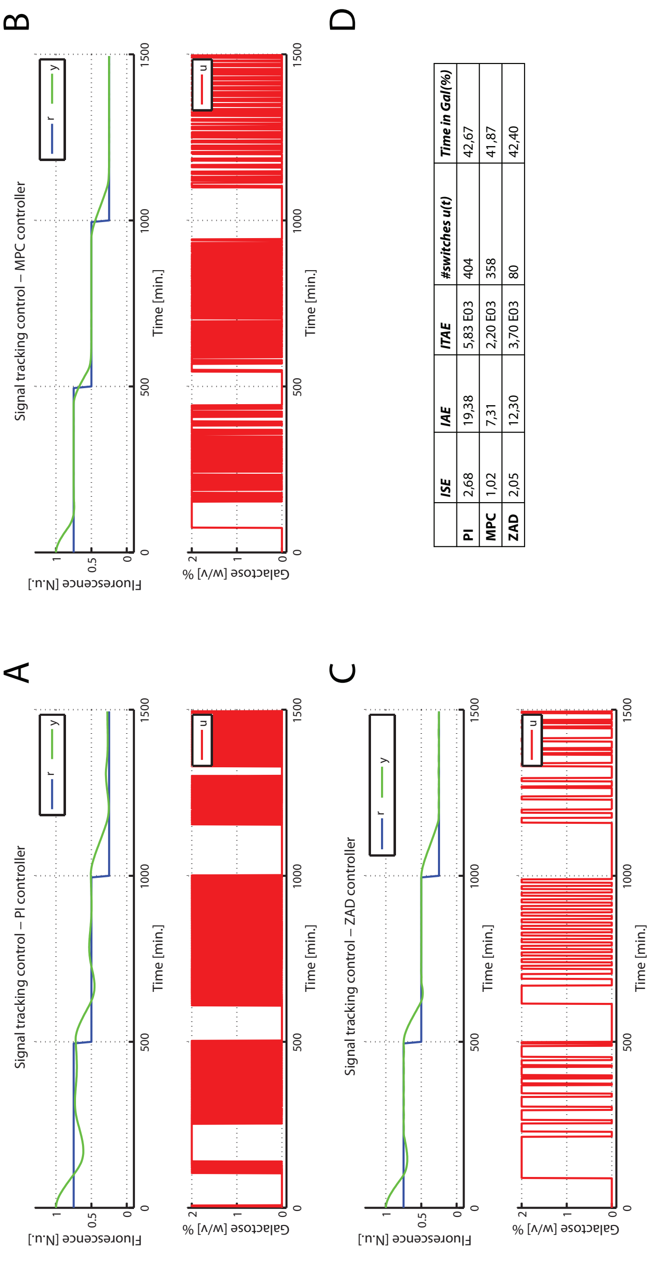
The improved PI regulator, when applied to the set-point control (Figure 6.15 A) is able to reach the control reference and to maintain its value without appreciable oscillations at the steady-state, thus improving the performances of the original PI (Figure 6.4). Comparing this result with ones achieved by MPC and ZAD regulators (Figure 6.15 B-C), the performances indices calculated (ISE, IAE, ITAE see Figure 6.15 D) are of the same order of magnitude for all the control strategies; interestingly the ZAD controller is able to achieve satisfying results with a reduced number of input switches (five and six fold less than respectively MPC and PI). This, in a practical control implementation, could result in a reduced energy for the control (if a cost is applicable to each input switch), in other words the ZAD strategy is the cheapest, at least in theory, among the tested feedback control laws.



	ISE	IAE	ITAE	#switches u(t)	Time in Gal(%)
PI	3,60	11,85	4,75 E03	722	53,05
MPC	3,69	10,30	4.13 E03	678	47,5
ZAD	3,69	10,42	4.18 E03	120	47

Figure 6.15: ***GAL1* promoter *in-silico* set point control** (A-C) Three *in-silico* set point control experiments performed on the *GAL1* promoter mathematical model by the mean of the PI (A), of the MPC (B) and of the ZAD (C) regulators. The control action starts at time $t = 0$ min. and finishes at $t = 1000$ min. The desired level of the output (r in blue) is calculated as a percentage (50%) of the high steady state of the model's output here normalised to 1. The model's output during the control (y in green) and the control input (u in red) are shown for the whole duration of the experiment. (D) Performance indices calculated over the control time interval: Integral Square Error (ISE), Integral Absolute Error (IAE), Integral Time Absolute Error (ITAE), number of switches of the control input, and the percentage of time during which the model is fed with the 'ON' input.

In the case of *in-silico* signal tracking control, as expected, the performance of the PI is the worst (Figure 6.16 A); this is due to the fact that PI regulators are meant to account for steady-state regulations and not to track time varying signals. The intrinsic predictive structure of the MPC allows this strategy to achieve good performances in particular in the proximity of reference discontinuities; the controller predicts these changes in the reference signal and, adjusts accordingly the control input starting to "switch off" the system before than the ZAD and the PI do (Figure 6.16). Once again the control implemented with the ZAD technique achieves satisfying results with a lower number of input switches.



	ISE	IAE	ITAE	#switches u(t)	Time in Gal(%)
PI	2,68	19,38	5,83 E03	404	42,67
MPC	1,02	7,31	2,20 E03	358	41,87
ZAD	2,05	12,30	3,70 E03	80	42,40

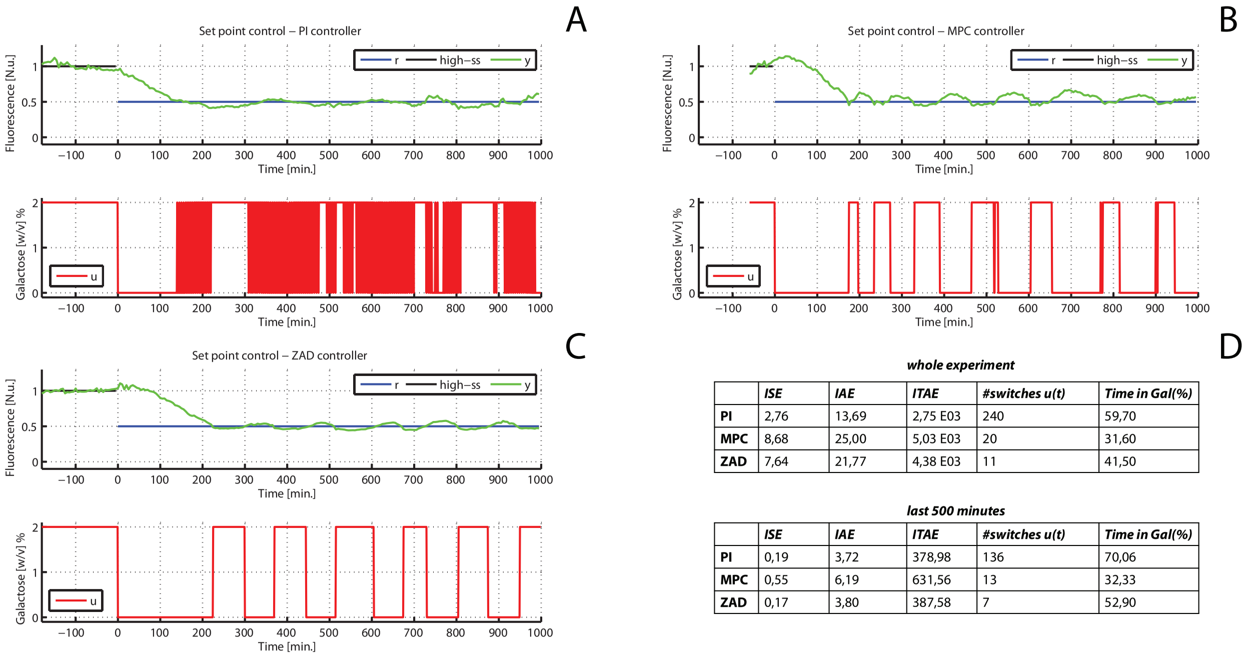
Figure 6.16: *GAL1* promoter *in-silico* signal tracking control (A-C) Three *in-silico* signal tracking control experiments performed on the *GAL1* promoter mathematical model by the mean of the PI (A), of the MPC (B) and of the ZAD (C) regulators. The control action starts at time $t = 0$ min. and finishes at $t = 1500$ min. The desired level of the output (r in blue) is a three-steps signal where, the value of each step (lasting for 500 min.) is calculated as a percentage (75%, 50% and 25%) of the high steady state of the model's output here normalised to 1. The model's output during the control (y in green) and the control input (u in red) are shown for the whole duration of the experiment. (D) Performance indices calculated over the control time interval: Integral Square Error (ISE), Integral Absolute Error (IAE), Integral Time Absolute Error (ITAE), number of switches of the control input, and the percentage of time during which the model is fed with the 'ON' input.

6.2.6 *In-vivo* validation

The results obtained *in-silico* confirmed that the selected control strategies, at least in simulation, were suitable to regulate the system under investigation.

When applied *in-vivo* to perform set-point control (Figure 6.17), they are able to accomplish the control objective as predicted by the numerical simulations.

In particular, the PI controller (Figure 6.17 A) regulates cell fluorescence with less oscillations than its previous implementation (Figure 6.6); moreover, unlike MPC and ZAD regulators, the PI does not use a mathematical description of the process being controlled to calculate the control input; this results in a higher robustness than the other controllers as confirmed by the performances indices calculated for this set of experiments (Figure 6.17 D). The ZAD regulator, as suggested by *in-silico* results, achieves the control objective with very few input switches (Figure 6.17 C); furthermore it gets fluorescence oscillations with a reduced amplitude of those obtained by the MPC feedback strategy (Figure 6.17 B)



whole experiment

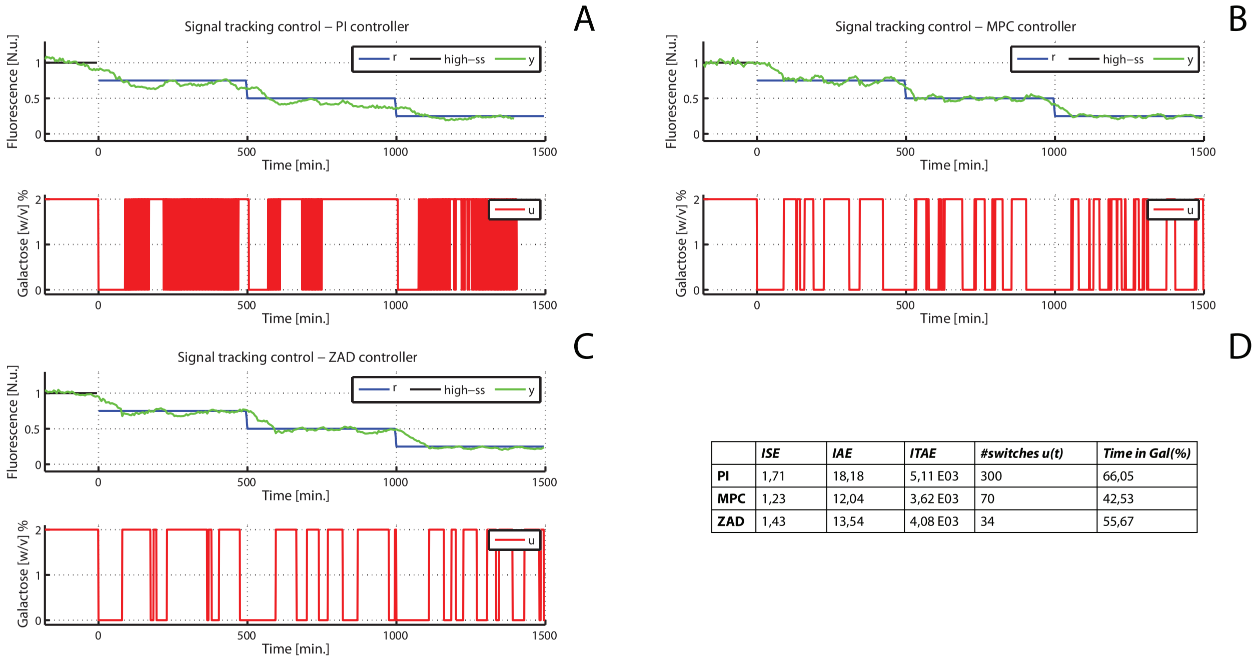
	ISE	*IAE*	*ITAE*	*#switches u(t)*	*Time in Gal(%)*
PI	2,76	13,69	2,75 E03	240	59,70
MPC	8,68	25,00	5,03 E03	20	31,60
ZAD	7,64	21,77	4,38 E03	11	41,50

last 500 minutes

	ISE	*IAE*	*ITAE*	*#switches u(t)*	*Time in Gal(%)*
PI	0,19	3,72	378,98	136	70,06
MPC	0,55	6,19	631,56	13	32,33
ZAD	0,17	3,80	387,58	7	52,90

Figure 6.17: ***GAL1*** **promoter** ***in-vivo*** **set point control** (A-C) Three *in-vivo* set point control experiments performed on the *GAL1* promoter by the mean of the PI (A), of the MPC (B) and of the ZAD (C) regulators. The control action starts at time $t = 0$ min. and finishes at $t = 1000$ min. The value of the desired level of fluorescence (r in blue) is calculated as a percentage (50%) of the GFP fluorescence quantified before time $t = 0$ min. during the period of 180 min in which cells are kept in galactose enriched medium to let them adapt to the microfluidic environment. The measured GFP fluorescence during the control (y in green) and the control input (u in red) are shown for the whole duration of the control. (D) Performance indices calculated over the whole control time interval (upper table) and for the last 500 min. of regulation: Integral Square Error (ISE), Integral Absolute Error (IAE), Integral Time Absolute Error (ITAE), number of switches of the control input, and the percentage of time during which cells are fed with galactose enriched medium starting from time $t = 0$ min.

Signal-tracking performed *in-vivo* (Figure 6.18) confirms the results achieved in simulation. The PI controller (Figure 6.18 A) despite the high number of input switches, poorly tracks the reference signal. The intrinsic forecasting structure of the MPC, as already demonstrated *in-silico*, allows this regulator to obtain best performances as confirmed by performance indices (Figure 6.18 B and D). Even in this case the ZAD controller achieves the control objective with less input switches than the PI and the MPC.



	ISE	*IAE*	*ITAE*	*#switches u(t)*	*Time in Gal(%)*
PI	1,71	18,18	5,11 E03	300	66,05
MPC	1,23	12,04	3,62 E03	70	42,53
ZAD	1,43	13,54	4,08 E03	34	55,67

Figure 6.18: ***GAL1* promoter *in-vivo* signal tracking control.** (A-C) Three *in-vivo* signal tracking control experiments performed on the *GAL1* promoter by the mean of the PI (A), of the MPC (B) and of the ZAD (C) regulators. The control action starts at time $t = 0$ min. and finishes at $t = 1500$ min. The desired level of fluorescence (r in blue) is a three-steps signal where, the value of each step (lasting for 500 min.) is calculated as a percentage (75%, 50% and 25%) of the GFP fluorescence quantified before time $t = 0$ min. during the period of 180 min in which cells are kept in galactose enriched medium to let them adapt to the microfluidic environment. The measured GFP fluorescence during the control (y in green) and the control input (u in red) are shown for the whole duration of the control. (D) Performance indices calculated over the control time interval: Integral Square Error (ISE), Integral Absolute Error (IAE), Integral Time Absolute Error (ITAE), number of switches of the control input, and the percentage of time during which cells are fed with galactose enriched medium starting from time $t = 0$ min.

Summarising the results achieved, simulations as well as *in-vivo* experiments, confirm that MPC and ZAD strategies can achieve successfully the regulation of gene expression in living cells for both set-point and tracking control, as long as an accurate dynamical model is able to predict process dynamics. The PI control, as expected from the theory (37), has a worse performance in the case of signal-tracking regulation whereas, for the set-point control, is more robust than the other two regulators. Moreover the ZAD allows to accomplish both the control tasks with a lower actuation effort.

7

In vivo feedback control of inducible promoters in mammalian cells

The results achieved controlling gene expression in a population of *S. Cerevisiae*, confirm that it is possible to steer the expression of a target gene , in real time, with an accuracy strongly related to the control strategy adopted. The use of microfluidics device allows to precisely administer inducer molecule to cells while all living conditions are guaranteed to them. The experimental platform developed is highly modular and can be used to manipulate other cellular models apart from yeasts. In our Laboratory we borrowed the design of a microfluidic device for mammalian cells developed in the Biodynamics Laboratory at the University of San Diego (CA) and described in (71). This device has been designed to load cells in a microfluidic environment and to feed them with two different compounds with the actuation strategy described in §4 and §8. We integrated this device in our experimental platform to set up a negative feedback control strategy for gene expression in mammalian cells. We chose as a testbed for the control the NOPFL circuit already described in §5.4.

7.1 Back to the Mathematical model

The NOPFL circuit was presented in (25) and its mathematical model here discussed in §5.4. The model can be rewritten in a more compact form:

$$\frac{dx_1}{dt} = v_1\left(\alpha_1 + \left(1 - \alpha_1\right)0.6\right) + v_1\left(\left(1 - \alpha_1\right)0.4\right)D - d_1 x_1 \tag{7.1}$$

$$\frac{dx_2}{dt} = v_2 x_1 - \left(d_3 + K_f\right)x_2 \tag{7.2}$$

$$\frac{dx_3}{dt} = K_f x_2 - d_3 x_3 \tag{7.3}$$

where x_1 is the production of the d2EYFP mRNA, x_2 is the unfolded reporter protein, and x_3 is the mature, fluorescent d2EYFP (the output of the system). D accounts for the presence, or absence, of Doxycyline (or Tetracycline) in the growing medium, and it can be either 0 (switch-off signal) or 1 (switch-on) In Table 7.1 all the values for the parameters that were fitted in (25) are reported.

Parameter	Fitted value
$\alpha_1[nMmin^{-1}]$	$1.13E-05$
$v_1[min^{-1}]$	$7.54E-02$
$v_2[min^{-1}]$	$2.71E-02$
$d_1[min^{-1}]$	$1.01E-02$
$d_3[min^{-1}]$	$3.24E-03$
$K_f[min^{-1}]$	$1.24E-03$

Table 7.1: NOPFL model parameter values.

From a control perspective the system is linear and the input, as in the case of the *GAL1* promoter, can assume only two values. As first step towards the *in-vivo* control of this biological system, we decided to apply the simple On-off control strategy described in §2.3 to accomplish a set-point regulation. We tested this strategy *in-silico* to analyse the results achieved by this controller in steering the output of the mathematical model above discussed.

7.2 *In-silico* set point control

The control objective consisted in the regulation of the model output at the 50% of the value of its high steady-state over 5000 minutes. To account for measurement noise, a gaussian noise with zero mean and standard deviation equal to 1 is added to the model output prior of being fed back to the controller. As expected, the control strategy adopted allows to reach the set point that is maintained by the output with an oscillating behaviour (Figure 7.1). When no hysteresis is added to the regulator (Figure 7.1 A), the control input switches more than when applying an hysteresis equal to the 5% of the set-point (Figure 7.1 B); the drawback in the latter case is that the amplitude of the oscillations around the set-point increases.

While I am writing, my colleagues are testing this control strategy *in-vivo* to verify whether the outcome of the numerical simulations can be confirmed controlling living cells. The next step will be to test *in-silico* and implement *in-vivo* the controllers I have devised for the *GAL1* promoter.

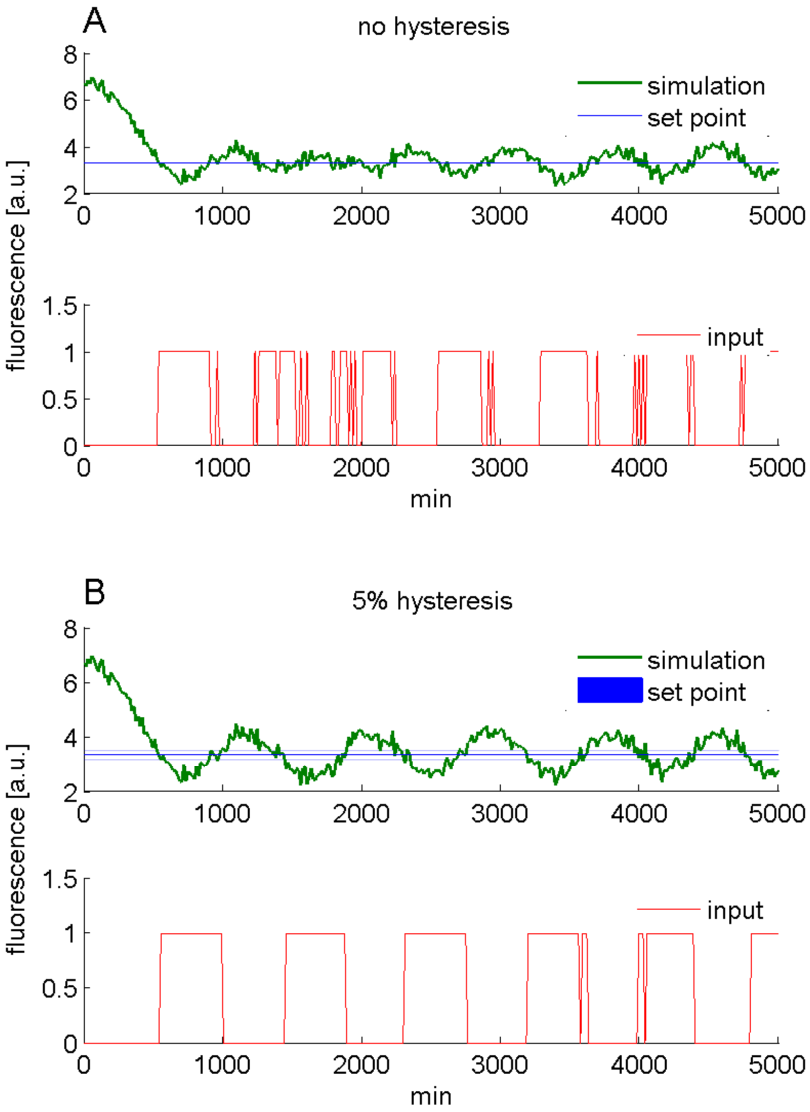


Figure 7.1: NOPFL *in-silico* signal tracking control. Green line is the simulated model output, the blu signal is the control objective and the red line is the control input computed by the controller. (A) *Control simulation carried out without any hysteresis applied to the regulator.* (B) *Control simulation carried out with an hysteresis equal to the* 5% *of the set-point applied to the regulator.*

8

Materials and Methods

8.1 Microfluidics: from fabrication to living cells handling

Experimental results presented in this work have been achieved by the means of the microfluidic device MFD0005a designed in the Jeff Hasty's Biodynamics Laboratory at the University Of California, San Diego (44). This device was instrumental in carrying out the proposed experiments since allowed cells to grow in a monolayer inside a dedicated chamber (trap), to refresh the growing medium and to administer precisely the concentrations of the inducer compounds provided to yeasts growing in the chamber via a complex topology of channels connecting 5 inlets between each other and the cells' trap (a detailed description of its topology is provided in §4.1.1 and (44)).

8.1.1 Microfluidic devices fabrication process

I have used replica molding technique to obtain replicas of the device presented in (44) thanks to the master-mold Prof. Jeff Hasty kindly provided us as a blueprint.

Before the fabrication of the microfluidic devices the master is exposed to chlorotrimethylsilane (Sigma-Aldrich Co.) vapours for 10 min so as to create an anti-sticking silane layer for PDMS. A 10 : 1 mixture of PDMS prepolymer and curing agent (Sylgard 184, Dow Corning) is prepared and degassed under vacuum for 1 hour. Then the mixture is poured on the patterned, and to facilitate the polymerization and the cross-linking, it is cured in a standard oven at 80°C for 2h. After this step the PDMS layer, containing the microfluidic channels, is peeled from the master and it is cut with a scalpel to separate the single devices; holes are bored through them with a 20-gauge blunt needle in order to create fluidic ports for the access of cells and liquid substances. The PDMS layers obtained are rinsed in isopropyl alcohol in a sonic bath for 10 min to remove debris. For each PDMS piece containing microchannels a thin glass slide (150um) is cleaned in acetone and isopropyl alcohol in a sonic bath for 10 min for each step. Finally the PDMS layers and glass slides are exposed to oxygen plasma in Plasma Cleaner machine (ZEPTO version B, Diener electronic GmbH) for 2 minutes and brought into contact forms a strong irreversible bond between two surfaces. As last step all devices were checked for faults inside and outside the channels.

8.2 Actuators, design and sizing

Actuation aim is to establish a difference in the hydrostatic pressure at the two ports (1 and 2) of the microfluidic device in order to appropriately modulate, according to the desired goal, the inputs concentration in the fluid reaching the cell trap. To accomplish this task, I designed and built two vertically mounted linear actuators; using this system it is possible to change heights of liquid filled syringes that feed into the DAW junction. The actuation system comprises two linear actuators both designed to move independently; the motion is achieved through a stepper motor while the transmission by using a timing belt and two pulleys (for each of the rail).

To ensure an effective regulation of the difference in the hydrostatic pressure at the inlet ports of the MFD0005a device, I designed the actuators sizing accurately the transmission system (pulleys and belts) and the stepper motor used.

Transmission system Considering typical 60ml syringes dimensions the distance between the centres of the two pulleys must be at least 600 mm. Physically this length is a function of the timing pulleys and of the particular timing belt adopted, from the SDP/SI on-line catalogue (http://www.sdp-si.com) I chose the following pulley and belt whose details are reported in Tables 8.1 and 8.2.

Material	Neoprene
Pitch	$3mm$
Teeth number	415
Width	$9mm$
Total length	$1245mm$

Table 8.1: Belt specifications

Material	Aluminium
Pitch	$3mm$
Teeth number	12
Toothed diameter	$11.50mm$
Total diameter	$14.70mm$
Total length	$17.50mm$

Table 8.2: Pulley specifications

Analysing data from Tables 8.1 and 8.2, considering belt total length, and pulley toothed circumference (number of teeth multiplying pitch dimension), the resulting center distance is of 604.5 mm, this result is consistent with the design constraints; thereby, each rail length must be long at least as the center distance obtained. Syringes are attached to the belt, using a plastic belt clamp.

Stepper motor The choice of motor is bound to the static and dynamic behavior that it must assume. Thus it is important to define the load, fixed to the belt, and its acceleration profile. Approximately the maximum load is about 0.2Kg (filled syringe or glass beaker), and the rising time (equal to the falling time) is 10 seconds (negligible if compared to the time interval needed to acquire images). Furthermore, to size the motor, it is necessary to transfer all the loads (pulleys, load, etc.) to the motor shaft. Given the values in Table 8.2, the calculus of pulleys' weight and inertia and load inertia is reported in formulas 8.1, 8.2 and 8.3

$$M_P = \left[\pi \frac{14.7 \cdot 10^{-3}}{2} \cdot 17.5 \cdot 10^{-3}\right] m^3 \cdot 2700 \frac{Kg}{m^3} = 8.02 \cdot 10^{-3} Kg \tag{8.1}$$

Rated voltage	$12Vdc$
Phase current	$0.6A$
Holding torque	$50Ncm$
Detent torque	$3.5Ncm$
Rotor Inertia	$120gcm^2$
Shaft diameter	$6.35mm$
Shaft length	$19mm$
Step angle	1.8
Step accuracy	0.09

Table 8.3: Motor specifications

$$J_P = 2 \cdot \left[\frac{1}{2} M_p \cdot r_P^2\right] = 8. - 2 \cdot 10^{-3} Kg \cdot \left(7.35 \cdot 10^{-3}\right)^2 = 4.33 \cdot 10^{-7} kg \cdot m^2 \quad (8.2)$$

$$J_L = M_L \cdot r_P^2 = 0.2Kg \cdot \left(5.75 \cdot 10^{-3} m\right)^2 = 6.62 \cdot 10^{-6} Kg \cdot m^2 \quad (8.3)$$

Thus, given load mass M_L, the load weight force P_L is calculated with formula 8.4; thus it is possible to calculate the total torque for the load T_L , by using formula 8.5. T_L must be less or equal than the torque that the motor is able to produce when it is not powered (detent torque), this to be sure that even if the motor is not powered it is able to hold the load in its actual position.

$$P_L = \left(0.2Kg \cdot 9.81 \frac{m}{s^2}\right) = 2N \quad (8.4)$$

$$T_L = P_L \cdot r_P = 0.0115Nm = 1.15Ncm \quad (8.5)$$

Considering all these requirements and constraints I chose the stepper motor (Pc Control Ltd) whose specifications are reported in Table 8.3.

The holding torque, in Table 8.3, is the torque that the motor is able to produce when it is powered. This torque must be greater or at least equal than the torque required to move the load with the desired speed and acceleration. Considering an activation time of 10sec (time span needed by the load to complete a full excursion along pulleys' distance), an acceleration time and a deceleration time both equal to 3sec (trapezoidal velocity profile (46)), I calculated the maximum velocity v_M and the maximum acceleration a_M with the formulas 8.6 and 8.7.

$$v_M = 0.086\frac{m}{s} \tag{8.6}$$

$$a_M = \frac{v_M}{3} = 0.029\frac{m}{s^2} = \frac{d\omega_p}{dt} = 5\frac{rad}{s^2} \tag{8.7}$$

The total system (motor rotor + pulleys + load) inertia J_T is calculated via formula 8.8. Thereby the total torque needed to move the load with the desired velocity profile is T_T, expressed in formula 8.9, is less than the motor holding torque.

$$J_T = J_R + J_P + J_L = 1.906 \cdot 10^{-5} kgm^2 \tag{8.8}$$

$$T_T = J_T \cdot \frac{d\omega_p}{dt} + T_L = 0.116Nm = 11.6Ncm \tag{8.9}$$

In order to drive the motors with the appropriate sequence of pulses, a commercial electronic board StepperBee+ (Pc Control Ltd) has been used. This driver, with an appropriate dynamic link library (DLL), gives the possibility to control both the motors through the USB pc port; to perform this task the routines, written in C++ programming language, have been included in the control algorithm written in Matlab environment.

8.3 Fluorescence Microscopy

The closed-loop control platform described in §4, employs an inverted fluorescence Nikon-TI Eclipse microscope to acquire images from cells trapped in the microfluidic device. To overcome the problem of the focus drift (due to cells growing and replicating and to the length of the experiments performed), the microscope has been equipped with the Nikon Perfect Focus System (PFS) that is able to compensate for axial focus fluctuations in real time during long-term imaging experiments.

I programmed the microscope to acquire two types of images: *(a)* a phase contrast image (PhC) and *(b)* two fluorescence images (one for the fluorescent reporter used to track cell state and, one in the red spectrum for Sulforhodamine B). The red dye Sulforhodamine B is added to the galactose medium and it is used to check (off - line) for the proper administration of the control input. Image acquisition is carried out with NIS Elements v. 3.22 software controlling an Andor iXon Ultra897 EMCCD camera. Both PhC and fluorescence images are acquired with the same objective (Nikon $40X$ dry objective, NA 0.63) at intervals of 5 min. An automated shutter is used to finely control the exposure times for each type of image acquired: *(a)* phase contrast exposure time of 286 ms, *(b)* green spectrum, with a Nikon FITC Filter (Ex $465-495$ nm, Em $515-555$ nm), exposure time of 900 ms, *(c)* red spectrum, with a Nikon TRITC HYQ filter (Ex $530-560$ nm, Em $590-650$ nm), exposure time of 100 ms. The exposure times and the acquisition interval of 5 min have been chose to avoid phototoxicity damages to cells and photobleaching of the fluorescent proteins/dyes (44).

8.4 Image Analysis

Image Analysis, together with the microscope, composes the sensing apparatus of the experimental platform I designed and assembled. The outcome of the experiments shown in this study is strictly dependent on the accuracy of the real time image analysis performed. For this reason I developed an image processing algorithm as much reliable and precise as possible, by exploiting well established principles of image processing (72) . Once bright field (PhC) and fluorescence images are acquired by the camera connected to the microscope, then they are fed to the developed algorithm that is meant to locate cells within each PhC frame, and to use this information to calculate the fluorescence (corresponding to each reporter or fluorophore). For real time image analysis of yeast cells I have adapted and improved an algorithm developed by La Brocca and colleagues (described in (47)); Furthermore I devised an algorithm that my colleagues have been using to quantify, in real time, the fluorescence emitted, upon excitation, by mammalian cells line. Implementation details of both these algorithms are broadly described in the following paragraphs.

8.4.1 Yeast cells

Yeast cells in phase contrast images occur in clustered, low intensity, convex and often quasi-circular shapes surrounded by a white halo (Figure 8.1 A). The contrast between the pixels belonging to the cells and the pixels belonging to the halos is usually so high that edge points can be detected by the evaluation of the magnitude of the gradient calculated in each point of the image Due to the shape of yeast cell, edge points can be connected with the Circular Hough Transform (CHT) (73). CHT can detect almost all cells within the image, even when cells edges overlap.

I implemented a custom image processing algorithm, written in MATLAB programming language, able to discriminate cells from the background of each image in order to calculate the value of the fluorescence emitted by the entire population and even by each single cell; thus the algorithm, from a conceptual perspective, can be divided into two parts: *(a)* cell locating, *b)* fluorescence calculation.

In its first part, the algorithm works on the phase contrast image and employs a sequence of commands meant to filter (remove grainy effects) and to enhance the contrast at cells edges (72), then locates the cells taking advantage of a MATLAB built-in function (*imfindcircles.m*) that implements CHT looking for quasi-circular objects (cells) within a certain radii range. Taking into account the typical dimensions of yeast cells and the magnification used to acquire images, the search radii span can be easily estimated. This MATLAB function returns, as output, the coordinates of the centres and the corresponding radius of each round object identified; The algorithm uses these informations to calculate a binary filter meant to select only pixels within cells (Figure 8.1), this selective binary filter is applied to the GFP field image to select only fluorescence intensity emitted by cells.

Defining the fluorescence field image I as:

$$I : (p) \in \Omega \subset \mathbf{N}^2 \tag{8.10}$$

then

$$I(x, y) \in \left[0, 2^{-L} - 1\right] \subset \mathbf{N} \tag{8.11}$$

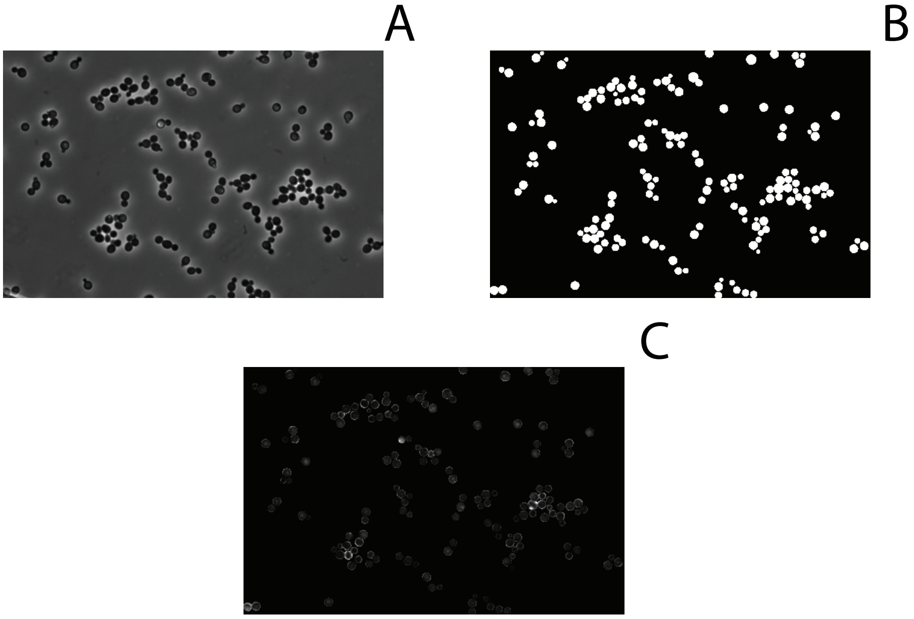


Figure 8.1: Image segmentation of yeast cells. *(A) Yeast cells imaged in phase contrast; all the cells, although of different dimensions, have a quasi-circular shape. (B) Binary filter built finding all the circular objects (cells) whose radii are in a certain search range. (C) Cells selected with the binary image of panel B.*

with x and y generic coordinates and L the number of bits used for image encoding and ω the set of intensity values the pixel in the image can assume.

The mask image $\hat{M}$ can be similarly defined as:

$$\hat{M} : (p) \in 0, 1 \tag{8.12}$$

where $\hat{M}(x, y) = 1$ denotes a cell belonging pixel while $\hat{M}(x, y) = 0$ indicates background pixels.

The latter class of pixels is useful to estimate the amplitude of the background signal, which can be subtracted from the raw signal to obtain a normalised fluorescence intensity. In order to compute the normalised signal, I used the following equation:

$$\mathrm{GFP_{avg}} = \underbrace{\sum_i \sum_j I(x, y) \cdot \hat{M}(x, y)}_{\text{raw GFP signal}} - \underbrace{\sum_i \sum_j I(x, y) \cdot (1 - \hat{M}(x, y))}_{\text{background signal}} \tag{8.13}$$

with i and j spanning the rows and columns, respectively, of the arrays. $\neg\hat{M}(x, y)$ is a transformation of $\hat{M}$ that is simply meant to complement the binary values of the original matrix (so as to select image areas not belonging to cells). The quantity GFP_{avg} is the quantified fluorescence output y used by the control algorithm to define the control input to the cells.

The same algorithm is used to calculated the fluorescence intensity of each single cell within the imaging field. In this case the mask $\hat{M}_k$ is calculated for each identified cell k and, as seen before, it is used to calculate the intensity of the fluorescent pixels belonging only to the $k-th$ cell. The complementary mask is used to calculate the intensity of the background signal that is subtracted from the fluorescence intensity of the single cell.

8.4.2 Mammalian cells

In this study I have analysed microscopy images of Chinese Hamster Ovary (CHO) cells. This cell type, as other mammalian cells, does not have a particular geometrical shape (74), for this reason, the image processing algorithm that I developed is not based on cells' morphological features.

Since the microscope used (Figure 8.3 A) employs an optical apparatus meant to acquire images in phase contrast, cells imaged exhibit a white halo at their edges 8.2; thus I exploited this property of phase contrast images to locate cells within an image. By using the thresholding technique described in (75) it is possible to define a threshold and use it to generate a binary image to select only pixels belonging to cells' edges (Figure 8.2 B). Using morphological operators (dilation and filling) (72) it is possible to obtain a binary image, a mask, that overestimates the area occupied by cells; by subtracting from the mask thus obtained the one retrieved after the thresholding process, it is possible to derive a binary image (Figure 8.2 C) that tightly select the portion of the original image covered by cells (8.2 D).

The latter binary filter obtained, is used as a selection filter to calculate the average intensity fluorescence of pixels belonging to cells, subtracting the background signal (as described in §8.4.1).

8.5 Yeast strains

In this study I've used two different yeast strains yGIL337 (*GAL1* promoter cells) and IC18 (IRMA cells).

GAL1 promoter cells (yGIL337, Gal1-GFP::KanMX, Gal10-mCherry::NatMX) are yeasts constructed by Lang et al. (48) in which the Gal1 protein, expressed by the *GAL1* promoter, was fused to a green fluorescent protein (Gfp) (system dynamics are discussed in §5.1).

IRMA cells (IC18), obtained by Cantone et al. (11), stably integrate in their genome a synthetic network consisting of 5 genes regulating each other via positive and negative feedback loops, and represents one of the most complex synthetic networks built so far (55) (network dynamics are described in §5.3).

8.6 Experimental protocol

In this section all the steps needed to perform the experiments reported in this study are explained. For both identification and control experiments the steps needed to prepare cells and the microfluidic device are the same. Moreover the

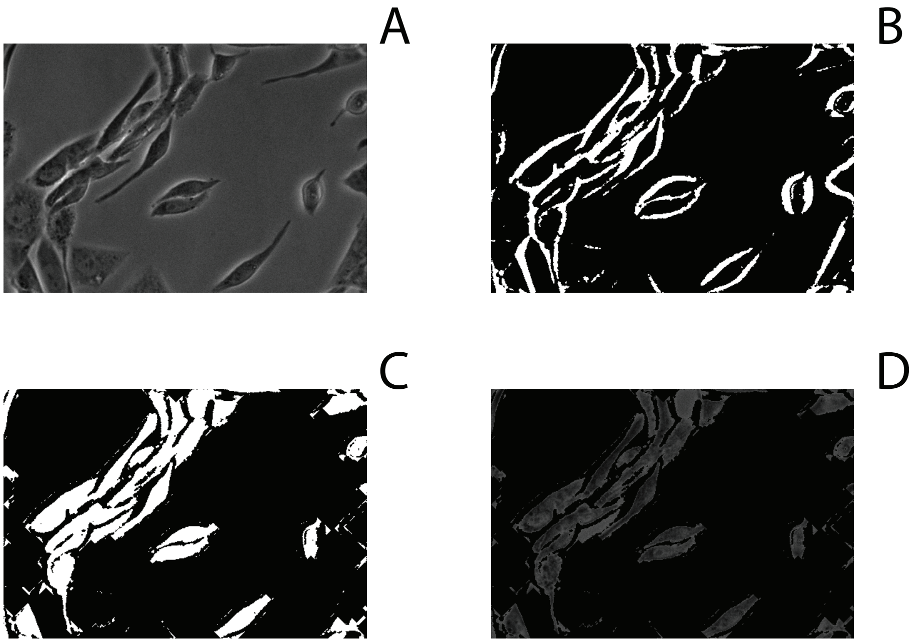


Figure 8.2: Image segmentation of CHO cells. *(A) CHO cells imaged in phase contrast, all cells exhibit different shapes the only characteristic they have in common is the white halo surrounding them, this property is exploited to locate cells within the image. (B) Binary image obtained selecting cells' edges with Otsu's method. (C) Binary image obtained by subtracting the binary image of panel B, from the binary image obtained dilating and filling the same binary image; the result is the binary filter used to select cells within the frame. (D) Selected cells in the frame.*

two yeast strains used in this study exhibit same growth curves when cultured in a Galactose enriched medium; for this reason the culturing protocol applies to both cell types.

8.6.1 Cells and microfluidic device set up

On day 0 batch cultures are inoculated in 10 mL GAL/RAF+Sulforhodamine B (Sigma-Aldrich) (2%) Synthetic Complete medium (SC). On day 1 the batch culture is diluted at intervals of 12 hours (final OD_{600} 0.01). On day 2, 60mL syringes (Becton, Dickinson and Company, NJ) filled with 10 mL SC+GAL/RAF (2%) and SC+GLC (2%) media are prepared, as well as sink syringes (filled with 10 mL ddH2O); capillaries and needles are used to allow connection to the microfluidic device. Temperature in the micro-environment surrounding the moving stage of the microscope is allowed to settle at 30 °C. Before connecting media and sink syringes, the microfluidic device MFD0005a wetting is carried out as described in (44). After air bubbles are removed, media and water filled 60 mL syringes are attached to the device and correct functioning is checked by inspecting the red-fluorescence emitted by Sulforhodamine B as a result of the automatic height control of syringes. This allowed us to carry out a correct calibration of the actuation strategy before the actual experiment is run. At this point cells (IC18 or yGIL337 strain) are injected in the microfluidic device by pouring the batch culture in a 60 mL syringe similar to the ones used to media and sinks. Once cells are trapped in the defined area (see (44) for details) Perfect Focus System is activated to assist autofocusing during the experiment and the acquisition routine of the microscope software is started to initiate image acquisition as explained in paragraph 8.3.

8.6.2 *GAL1* promoter identification experiments

Once cells are loaded in the microfluidic device, they are kept in a Galactose enriched growing medium for 180 minutes (to allow cells adapt to the microchemostat environment) simply by holding the syringe filled with Galactose in a higher position with respect to the one carrying Glucose. After this *calibration phase* of the experiment a sequential MATLAB script controls syringes' positions over the time to obtain desired time profile for the input fed to cells. The image processing algorithm, running in real time, calculates the absolute fluorescence emitted by the entire cell population as well as its normalised value by dividing the time course of fluoresce by the average fluorescence intensity measured during the initial *calibration phase*. Input and output time series thus generated are used, as discussed in §5.1-5.2, to apply System Identification techniques.

8.6.3 *GAL1* promoter and IRMA control experiments

The same experimental procedure, unless explicitly reported, applies to control experiments carried out on *GAL1* promoter and IRMA cells (IC18 and yGIL337 strains).

Set - point control experiments (*GAL1* promoter and IRMA): once cells are loaded in the microfluidic device, the user has to start a custom MATLAB script, that manages the entire experimental platform (controller implementation, actuation, image analysis), and has to set the duration (in minutes) of the control. The script is built to calculate the set point for the control as a percentage (indicated by the user at the beginning of the experiment) of the average of the fluorescence measured by the image processing algorithm during the calibration phase previously described. After this the implemented script proceeds in executing all the code blocks necessary to reach and maintain the fluorescence reference.

Signal - tracking control experiments (*GAL1* promoter): the length and the values of the steps in the step - like time varying reference used in signal tracking control experiments with *GAL1* promoter cells is calculated by a custom MATLAB script that manages the entire experimental platform. The script is built to calculate the values of the step - like reference as percentages (indicated by the user at the beginning of the experiment) of the average of the fluorescence measured by the image processing algorithm during the calibration phase. At the end of the calibration, the implemented script proceeds in executing all the code blocks necessary to reach and maintain the fluorescence reference.

9

Conclusions and future works

A pressing open problem in quantitative biology is to develop integrated experimental and computational system identification approaches to biological processes, such as transcriptional control of gene expression, to derive quantitative dynamical models of complex molecular mechanisms. I believe that the use of a microfluidics based platform, such as the one I developed, can be instrumental for the design of innovative identification strategies and address the need for better quantitative models of biological processes.

The results here described confirm that complex mechanisms underlying transcriptional control from a eukaryotic promoter, which requires the coordinated action of several protein complexes and it is still not completely understood (76), can indeed be identified using standard System Identification strategies without requiring any detailed *a priori* knowledge of the promoter to be modelled (12, 13, 14, 15).

The experimental platform I devised represents a cheap but accurate technological solution that may be used to analyse and model any promoter of interest. However, in order to use the platform to model promoters which are not inducible by small molecules, an extra step is required as depicted in Figure 9.1. The transcription factor, together with a reporter fluorescent protein, has to be cloned downstream of an inducible promoter, in order to be able to generate a time-varying concentration of the transcription factor, which can be used as input for the system identification procedure.

The different identification strategies used show comparable performances indicating that even a linear model with the simplest structure, such as the linear first-order transfer function with a delay, can be effectively used to model promoter dynamics. A linear model may be preferred to nonlinear ones, because of its simplicity and versatility for control purposes as here demonstrated. Moreover, since nonlinear model parameters can be identified only using heuristic method, the extra effort required to tune heuristic algorithm parameters is worthwhile only when a linear model fails to satisfactorily capture the promoter dynamics.

Indeed, there will be cases in which a linear model will not be able to capture the promoter behaviour, such as when adaptation to the input occurs (7). For example, the promoter of the *STL1* gene, which encodes a glycerol protein symporter, is controlled by the transcription factor Hog1, which is activated by osmotic shock. Our experimental system identification platform may be applied using as input the osmotic stress (i.e. increasing extracellular pressure) and, as output, a Gfp reporter protein downstream the *STL1* promoter (7). However, a sustained osmotic stress will at first activate the promoter but then, once enough glycerol is produced to counteract the external pressure, the cell will stop expressing Gfp from *STL1* promoter, because of cell adaptation to osmotic stress (for details refer to (7, 10)). Hence, a linear model would fail in capturing the promoter dynamics and a nonlinear model is needed instead (7).

The experimental system identification platform here described allows fast prototyping of eukaryotic promoters to probe their dynamical behaviour and to identify input-output quantitative models.

The high degree of modularity of the experimental set up implemented has allowed to use it for the external control of gene expression in population of living yeasts.

The control experimental results described here convincingly demonstrate that the expression of a protein can be controlled *in-vivo* in real-time, using an inducer molecule acting directly or indirectly on protein expression, by applying principles drawn from classical control theory, and without requiring detailed quantitative knowledge of the process to be controlled, at least in the case of set-point regulation.

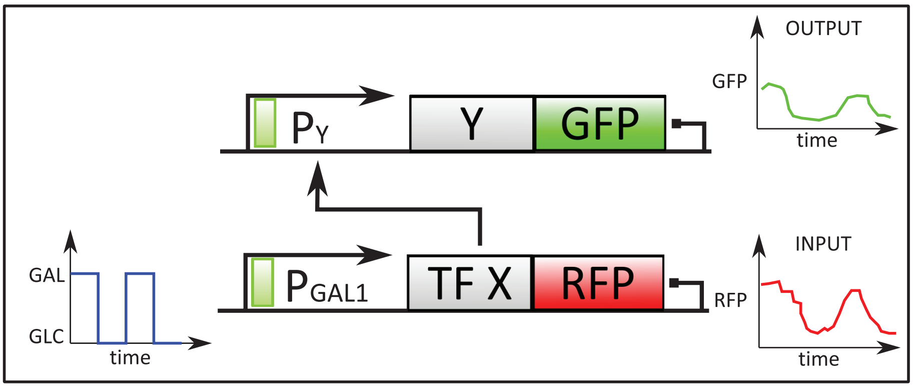


Figure 9.1: Modelling of promoters which are not inducible by small molecules. *In order to use the platform to model promoters that are not inducible by small molecules, the transcription factor (X), together with a reporter fluorescent protein (Rfp), has to be cloned downstream of an inducible promoter (GAL1), in order to be able to generate a time-varying concentration of the transcription factor, which can be used as input for the system identification procedure.*

Moreover for the first time a comparative analysis, of different control strategies, has been proposed by carrying out *in-vivo* experiments meant to assess and compare their performances. In the broad context of the external control of living cells (4, 5, 6, 7, 8, 9) I have proposed a control strategy, Zero Average Dynamics control (67), that has never been applied before to steer gene expression, demonstrating its effectiveness and ability of optimising the stimuli provided to cells.

One of the further advantages of the control approach here proposed is that it can use as input any molecule and thus it may be easily transferred to the control of any other endogenous promoter, or gene network, whose dynamics can be elicited by external molecules and for which a measurable estimate of the output is available.

In addition to providing an innovative platform to control protein expression in a completely automatic fashion, the proposed results show also that binary digital pulses of an inducer molecule can be encoded and interpreted by the cell population to produce an ”analog” response. Digital-to-analog and analog-to-digital conversions are key features of signaling pathways. Gradients of extracellular stimuli are converted into an all-or-none responses by signaling pathways

(77). These digital responses, in turn, are decoded by the cells to generate analog time-varying transcriptional responses (digital-to-analog conversion). Here I have shown that this core mechanism can be exploited by artificial control systems to modify at will gene and protein expression.

I do strongly believe that experimental biologists will find new and clever ways to apply the proposed approach to study trafficking or signalling pathways and the endogenous control mechanisms of a cell. Indeed the ability to simply overexpress a protein has led to innumerable new discoveries, and with my work I have provided a new ability which could be beneficial to many.

References

[1] Lennart Ljung. *System identification.* Wiley Online Library, 1999. 1, 11, 42, 45, 46, 47, 48, 49, 59, 60, 62

[2] Gaetan Kerschen, Keith Worden, Alexander F Vakakis, and Jean-Claude Golinval. **Past, present and future of nonlinear system identification in structural dynamics**. *Mechanical Systems and Signal Processing*, **20**(3):505–592, 2006. 1

[3] Oliver Nelles. *Nonlinear system identification: from classical approaches to neural networks and fuzzy models.* Springer, 2000. 1

[4] Tal Danino, Octavio Mondragón-Palomino, Lev Tsimring, and Jeff Hasty. **A synchronized quorum of genetic clocks**. *Nature*, **463**(7279):326–330, 2010. 1, 2, 24, 25, 26, 132

[5] Andreas Milias-Argeitis, Sean Summers, Jacob Stewart-Ornstein, Ignacio Zuleta, David Pincus, Hana El-Samad, Mustafa Khammash, and John Lygeros. **In silico feedback for in vivo regulation of a gene expression circuit**. *Nature biotechnology*, **29**(12):1114–1116, 2011. 1, 2, 24, 27, 29, 32, 97, 132

[6] Jared E Toettcher, Delquin Gong, Wendell A Lim, and Orion D Weiner. **Light-based feedback for controlling intracellular signaling dynamics**. *Nature methods*, **8**(10):837–839, 2011. 1, 2, 24, 29, 30, 132

[7] Jannis Uhlendorf, Agnès Miermont, Thierry Delaveau, Gilles Charvin, François Fages, Samuel Bottani, Gregory Batt, and Pascal Hersen. **Long-term model predictive control of gene expression at the population and single-cell levels**. *Proceedings of the National Academy of Sciences*, **109**(35):14271–14276, 2012. 1, 2, 24, 26, 28, 97, 131, 132

[8] Justin Melendez, Michael Patel, Benjamin L Oakes, Ping Xu, Patrick Morton, and Megan N McClean. **Real-time optogenetic control of intracellular protein concentration in microbial cell cultures**. *Integrative Biology*, **6**(3):366–372, 2014. 1, 2, 24, 31, 33, 132

[9] Evan J Olson, Lucas A Hartsough, Brian P Landry, Raghav Shroff, and Jeffrey J Tabor. **Characterizing bacterial gene circuit dynamics with optically programmed gene expression signals**. *Nature methods*, **11**(4):449–455, 2014. 1, 2, 24, 30, 31, 132

[10] Jerome T Mettetal, Dale Muzzey, Carlos Gómez-Uribe, and Alexander van Oudenaarden. **The frequency dependence of osmo-adaptation in Saccharomyces cerevisiae**. *Science Signaling*, **319**(5862):482, 2008. 2, 26, 131

[11] Irene Cantone, Lucia Marucci, Francesco Iorio, Maria Aurelia Ricci, Vincenzo Belcastro, Mukesh Bansal, Stefania Santini, Mario di Bernardo, Diego di Bernardo, Maria Pia Cosma, et al. **A yeast synthetic network for in vivo assessment of reverse-engineering and modeling approaches**. *Cell*, **137**(1):172, 2009. 2, 3, 41, 66, 67, 68, 69, 78, 83, 84, 90, 126

[12] Trey Ideker, Timothy Galitski, and Leroy Hood. **A new approach to decoding life: systems biology**. *Annual review of genomics and human genetics*, **2**(1):343–372, 2001. 2, 130

[13] Jeff Hasty, David McMillen, Farren Isaacs, and James J Collins. **Computational studies of gene regulatory networks: in numero molecular biology**. *Nature Reviews Genetics*, **2**(4):268–279, 2001. 2, 130

[14] Hiroaki Kitano. **Computational systems biology**. *Nature*, **420**(6912):206–210, 2002. 2, 130

[15] John J Tyson, Kathy Chen, and Bela Novak. **Network dynamics and cell physiology**. *Nature Reviews Molecular Cell Biology*, **2**(12):908–916, 2001. 2, 130

[16] Pablo A Iglesias and Brian P Ingalls. *Control theory and systems biology.* The MIT Press, 2010. 2

[17] Giovanni Russo, Mario Di Bernardo, and Eduardo D Sontag. **Global entrainment of transcriptional systems to periodic inputs**. *PLoS computational biology*, **6**(4):e1000739, 2010. 2

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

[19] Hiroyuki Kurata, Hana El-Samad, Rei Iwasaki, Hisao Ohtake, John C Doyle, Irina Grigorova, Carol A Gross, and Mustafa Khammash. **Module-based analysis of robustness tradeoffs in the heat shock response system**. *PLoS computational biology*, **2**(7):e59, 2006. 2

[20] Hana El-Samad and Mustafa Khammash. **Regulated degradation is a mechanism for suppressing stochastic fluctuations in gene regulatory networks**. *Biophysical journal*, **90**(10):3749–3761, 2006. 2

[21] Kevin F Murphy, Rhys M Adams, Xiao Wang, Gabor Balazsi, and James J Collins. **Tuning and controlling gene expression noise in synthetic gene networks**. *Nucleic acids research*, **38**(8):2712–2726, 2010. 2

[22] Joshua F Apgar, Jared E Toettcher, Drew Endy, Forest M White, and Bruce Tidor. **Stimulus Design for Model Selection and Validation in Cell Signaling**. *PLoS Comput Biol*, **4**(2), 02 2008. 2

[23] Neda Bagheri, Jörg Stelling, and FJ Doyle. **Circadian phase entrainment via nonlinear model predictive control**. *International Journal of Robust and Nonlinear Control*, **17**(17):1555–1571, 2007. 2

[24] Eric Klavins. **Proportional-integral control of stochastic gene regulatory networks**. In *Decision and Control (CDC), 2010 49th IEEE Conference on*, pages 2547–2553. IEEE, 2010. 2

[25] Velia Siciliano, Filippo Menolascina, Lucia Marucci, Chiara Fracassi, Immacolata Garzilli, Maria Nicoletta Moretti, and Diego di Bernardo. **Construction and modelling of an inducible positive feedback loop stably integrated in a mammalian cell-line**. *PLoS computational biology*, **7**(6):e1002074, 2011. 3, 41, 68, 71, 72, 73, 74, 114

[26] M Madan Babu, Nicholas M Luscombe, L Aravind, Mark Gerstein, and Sarah a Teichmann. **Structure and evolution of transcriptional regulatory networks.** *Current opinion in structural biology*, **14**(3):283–91, June 2004. 7

[27] Uri Alon. *Introduction to Systems Biology: And the Design Principles of Biological Networks*, **10**. CRC press, 2007. 7, 9

[28] Tong Ihn Lee, Nicola J Rinaldi, François Robert, Duncan T Odom, Ziv Bar-Joseph, Georg K Gerber, Nancy M Hannett, Christopher T Harbison, Craig M Thompson, Itamar Simon, et al. **Transcriptional regulatory networks in Saccharomyces cerevisiae**. *science*, **298**(5594):799–804, 2002. 7, 8, 10

[29] Nitzan Rosenfeld, Michael B Elowitz, and Uri Alon. **Negative Autoregulation Speeds the Response Times of Transcription Networks**. *Journal of Molecular Biology*, **323**(5):785–793, November 2002. 7

[30] Yann Dublanche, Konstantinos Michalodimitrakis, Nico Kümmerer, Mathilde Foglierini, and Luis Serrano. **Noise in transcription negative feedback loops: simulation and experimental analysis.** *Molecular systems biology*, **2**:41, January 2006. 9

[31] Yusuke T Maeda and Masaki Sano. **Regulatory dynamics of synthetic gene networks with positive feedback.** *Journal of molecular biology*, **359**(4):1107–24, June 2006. 9

[32] Velia Siciliano, Immacolata Garzilli, Chiara Fracassi, Stefania Criscuolo, Simona Ventre, and Diego di Bernardo. **MiRNAs confer phenotypic robustness to gene networks by suppressing biological noise**. *Nature communications*, **4**, 2013. 9

[33] James E Ferrell. **Self-perpetuating states in signal transduction: positive feedback, double-negative feedback and bistability**. *Current Opinion in Cell Biology*, **14**(2):140–148, April 2002. 9

[34] Carmine Settembre and Andrea Ballabio. **TFEB regulates autophagy**. *Autophagy*, **7(11)**(November):1379–1381, 2011. 10

[35] Artem K Velichko, Elena N Markova, Nadezhda V Petrova, Sergey V Razin, and Omar L Kantidze. **Mechanisms of heat shock response in mammals.** *Cellular and molecular life sciences : CMLS*, **70**(22):4229–41, November 2013. 10

[36] Jaime J Carvajal and Peter W J Rigby. **Regulation of gene expression in vertebrate skeletal muscle.** *Experimental cell research*, **316**(18):3014–8, November 2010. 10

[37] Karl Johan Aström and Richard M Murray. *Feedback systems: an introduction for scientists and engineers*. Princeton university press, 2010. 13, 15, 16, 17, 18, 63, 82, 96, 97, 99, 100, 112

[38] Fuzhong Zhang, James M Carothers, and Jay D Keasling. **Design of a dynamic sensor-regulator system for production of chemicals and fuels derived from fatty acids**. *Nature biotechnology*, **30**(4):354–359, 2012. 21, 22

[39] Eric J Steen, Yisheng Kang, Gregory Bokinsky, Zhihao Hu, Andreas Schirmer, Amy McClure, Stephen B Del Cardayre, and Jay D Keasling. **Microbial production of fatty-acid-derived fuels and chemicals from plant biomass**. *Nature*, **463**(7280):559–562, 2010. 21, 22

[40] Victoria Hsiao, Emmanuel LC de los Santos, Weston R Whitaker, John E Dueber, and Richard M Murray. **Design and implementation of a biomolecular concentration tracker**. *ACS synthetic biology*, 2014. 23

[41] Patrick Tabeling. *Introduction to microfluidics*. Oxford University Press, 2010. 25

[42] Manfred Morari and Jay H Lee. **Model predictive control: past, present and future**. *Computers & Chemical Engineering*, **23**(4):667–682, 1999. 27, 101

[43] Rudolph Emil Kalman. **A new approach to linear filtering and prediction problems**. *Journal of Fluids Engineering*, **82**(1):35–45, 1960. 27, 100

[44] MS Ferry, IA Razinkov, and J Hasty. **Microfluidics for synthetic biology from design to execution**. *Methods Enzymol*, **497**:295, 2011. 36, 37, 81, 117, 122, 128

[45] Manda S Williams, Kenneth J Longmuir, and Paul Yager. **A practical guide to the staggered herringbone mixer**. *Lab on a Chip*, **8**(7):1121–1129, 2008. 37

[46] Bruno Siciliano and Oussama Khatib. *Springer handbook of robotics*. Springer Science & Business Media, 2008. 38, 121

[47] Raffaele La Brocca, F Menolascina, D di Bernardo, and C Sansone. **Segmentation, tracking and lineage analysis of yeast cells in bright field microscopy images**. *PR PS BB, Ravenna, Italy*, 2011. 39, 122

[48] Gregory I Lang and David Botstein. **A Test of the Coordinated Expression Hypothesis for the Origin and Maintenance of the GAL Cluster in Yeast**. *PloS one*, **6**(9):e25290, 2011. 41, 42, 126

[49] G Fiore, F Menolascina, M di Bernardo, and D di Bernardo. **An experimental approach to identify dynamical models of transcriptional regulation in living cells**. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, **23**(2):025106–025106, 2013. 41

[50] Matthew R Bennett, Wyming Lee Pang, Natalie A Ostroff, Bridget L Baumgartner, Sujata Nayak, Lev S Tsimring, and Jeff Hasty. **Metabolic gene regulation in a dynamically changing environment**. *Nature*, **454**(7208):1119–1122, 2008. 42, 43, 77

[51] Uri Alon. **An Introduction to Systems Biology: Design Principles of Biological Circuits (Mathematical and Computational Biology Series vol 10)**, 2007. 47

[52] Scott Kirkpatrick, D. Gelatt Jr., and Mario P Vecchi. **Optimization by simulated annealing**. *science*, **220**(4598):671–680, 1983. 48

[53] R Apostu and Mackey MC. **Mathematical model of GAL regulon dynamics in Saccharomyces cerevisiae**. *Journal of Theoretical Biology*, pages 219–235, 2012. 58

[54] Diane Longo and Jeff Hasty. **Dynamics of single-cell gene expression**. *Molecular systems biology*, **2**(1), 2006. 58

[55] Diogo M Camacho and James J Collins. **Systems biology strikes gold**. *Cell*, **137**(1):24–26, 2009. 66, 126

[56] Maria Pia Cosma, Tomoyuki Tanaka, and Kim Nasmyth. **Ordered recruitment of transcription and chromatin remodeling factors to a cell cycle–and developmentally regulated promoter**. *Cell*, **97**(3):299–311, 1999. 66, 67

[57] Nir Yosef and Aviv Regev. **Impulse control: temporal dynamics in gene transcription**. *Cell*, **144**(6):886–896, 2011. 66, 75

[58] 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. 76, 87

[59] Hidde De Jong. **Modeling and simulation of genetic regulatory systems: a literature review**. *Journal of computational biology*, **9**(1):67–103, 2002. 78

[60] Jean-Luc Gouzé and Tewfik Sari. **A class of piecewise linear differential equations arising in biological models**. *Dynamical systems*, **17**(4):299–316, 2002. 78

[61] Jamil Ahmad, Gilles Bernot, J-P Comet, Didier Lime, and Olivier Roux. **Hybrid modelling and dynamical analysis of gene regulatory networks with delays**. *ComPlexUs*, **3**(4):231–251, 2007. 78

[62] John G Kassakian, Martin F Schlecht, and George C Verghese. *Principles of power electronics*, **46**. Addison-Wesley Reading, USA, 1991. 80

[63] Andrew R Teel, Luc Moreau, and D Nešić. **Input to state set stability for pulse width modulated control systems with disturbances**. *Systems & control letters*, **51**(1):23–32, 2004. 80, 85

[64] Soumitro Banerjee and George C Verghese. *Nonlinear phenomena in power electronics*. IEEE press New York, 2001. 80

[65] Gábor Balázsi, Alexander van Oudenaarden, and James J Collins. **Cellular decision making and biological noise: from microbes to mammals**. *Cell*, **144**(6):910–925, 2011. 87, 90

[66] E Fossas, R Grinó, and D Biel. **Quasi-Sliding control based on pulse width modulation, zero averaged dynamics and the L2 norm**. *Advances in Variable Structure System, Analysis, Integration and Applications*, pages 335–344, 2001. 97, 102

[67] Rafael R Ramos, Domingo Biel, Enric Fossas, and Francesc Guinjoan. **A fixed-frequency quasi-sliding control algorithm: application to power inverters design by means of FPGA implementation**. *Power Electronics, IEEE Transactions on*, **18**(1):344–355, 2003. 97, 103, 132

[68] Gene F Franklin, J David Powell, and Michael L Workman. *Digital control of dynamic systems*, **3**. Addison-wesley Menlo Park, 1998. 98

[69] David E Golberg. **Genetic algorithms in search, optimization, and machine learning**. *Addion wesley*, **1989**, 1989. 102

[70] Jean-Jacques E Slotine, Weiping Li, et al. *Applied nonlinear control*, **60**. Prentice-Hall Englewood Cliffs, NJ, 1991. 102

[71] Martin Kolnik, Lev S Tsimring, and Jeff Hasty. **Vacuum-assisted cell loading enables shear-free mammalian microfluidic culture**. *Lab on a chip*, **12**(22):4732–4737, 2012. 113

[72] Rafael C.. Gonzalez, Richard E.. Woods, and Steven L.. Eddins. *Digital Image Processing Using MATLAB®*. Tata McGraw Hill Education, 2010. 122, 123, 126

[73] Simon Just Kjeldgaard Pedersen. **Circular hough transform**. *Aalborg University, Vision, Graphics, and Interactive Systems*, 2007. 123

[74] Yu P Petrov and NV Tsupkina. **Growth characteristics of CHO cells in culture**. *Cell and Tissue Biology*, **7**(1):72–78, 2013. 125

[75] Nobuyuki Otsu. **A threshold selection method from gray-level histograms**. *Automatica*, **11**(285-296):23–27, 1975. 126

[76] Roger D Kornberg. **The molecular basis of eukaryotic transcription**. *Proceedings of the National Academy of Sciences*, **104**(32):12955–12961, 2007. 130

[77] Boris Kholodenko, Michael B Yaffe, and Walter Kolch. **Computational approaches for analyzing information flow in biological networks**. *Science signaling*, **5**(220):re1, 2012. 133

Declaration

I herewith declare that I have produced this study without the prohibited assistance of third parties and without making use of aids other than those specified; notions taken over directly or indirectly from other sources have been identified as such.

The thesis work was conducted from 01/03/2012 to 28/02/2015 under the supervision of Dr. Diego di Bernardo in the Systems and Synthetic Biology Laboratory at the TeleThon Institute of Genetics and Medicine (Naples, ITA), and of Prof. Mario di Bernardo at the University of Naples "Federico II" (Naples, ITA)

Naples, 31/03/2015