University of Pennsylvania
ScholarlyCommons

Departmental Papers (MEAM)

Department of Mechanical Engineering & Applied Mechanics

7-11-2007

Controlling Biological Systems: The Lactose Regulation System of Escherichia Coli

Anak Agung Julius
University of Pennsylvania

Ádam Halasz
University of Pennsylvania

Vijay Kumar
University of Pennsylvania, kumar@grasp.upenn.edu

George Pappas
University of Pennsylvania, pappasg@seas.upenn.edu

Suggested Citation:
Julius, Anak Agung, Adam Halasz, Vijay Kumar and George Pappas. (2007). *Controlling biological systems: the lactose regulation system of Escherichia coli.* 2007 American Control Conference. New York, New York. July 11-13, 2007.

©2007 IEEE. Personal use of this material is permitted. However, permission to reprint/republish this material for advertising or promotional purposes or for creating new collective works for resale or redistribution to servers or lists, or to reuse any copyrighted component of this work in other works must be obtained from the IEEE.

This paper is posted at ScholarlyCommons. http://repository.upenn.edu/meam_papers/233
For more information, please contact repository@pobox.upenn.edu.

Controlling Biological Systems: The Lactose Regulation System of Escherichia Coli

Abstract
In this paper we present a comprehensive framework for abstraction and controller design for a biological system. The first half of the paper concerns modeling and model abstraction of the system. Most models in systems biology are deterministic models with ordinary differential equations in the concentration variables. We present a stochastic hybrid model of the lactose regulation system of E. coli bacteria that capture important phenomena which cannot be described by continuous deterministic models. We then show that the resulting stochastic hybrid model can be abstracted into a much simpler model, a two-state continuous time Markov chain.

The second half of the paper discusses controller design for a specific architecture. The architecture consists of measurement of a global quantity in a colony of bacteria as an output feedback, and manipulation of global environmental variables as control actuation. We show that controller design can be performed on the abstracted (Markov chain) model and implementation on the real model yields the desired result.

Disciplines
Engineering | Mechanical Engineering

Comments
Suggested Citation:
Julius, Anak Agung, Adam Halasz, Vijay Kumar and George Pappas. (2007). *Controlling biological systems: the lactose regulation system of Escherichia coli.* 2007 American Control Conference. New York, New York. July 11-13, 2007.

©2007 IEEE. Personal use of this material is permitted. However, permission to reprint/republish this material for advertising or promotional purposes or for creating new collective works for resale or redistribution to servers or lists, or to reuse any copyrighted component of this work in other works must be obtained from the IEEE.

This journal article is available at ScholarlyCommons: http://repository.upenn.edu/meam_papers/233

Proceedings of the 2007 American Control Conference
Marriott Marquis Hotel at Times Square
New York City, USA, July 11-13, 2007

Controlling biological systems: the lactose regulation system of *Escherichia coli*

A. Agung Julius, Adam Halasz, Vijay Kumar and George J. Pappas

***Abstract*— In this paper we present a comprehensive framework for abstraction and controller design for a biological system. The first half of the paper concerns modeling and model abstraction of the system. Most models in systems biology are deterministic models with ordinary differential equations in the concentration variables. We present a stochastic hybrid model of the lactose regulation system of E. coli bacteria that capture important phenomena which cannot be described by continuous deterministic models. We then show that the resulting stochastic hybrid model can be abstracted into a much simpler model, a two-state continuous time Markov chain.**

The second half of the paper discusses controller design for a specific architecture. The architecture consists of measurement of a global quantity in a colony of bacteria as an output feedback, and manipulation of global environmental variables as control actuation. We show that controller design can be performed on the abstracted (Markov chain) model and implementation on the real model yields the desired result.

I. Introduction

In this paper we present a framework that consists of modeling, abstraction and control of a biological system, namely, the lactose regulation system of the *Escherichia coli* bacteria. The conceptual idea behind the paper is captured in Figure 1. Roughly speaking, the paper can be divided into two parts. The first part corresponds to the lower half of the hourglass in Figure 1, that discusses modeling of the *lac* regulation system as a stochastic hybrid system. The model presented in this paper is a slight modification of the one presented in our earlier work [1]. We also discuss how the stochastic hybrid model can be abstracted into a two-state continuous time Markov chain, and demonstrate how this abstraction is consistent with the macroscopic behavior of a colony of bacteria.

The second part of the paper pertains to the upper half of the hourglass, as it discusses feedback controller synthesis with the abstract model (two-state continuous time Markov chain) as the plant model. We also demonstrate that the controller designed for the abstract model yields the desired behavior when implemented on the actual system, which is a colony of bacteria. The control goal, in this case, is to make a certain fraction of the population induced.

The first part of the paper thus resides in the domain of biology/biophysics, where model building is performed. The second part of the paper is arguably of control/engineering nature, where the problem of controller design and synthesis for a particular control problem is discussed. The abstract model can be viewed as a divider between the two domains, where controller design can be done with the abstract model, without the control engineer having to know about the biology/biophysics aspect of the system.

This work was partially supported by the NSF Presidential Early CAREER (PECASE) Grant 0132716, NIH-NLM Individual Biomedical Informatics Fellowhip award 5-F37-LM008343-1, and the ARO Grant W911NF-05-1-0219

The authors are with the GRASP Laboratory, School of Engineering and Applied Sciences, University of Pennsylvania, Philadelphia, PA 19104, U.S.A. agung,halasz,kumar,pappasg@grasp.upenn.edu

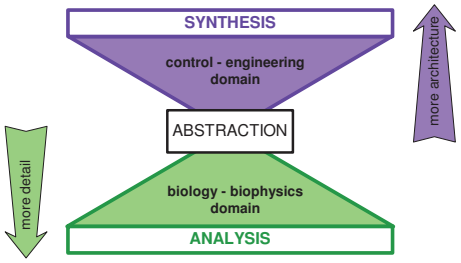

Fig. 1. The hourglass paradigm presented in this paper.

The lactose regulation system in *E. coli* [2] is one of the most extensively studied examples of positive feedback in a naturally occurring gene network. Two of its three component genes encode enzymes (β-galactosidase and permease) which contribute to the synthesis of allolactose which in turn acts as an inducer for the operon itself. Hysteresis and bistability on the level of the entire bacterial population was identified early on by Monod and Pappenheimer [3]. Novick and Weiner [4] discovered bistability at the level of individual cells by studying the expression of β-galactosidase in a population of identical *E. coli* cells. They showed that cells were essentially in one of two discrete states: either fully induced, with enzyme levels close to maximum or uninduced, with negligible enzyme levels. The observation of intermediate activity on the level of the entire population reflects comparably sized sub-populations of induced and uninduced bacteria.

The lactose control system, encoded by the *lac* operon, is often used as a switch to control genes in genetically engineered systems [5], [6].

Having in mind applications where a graded setting of a protein level is needed, we would like to investigate whether such a response at the level of a bacterial population can be ensured by a macroscopic feedback controller design while maintaining the underlying bistable behavior on the level of individual cells. Possible applications are in using bacterial populations in large scale synthesis in drug production [7], novel energy sources [8]. Although control of biological and biochemical systems is not a new field, to the authors' knowledge there has not been other attempts at controlling the induction level of *E. coli* bacteria using macroscopic feedback controller.

The rest of this paper is organized as follows. In Section II we present the mathematical model of the system under consideration, together with its abstraction. In Section III, we discuss the construction of a feedback control mechanism based on the abstract model of the system. Some simulation results for the controller described in Section III are presented in Section IV. We conclude the paper with Section V, where we present a few potential future research directions.

II. Modeling the Lactose Regulation System

A. Deterministic continuous model

Our starting point is a model of the lactose system due to Yildirim and Mackey [1], [9], adapted to the use of *thiomethyl galactosidase* (TMG) as inducer. Briefly, the mRNA (M) transcribed from the lactose operon is translated into three different gene products, among them permease (P) and β-galactosidase (B). Permease facilitates the influx of inducer TMG (T) from the exterior and also an opposing process, equilibrating the concentration of lactose inside the cell with the external lactose. The enzyme β-galactosidase plays a role in the lactose induced network. However, in the TMG induced network, it is effectively decoupled from the equations of motion and has no effect on the dynamics of the remaining three variables. It is however the experimental quantity that is traditionally used as the observable, for example in [4]. A feedback loop is formed by the effect of TMG on the transcription of the lac operon. This complicated relationship involves substances not explicitly considered in the Yildirim-Mackey model, and results in the nonlinear activation function summarized by the first and second terms in Equation (1a).

The equations of motion for the TMG induced network are as follows:

$$\frac{dM}{dt} = \alpha_M \frac{1 + K_1(e^{-\mu\tau_M}T(t-\tau_M))^n}{K + K_1(e^{-\mu\tau_M}T(t-\tau_M))^n} + \Gamma_0 - \tilde{\gamma}_M M, \quad (1a)$$

$$\frac{dB}{dt} = \alpha_B e^{-\mu\tau_B} M(t-\tau_B) - \tilde{\gamma}_B B, \quad (1b)$$

$$\frac{dT}{dt} = \alpha_L P \frac{T_e}{K_{T_e} + T_e} - \beta_L P \frac{T}{K_{L_1} + T} - \tilde{\gamma}_L T, \quad (1c)$$

$$\frac{dP}{dt} = \alpha_P e^{-\mu(\tau_P+\tau_B)} M(t-\tau_P-\tau_B) - \tilde{\gamma}_P P. \quad (1d)$$

We take into account time delays due to transcription and translation. Variables without an argument are taken at time t, time delays are indicated by an explicit argument, e.g., $M(t-\tau_B)$ is the value of the variable M delayed with τ_B.

The symbol T_e in equation (1c) signifies the external TMG concentration. If the system is to be viewed as an input-state system, then T_e can be thought of as an input to the system, while the other four concentrations are the state variables. The other symbols in the equation are constant parameters. Because of space limitation, we refer the reader to [10] for the values of these parameters.

For the value of T_e between 1.4 - 32 μM, the system has three equilibria. Two of these equilibria are stable, giving rise to bistability of the system. Varying the value of T_e causes a hysteresis behavior. See Figure 2 for the bifurcation diagram of the system. The model (1) qualitatively reproduces the observed experimental behavior. The higher-dimensional version defined in [9] and discussed in [1], very closely approximates that behavior. However, that model (with its original parameter set) is correct only as a description of the average behavior of a large number of cells described as a single "reactor". This is because stochastic behavior on the level of individual cells is ignored both in its construction and in its validation.

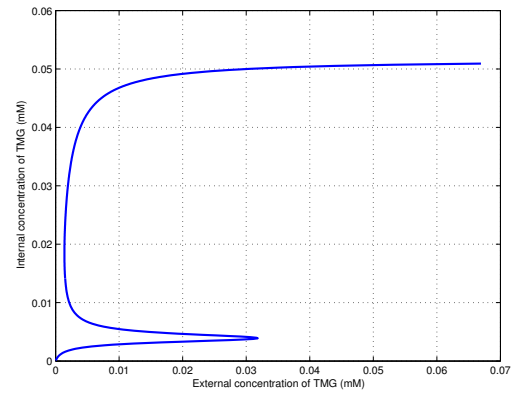

Fig. 2. The equilibria of the system given by (1). The middle range of T_e has three branches of equilibria.

It has been recognized early on that the observed concentrations on the level of a very large number of cells is actually an average over two distinct sub-populations of cells, whose β-galactosidase level takes one of two extreme values. This microscopic bistability was termed the "all-or-none" phenomenon [4]. Although individual cells spend most of the time in one of the two stable equilibria, a closer examination of traditional induction experiments reveals that the population average cannot be maintained at these extreme values. Rather, it can converge to intermediate values between the two extremes.

Following the discussion in [1], we identify the necessity of introducing stochasticity in the model to bridge the gap between the macroscopic behavior (population average) and microscopic behavior.

B. Stochastic model

While stochasticity is sometimes thought of as leading to small deviations from the ODE prediction, it actually may often lead to qualitatively different behavior [11]. There are several sources of randomness or noise in biochemical processes [12], [13]. Here we focus on "intrinsic" stochastic phenomena related to the *small copy number* of molecule species inside individual cells. Chemical reactions, at the microscopical level, amount to creation and breaking up of chemical molecules. These processes can be modeled as Poisson random processes [14], [15], whose rates depend on the state of the system, i.e. the number of molecules in the reaction. We develop a hybrid stochastic model for the system, where only a subset of the species is treated stochastically and the remainder are treated as continuous variables that obey traditional ODEs [1]. This computationally less expensive approach allows us to perform quasi-simultaneous simulations for many cells. The connection between the ODE and the stochastic description is through the the conversion constant $C_N = 6.023 \cdot 10^4 \frac{\text{molecules}}{\text{mM}}$.

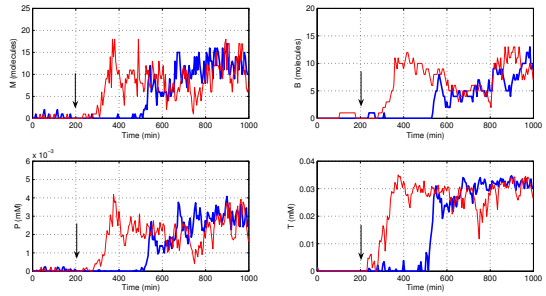

Fig. 3. Simulation results of two cells. In this plot, the external concentration of TMG is increased at $t = 200$ minutes (marked by the arrows). We observe that spontaneous inductions occur approximately 200 and 400 minutes later.

In terms of stochastic differential equations, our hybrid stochastic model can be written as follows.

$$dM_t = d\hat{M}_t - d\tilde{M}_t, \tag{2a}$$

$$dB_t = d\hat{B}_t - d\tilde{B}_t, \tag{2b}$$

$$\frac{dT_t}{dt} = \frac{T_e \alpha_L P_t}{K_{L_e} + T_e} - \frac{\beta_L P_t T_t}{K_{L_1} + T_t} - \tilde{\gamma}_L T_t, \tag{2c}$$

$$\frac{dP_t}{dt} = \alpha_P e^{-\mu(\tau_P + \tau_B)} \frac{M_{(t-\tau_P-\tau_B)}}{C_N} - \tilde{\gamma}_P P_t. \tag{2d}$$

Here the processes $\hat{M}_t$ and $\tilde{M}_t$ are the Poisson processes that are responsible for the creation and breaking up of the messenger RNA molecules, respectively. Similarly, $\hat{B}_t$ and $\tilde{B}_t$ are the Poisson processes that are responsible for the creation and breaking up of the β - galactosidase molecules, respectively. The rates of these processes are given by the reaction rates of the ODE model (1). We simulate the model (2) using a numerical scheme similar to the explicit tau-leaping method for Gillespie simulation, combined with a fixed time step Euler scheme for the continuous variables. The size of the time steps is fixed at 0.1 minute. Our implementation also takes into account the time delays. For more details we refer to [1]. A pair of typical simulation traces for individual cells are shown in Figure 3.

C. Two state Markov chain model

Similarly to [1], the bulk behavior can be well described by a simple two-level abstraction of the states of an individual cell, as illustrated in Figure 7, as a continuous time Markov chain [16]. The states of the Markov chain correspond to the low and high stable equilibria of the systems, also known as the induced and uninduced states. The rates of switching between the two states are given as a function of the external TMG concentration T_e. See Figure 4 for a diagram of the system.

This simple model can closely reproduce the bulk behavior of a large number of cells. We run the simulation of the full model (given in (2)) to simulate a colony of 1000 cells. The macroscopic behavior of the colony, which is computed as the average across the 1000 samples are plotted in Figure 5. It is this colony-level average behavior that is observed in macroscopic experiments like Novick and Weiner's. The experimental curves as well as those corresponding to the

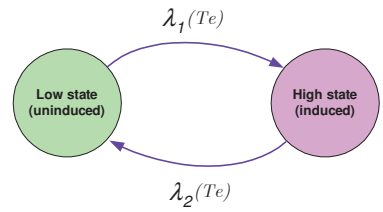

Fig. 4. The two-state continuous time Markov chain model.

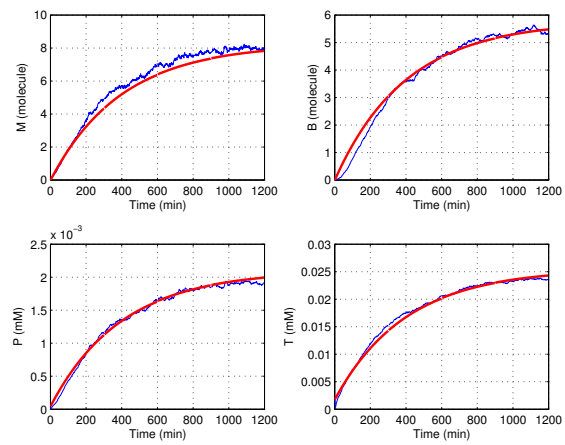

Fig. 5. The average behavior of a colony with 1000 bacteria from stochastic simulations. The exponential curve is plotted to show that the macroscopic behavior can be fitted quite well with a first order dynamics (given in (3)).

average of many individual simulations are well matched by a two-exponential behavior. This is a strong argument for the validity of our simple two-level model, where the evolution of the probability of the two states is given by

$$\frac{d}{dt}\begin{bmatrix} x_{\text{lo}} \\ x_{\text{hi}} \end{bmatrix} = \begin{bmatrix} -\lambda_1(T_e) & \lambda_2(T_e) \\ \lambda_1(T_e) & -\lambda_2(T_e) \end{bmatrix} \begin{bmatrix} x_{\text{lo}} \\ x_{\text{hi}} \end{bmatrix}. \tag{3}$$

The connection between the full model and the abstraction is as follows. The rate of the exponential curve, λ_1, should match $1/\tau$, which is the mean time to transit from the low stable state to the high stable state. We compute this time average from the 1000 samples and use its value to compute the exponent of the curves in Figure 8. We can observe that the fit is good.

We will rely on the two-level abstraction (3) to design a control strategy for obtaining a graded response from a colony of individually bacteria, by modulating the external TMG. We will use simulation on the full model for validation.

III. Controller Synthesis

The architecture of the control system that we discuss in this paper is illustrated in Figure 6. The plant to be controlled in a colony of *E. coli* bacteria. The controller affects the plant by adjusting of the external concentration of TMG in the environment. Feedback information is read from the plant in the form of a global quantity, which we consider as the *output* of the control system. By this, we mean the controller does not have any information about the individual cells in the

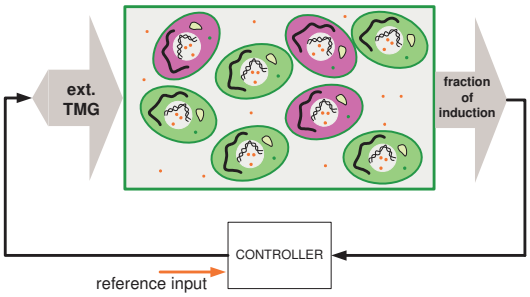

Fig. 6. The control block diagram.

colony. Rather, a global quantity, for example, the fraction of induced cells in the population, is fed into the controller. The control goal is to make the output track a given reference trajectory or attain a desired level.

Feedback control of a group of Markov chains by adjusting the transition rates has been studied, for example in [17]. There, the plant is a group of artificial muscle cells that can switch between contracting and noncontracting states.

Recall the two state Markov chain model of the bacteria. We denote the probability of finding the cell at time t in the induced state as $x_{hi}(t)$, and in the uninduced state as $x_{lo}(t)$. The evolution of the variables x_{hi} and x_{lo} satisfies differential equation (3). Suppose that we have N cells, and for each cell, we introduce an output/observation map

$$y_t^i = \begin{cases} y_{hi}, & \text{if the } i\text{-th cell is induced at time } t, \\ y_{lo}, & \text{if the } i\text{-th cell is uninduced at time } t. \end{cases} \tag{4}$$

with y_{hi} and y_{lo} both fixed real numbers. Obviously $\{y_t^i\}_{i\in\{1,\dots,N\}}$ are random processes. Furthermore, we denote the average output across the population as another random process $\bar{y}_t^N$,

$$\bar{y}_t^N := \frac{1}{N}\sum_{i=1}^{N} y_t^i.$$

Notice that we explicitly write down the dependency of the average output on the size of the population.

Suppose that we are given a control problem, namely we want to make $\bar{y}_t$ track a certain given trajectory $\eta(t)$. We propose the following solution. Consider the following model of control system

$$\frac{d}{dt}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} -\lambda_1(u) & \lambda_2(u) \\ \lambda_1(u) & -\lambda_2(u) \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix},$$
$$y(t) = y_{lo}x_1(t) + y_{hi}x_2(t)$$

Suppose that we have an output feedback law

$$u(t) = f(y(t)), \tag{5}$$

such that the closed loop system

$$\frac{d}{dt}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} -\lambda_1(f(y)) & \lambda_2(f(y)) \\ \lambda_1(f(y)) & -\lambda_2(f(y)) \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}, \tag{6}$$

$$\begin{bmatrix} x_1(0) \\ x_2(0) \end{bmatrix} = \begin{bmatrix} x_{1,0} \\ x_{2,0} \end{bmatrix}, y(t) = y_{lo}x_1(t) + y_{hi}x_2(t), \tag{7}$$

produces the output trajectory

$$y(t) = \eta(t). \tag{8}$$

We can state the following theorem about the effect of a feedback control law on the behavior of the system.

Theorem 3.1: If we apply the following feedback

$$T_e(t) = f(\bar{y}_t),$$

on a colony of N cells with independently identically distributed initial states,

$$\begin{bmatrix} x_{lo}(0) \\ x_{hi}(0) \end{bmatrix} = \begin{bmatrix} x_{1,0} \\ x_{2,0} \end{bmatrix},$$

then the expected value of $\bar{y}_t$ satisfies

$$\lim_{N\to\infty} E\bar{y}_t^N = \eta(t). \tag{9}$$

Because of space limitation, we refer the reader to [10] for a proof of Theorem 3.1. This theorem provides us with a guarantee that if we design a suitable output feedback law based on the model (3), implementing the feedback law on the colony of N cells will make the expected value of the average output $E\bar{y}_t^N$ as N tends to infinity.

In the remainder of the paper we shall address the following control problem. Given the control architecture in 6, we want to design a controller such that the fraction of induced cells attains a certain level, for example 50%. Declaring the fraction of induced cells as output is equivalent to setting $y_{lo} = 0$ and $y_{hi} = 1$.

Before we proceed to propose a feedback control algorithm, notice that the equilibria of (3) is given by:

$$\frac{d}{dt}\begin{bmatrix} x_{lo} \\ x_{hi} \end{bmatrix} = \begin{bmatrix} -\lambda_1(T_e) & \lambda_2(T_e) \\ \lambda_1(T_e) & -\lambda_2(T_e) \end{bmatrix}\begin{bmatrix} x_{lo} \\ x_{hi} \end{bmatrix} = 0,$$
$$x_{hi} = \frac{\lambda_1(T_e)}{\lambda_2(T_e)}x_{lo},$$

such that the fraction of induced cells at the equilibria is given by

$$n = \frac{\lambda_1(T_e)}{\lambda_1(T_e) + \lambda_2(T_e)}. \tag{10}$$

Figure 7 captures the relation between the transition rates λ_1 and λ_2, and the external concentration of TMG. Notice that λ_1 is a monotonously increasing function of T_e, while λ_2 is monotonously decreasing. Also notice that at $T_e = 1.4$ μM, λ_2 is about 9 times bigger than λ_1, while at $T_e = 2$ μM λ_1 is about 4.5 times bigger than λ_2. Therefore, if the external concentration of TMG is kept at 1.4 μM, the fraction of induced cells is going to converge to around 10%, while if the external concentration of TMG is kept at 2 μM, the fraction of induced cells is going to converge to around 80%.

Based on this knowledge, we propose the following simple on-off feedback control strategy.

On-off controller. The external concentration can assume only two values, $T_{lo} = 1.4$ μM and $T_{hi} = 2$ μM. If the fraction of induced cells, $\bar{y}_t$, is higher than 0.52, then $T_e = T_{lo}$. If $\bar{y}_t$ is less than 0.48 then $T_e = T_{hi}$. If $\bar{y}_t$ is between 0.48 - 0.52, then T_e is kept at its current value. We therefore create a deadzone that will prevent the controller from switching indeterminately around the desired level of $\bar{y}_t = 0.5$.

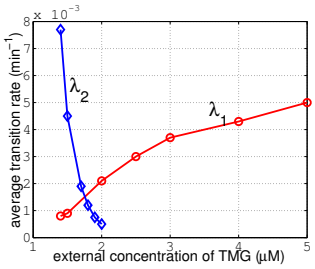

Fig. 7. Relationship between the external concentration of TMG (T) and the average transition rates (induction (λ_1) and deinduction (λ_2)). The points are data obtained from Monte Carlo simulations.

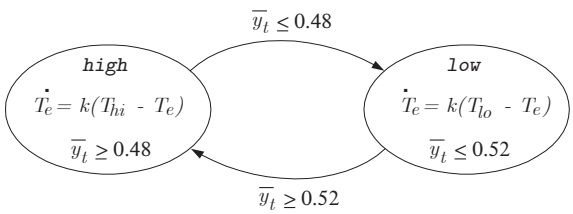

Fig. 8. The hybrid automaton representing the flow controller. The value of k is chosen such that the time constant of the exponential is 10 minutes ($k = 0.1$).

The on-off controller algorithm assumes that the external concentration of TMG can change between T_{lo} and T_{hi} instantaneously. This is not physically feasible if the controller is to be actually implemented. We therefore propose another controller that is more feasible.

Flow controller. The controller that we propose is essentially a hybrid system with two modes of dynamics. The continuous dynamics of the first mode, `high`, is such that the concentration of T_e converges exponentially to T_{hi}, while in the other mode, `low`, T_e converges exponentially to T_{lo}. The scheme of the dynamics is shown in Figure 8. If the fraction of induced cells, $\bar{y}_t$, is higher than 0.52, then the controller is switched to the `low` mode. If $\bar{y}_t$ is less than 0.48 then the controller is switched to the `high` mode. If $\bar{y}_t$ is between 0.48 - 0.52, then T_e is kept at its current value. Again, here we create a deadzone that will prevent the controller from switching indeterminately around the desired level of $\bar{y}_t = 0.5$.

Actuation in the flow controller is indeed physically feasible, since we assume that the change in the external concentration of TMG is done gradually in the order of minutes. Sensing of the fraction of induction is also possible to implement. One way of doing it is by inserting a new gene in the DNA of the bacteria in the *lac* operon. As the genes in the operon gets expressed, the new gene produces *gfp*, a fluorescent protein that emits green light [18]. The amount of emitted light can be used as an indication of the concentration of proteins in the cell, which in turn determines if the cell is classified as induced or uninduced.

Theorem 3.1 provides us with a convergence guarantee if the controller is a static feedback controller, which is not the case with the controllers that we propose. Nevertheless, in the following section we present some simulation results that show that the controllers indeed function as intended. Establishing a stronger convergence proof for dynamic feedback controller is one of the our future research goals.

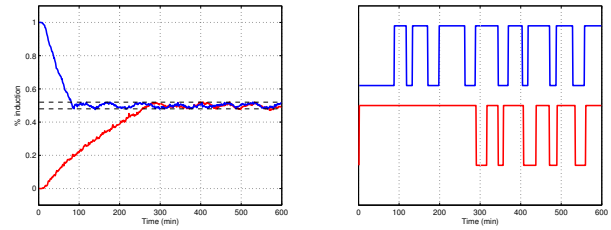

Fig. 9. Simulation results using the on-off controller. Left: Fraction of induced cells. Here the simulation is initiated at two different initial conditions. The dashed lines indicate the deadzone between 48% and 52%. Right: The level of T_e for both simulations. For clarity, the T_e trajectory of the fully induced initial condition is plotted higher.

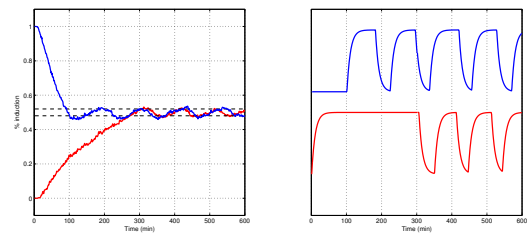

Fig. 10. Simulation results using the flow controller. Left: Fraction of induced cells. Here the simulation is initiated at two different initial conditions. The dashed lines indicate the deadzone between 48% and 52%. Right: The level of T_e for both simulations. For clarity, the T_e trajectory of the fully induced initial condition is plotted higher.

IV. Simulation Results

In this section, we present some simulation results on the application of the two controllers proposed in the previous section to a colony of bacteria with 1000 cells. Figure 9 shows the simulation results with two initial conditions, fully induced and fully uninduced colony, when the on-off controller is used. We can see that the desired fraction of activation of 50% can be attained and maintained within the deadzone. On the right side of Figure 9 we can see the level of T_e switches between T_{lo} and T_{hi} in both simulations.

The same simulations are repeated with the flow controller and the results are shown in Figure 10. We can see that the desired fraction of activation of 50% can be attained and maintained close to the deadzone. The variation of the fraction of induction is larger than that of the on-off controller, which can be expected since the flow controller is more sluggish.

In Figure 11 we can see a dynamic histogram that shows the distribution of the internal concentration of TMG in the cells, when the flow controller is used with fully uninduced initial condition. We can see that initially (at time $t = 0$) the distribution is concentrated at the bottom level. As time progresses, a second cluster, which corresponds to the induced cells appears. After $t = 300$ minutes, we can see that the higher cluster moves up and down because of the modulation of T_e, as it is also shown in Figure 10.

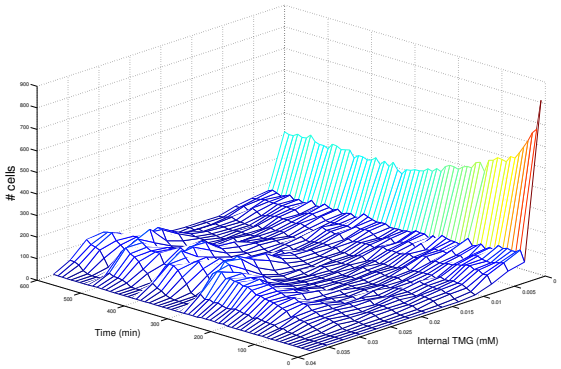

Fig. 11. A dynamic histogram of the distribution of the internal concentration of TMG in the cells. The colony size is 1000 cells. At each time instant, the classes in the histogram are constructed using 15 equal length intervals of internal TMG concentration between 0 and 40 μM.

V. Concluding Remarks

In this paper, we present a comprehensive framework for abstraction and controller design for the lactose regulation system of the *E. coli* bacteria. The abstraction framework is based on the idea that two stable equilibria of the systems can be thought of as states of a continuous time Markov chain, and that the transition rates of the Markov chain can be obtained through Monte Carlo simulations of the actual system.

Because of the simplicity of the abstract model and its demonstrated accuracy in predicting the average behavior of a colony with many cells, we can use the abstract model as a *building block* for designing, for example, a feedback control system for the biological systems. By feedback control here we mean influencing the average behavior of the colony using an environmental variable (external concentration of TMG) as control actuation. In the (future) implementation, the role of the flow controller may be played by another genetically engineered module, such as a toggle switch [5] that results in the production or consumption of the inducer. This may be implemented in the same organism or in another strain which is present in the same bioreactor. Thus, one might be able to construct a network using specifically engineered organisms as circuit elements.

Control of a large number of Markov chains by adjusting the transition rates is quite a versatile framework. For example, it has been studied in [17] for artificial muscle fibers. Considering the generality of the framework, we see potential application of it in other fields such as active materials and networked engineering systems with a large number of autonomous agents such as sensor networks and robotic swarms [19].

Concluding the discussion in this paper, we point out two future goals for our research. The first goal is of theoretical nature, namely, we want to establish a stronger proof of convergence than Theorem 3.1. By stronger we mean we want to have a proof of stochastic convergence of the average output for system with dynamic feedback. The second goal is very much related to the spirit of this paper, as captured in Figure 1. We want to develop the upper part of the hourglass in Figure 1, which means posing and solving more complicated control problems, for example, various types of optimal control problems that are relevant to the biological system, using strategies such as model predictive control.

Acknowledgements. We would like to thank Mahmut Selman Sakar and Marcin Imielinski for valuable discussions during the preparation of this paper, and Prof. Harvey Rubin for pointing us to the *lac* system.

References

[1] A. A. Julius, A. Halasz, V. Kumar, and G. J. Pappas, "Finite state abstraction of a stochastic model of the lactose regulation system of *Escherichia coli*," in *Proc. IEEE Conf. Decision and Control*, (San Diego, USA), 2006.

[2] J. Beckwith, *The lactose operon*, ch. 4, pp. 1444–1452. American Society for Microbiology, 1987.

[3] J. Monod and A. M. Pappenheimer, "The kinetics of the biosynthesis of beta-galactosidase in escherichia coli as a function of growth," *Biochim. et Biophys. Acta*, vol. 9, p. 648, 1952.

[4] A. Novick and M. Weiner, "Enzyme induction as an all-or-none phenomenon," *Proc. Natl. Acad. Sci. USA*, vol. 43, pp. 553–566, 1957.

[5] T. S. Gardner, C. R. Cantor, and J. J. Collins, "Construction of a genetic toggle switch in Escherichia coli," *Nature*, vol. 403, pp. 339–342, 2000.

[6] E. Andrianantoandro, S. Basu, D. K. Karig, and R. Weiss, "Synthetic biology: new engineering rules for an emerging discipline," *Molecular Systems Biology*, vol. 2, p. 2006.0028, 2006.

[7] D. Ro, E. M. Paradise, M. Ouellet, and J. Keasling et.al, "Production of the antimalarial drug precursor artemisinic acid in engineered yeast," *Nature*, vol. 440, no. 7086, pp. 940–943.

[8] "Genomics GTL: Systems biology for energy and environment." http://genomicsgtl.energy.gov/.

[9] N. Yildirim and M. C. Mackey, "Feedback regulation in the lactose operon: A mathematical modeling study and comparison with experimental data," *Biophys. J.*, pp. 2841–2851, 2003.

[10] A. A. Julius, A. Halasz, M. S. Sakar, V. Kumar, H. Rubin, and G. J. Pappas, "Controlling biological systems: the lactose regulation system of Escherichia coli." manuscript submitted to the IEEE Trans. Automatic Control and IEEE Trans. Circuits and Systems joint special issue on Systems Biology, 2007.

[11] M. Samoilov, S. Plyasunov, and A. P. Arkin, "Stochastic amplification and signaling in enzymatic futile cycles through noise-induced bistability with oscillations," *Proc. National Academy of Science*, vol. 102, no. 7, pp. 2310–2315, 2005.

[12] C. V. Rao, D. M. Wolf, and A. P. Arkin, "Control, exploitation and tolerance of intracellular noise," *Nature*, vol. 420, no. 6912, pp. 231–237, 2002.

[13] J. M. Pedraza and A. van Oudenaarden, "Noise propagation in gene networks," *Science*, vol. 307, no. 5717, pp. 1965–1969, 2005.

[14] D. T. Gillespie, "A general method for numerically simulating the stochastic time evolution of coupled chemical reactions," *Journal of Computational Physics*, vol. 22, pp. 403–434, 1976.

[15] T. B. Kepler and T. C. Elston, "Stochasticity in transcriptional regulation: origins, consequences, and mathematical representations," *Biophysical Journal*, vol. 81, pp. 3116 – 3136, December 2001.

[16] C. G. Cassandras and S. Lafortune, *Introduction to Discrete Event Systems*. Kluwer, 1999.

[17] J. Ueda, L. Odhnar, and H. Asada, "A broadcast-probability approach to the control of vast DOF cellular actuators," in *Proc. IEEE Int. Conf. Robotics and Automation*, (Florida, USA), pp. 1456–1461, 2006.

[18] E. M. Ozbudak, M. Thattal, H. N. Lim, B. I. Shraiman, and A. van Oudenaarden, "Multistability in the lactose utilization network of *Escherichia coli*," *Nature*, vol. 427, pp. 737 – 740, February 2004.

[19] S. Berman, A. Halasz, V. Kumar, and S. Pratt, "Algorithms for the analysis and synthesis of a bio-inspired swarm robotic system," 2006. submitted to the SAB'06 Swarms Robotics Workshop.