University of Pennsylvania
ScholarlyCommons

Departmental Papers (MEAM)

Department of Mechanical Engineering & Applied Mechanics

12-15-2006

# Finite state abstraction of a stochastic model of the lactose regulation system of *Escherichia coli*

Agung Julius
*University of Pennsylvania*

Ádám Halász
*University of Pennsylvania*

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

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

Copyright 2006 IEEE. Reprinted from *Proceedings of the 45th IEEE Conference on Decision & Control 2006*.

This material is posted here with permission of the IEEE. Such permission of the IEEE does not in any way imply IEEE endorsement of any of the University of Pennsylvania's products or services. Internal or 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 must be obtained from the IEEE by writing to pubs-permissions@ieee.org. By choosing to view this document, you agree to all provisions of the copyright laws protecting it.

This paper is posted at ScholarlyCommons. http://repository.upenn.edu/meam_papers/151
For more information, please contact repository@pobox.upenn.edu.

Proceedings of the 45th IEEE Conference on Decision & Control
Manchester Grand Hyatt Hotel
San Diego, CA, USA, December 13-15, 2006



# Finite state abstraction of a stochastic model of the lactose regulation system of *Escherichia coli*

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

***Abstract*— This paper focuses on the lactose regulation system in *Escherichia coli* bacteria, one of the most extensively studied examples of positive feedback in a naturally occurring gene network. State-of-the-art nonlinear dynamical system models predict a bi-stability phenomenon that is confirmed in experiments. However, such deterministic models fail to explain experimental observations of spontaneous transition between the two stable states in the system and the simultaneous occurrence of both steady states in a population of cells. In this paper, we propose a stochastic model that explains this phenomenon. Furthermore, we also extract a coarser two-state continuous-time Markov chain as a higher level abstraction of this model, and show that macroscopic properties are retained in the abstraction.**

## I. Introduction

Systems biology is an effort to use systems theoretic thinking, coupled with sophisticated computational methods, to gain better understanding of the functions of biological systems. Recent work by Sontag and collaborators [1] on monotone systems opens the possibility of representing complicated reaction networks as much simpler networks of interconnected monotone subsystems. Similarity between engineering design principles and the organization of a biological system has been investigated in [2]. Hybrid systems theory has been applied to multi-cellular networks by Ghosh and Tomlin [3], and genetic networks by Batt *et al* [4] as well as [5]. Stochastic hybrid systems were used to study genetic networks by Hespanha and Singh [6]. Important progress on the biochemical mechanism of domineering nonautonomy in the *Drosophila* wing has been made using optimization to tune the parameters of a differential equation based model [7].

In this paper, we present a stochastic hybrid model for the lactose regulation system in the *Escherichia coli* bacteria. The lactose operon [8] 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 ($\beta$-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 [9]. Novick and Weiner [10] discovered bistability at the level of individual cells by studying the expression of $\beta$-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 population heterogeneity was interpreted by Novick and Weiner as a result of a bistability of the gene expression mechanism of individual cells combined with stochastic fluctuations inherent to bio-molecular processes involving few molecules. A possibly related and yet unexplained phenomenon, discovered by Knorre [11] is that of transient oscillations (of a period on the order of an hour) of $\beta$-galactosidase activity upon diauxic shift from glucose to lactose medium (and vice versa, [12]).

The notion of autocatalytic gene expression in the *lac* system has motivated significant work in the context of dynamical models, starting from the early sixties [13]. It was well known that positive feedback and delays can result in multiple stable equilibrium points and limit cycles, as is the case in many other biological models. Different models were proposed to study the conditions for stability, the possibility of oscillations, and the effect of time delays due to transcription and translation in the *lac* system. This direction of research gradually led to more detailed dynamical models, which explicitly incorporate all relevant biochemical processes along with experimentally motivated kinetic constant values. The work of Yildirim and Mackey [14] is an example of this new generation of experimentally grounded dynamical modeling.

While much of the modeling of biochemical reactions is based on deterministic models with ordinary differential equations, it is well known that these models do not satisfactorily explain the behavior of systems with very low concentrations in which the continuum model is not applicable. Methods for stochastic simulations of biochemical reactions have been developed [15], [16], [17]. Autocatalytic gene expression has also been studied in the stochastic context, for example [18]. Recently it has been recognized that stochastic phenomena may have a crucial role in the fate of individual cells [19]. Multistability and stochastic transitions between equilibrium states were found to have a role in the phenomenon of bacterial persistence [20] and more generally are seen as an evolutionary strategy for survival in varying environments [21].

The motivation for a detailed stochastic model of the lactose induction mechanism comes from the work of Ozbudak *et al* [22] in which they used fluorescent labelling techniques allowing *in vivo* observation of individual cells. They showed a distinct bimodal distribution of the activity of the lac

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

operon in individual cells. Thus, population-averaged continuous changes with inducer concentration reflect changes in the relative size of the induced and uninduced populations, reinforcing the conclusions of Novick and Weiner [10].

The experimental results of Ozbudak *et al* were well summarized by a simple empirical model. A more elaborate model, highlighting the role of individual biochemical processes, is desirable. While Yilidirm and Mackey [14] provide a detailed account of the underlying processes, characterized by parameters that can be measured independently, their model can only predict the *bulk* behavior of a population of cells. While it predicts bistability and explains the hysteretic switching between steady states, it does not explain the bimodal distribution of $\beta$-galactosidase [10] and lactose operon activity [22] that have been experimentally observed.

In this paper we propose a stochastic hybrid model for the lactose regulation system. Stochasticity in the system naturally arises due to the low copy numbers of molecules involved in the reaction within the cell [23], [2]. In this situation, a model where the reactions are viewed as discrete Poisson random processes is more accurate than a deterministic one [15]. However, due to the incurred computational cost, we choose to use a model in which only those reactants with small copy numbers are modeled as discrete quantities, while the others are modeled as continuous concentrations.

We show that our stochastic hybrid model is able to reproduce the spontaneous transitions that are impossible to capture in the deterministic model. Further, the steady state behavior of a bulk of cells simulated with our model demonstrates agreement with the predicted equilibria of the Yildirim-Mackey model. Furthermore, we extract a finite state abstraction of the hybrid stochastic model, which is structured as a two-state continuous time Markov chain [24]. We demonstrate that despite of the simplicity of the abstraction, it can describe the average (macroscopic) behavior of a colony of *E. coli* bacteria, each of which is simulated with the hybrid stochastic model.

The remaining parts of this paper are organized as follows. The next section is devoted to the deterministic model, as given by Yildirim-Mackey [14]. In Section 3 we discuss the proposed hybrid stochastic model of the lactose regulation system. In Section 4 we describe the numerical simulation algorithm that we use to simulate the hybrid stochastic model. The finite state abstraction of the hybrid stochastic model using a two-state continuous time Markov chain is explained in Section 5. Finally, Section 6 contains some conclusions and directions for future work.

## II. The deterministic model

Our starting point is the time-delayed ordinary differential equation (ODE) model proposed in [14], describing the dynamics of the concentration of five substances that are involved in the lactose metabolism and its regulation. Briefly, the mRNA $(M)$ transcribed from the lactose operon is translated into three different gene products, among them permease $(P)$ and $\beta$-galactosidase $(B)$. Permease facilitates the influx of lactose $(L)$ from the exterior and also an opposing process, equilibrating the concentration of lactose inside the cell with the external lactose. The enzyme $\beta$-galactosidase has a dual role; it converts lactose to allolactose $(A)$ and also converts allolactose further to glucose and galactose. The control loop is closed by the effect of allolactose $(A)$ on the transcription of the lac operon. This complicated relationship involves substances not explicitely 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 are as follows:

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

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

$$\frac{dA}{dt} = \alpha_A B \frac{L}{K_L + L} - \beta_A B \frac{A}{K_A + A} - \tilde{\gamma}_A A, \quad (1c)$$

$$\frac{dL}{dt} = \alpha_L P \frac{L_e}{K_{L_e} + L_e} - \beta_L P \frac{L}{K_{L_1} + L} - \beta_{L_2} B \frac{L}{K_{L_2} + L} - \tilde{\gamma}_L L, \quad (1d)$$

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

The symbol $L_e$ in equation (1d) signifies the external lactose concentration. If the system is to be viewed as an input-state system, then $L_e$ can be thought of as an input to the system, while the other five concentrations are the state variables. The other symbols in the equation are constant parameters, given by the following table. 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 $\tau_B$.

| | Value | Unit | | Value | Unit |
|---|---|---|---|---|---|
| $n$ | 2 | | $K_{L_e}$ | 0.26 | mM |
| $\gamma_M$ | 0.411 | $\text{min}^{-1}$ | $\gamma_B$ | $8.33 \cdot 10^{-4}$ | $\text{min}^{-1}$ |
| $\gamma_A$ | 0.52 | $\text{min}^{-1}$ | $\Gamma_0$ | $7.25 \cdot 10^{-7}$ | mM/min |
| $K$ | 7200 | | $\alpha_M$ | $9.97 \cdot 10^{-4}$ | mM/min |
| $\tau_B$ | 2.0 | min | $\alpha_A$ | $1.76 \cdot 10^{4}$ | $\text{min}^{-1}$ |
| $K_{L_1}$ | 1.81 | mM | $\alpha_B$ | $1.66 \cdot 10^{-2}$ | $\text{min}^{-1}$ |
| $K_A$ | 1.95 | mM | $\beta_A$ | $2.15 \cdot 10^{4}$ | $\text{min}^{-1}$ |
| $\tau_M$ | 0.1 | min | $K_L$ | $9.7 \cdot 10^{-1}$ | mM |
| $\gamma_L$ | 0.0 | $\text{min}^{-1}$ | $\gamma_P$ | 0.65 | $\text{min}^{-1}$ |
| $\alpha_L$ | 2880 | $\text{min}^{-1}$ | $\alpha_P$ | 10.0 | $\text{min}^{-1}$ |
| $\tau_P$ | 0.83 | min | $\beta_{L_1}$ | $2.65 \cdot 10^{3}$ | $\text{min}^{-1}$ |
| $\mu$ | $3.47 \cdot 10^{-2}$ | $\text{min}^{-1}$ | $K_1$ | $2.52 \cdot 10^{4}$ | $(\text{mM})^{-2}$ |
| $K_{L_2}$ | $9.7 \cdot 10^{-1}$ | mM | $\beta_{L_2}$ | 2880 | $\text{min}^{-1}$ |

together with the following relations

$$\tilde{\gamma}_M = \gamma_M + \mu, \tilde{\gamma}_B = \gamma_B + \mu, \quad (2)$$

$$\tilde{\gamma}_A = \gamma_A + \mu, \tilde{\gamma}_P = \gamma_P + \mu. \quad , \quad (3)$$

where $\mu$ is the growth rate.

When the value of $L_e$ is maintained between 0.03 - 0.06 mM, the system has three equilibria. Two of these equilibria are stable, giving rise to bistability of the system. Also, varying the value of $L_e$ causes a hysteresis behavior. See Figure 1 for the illustration.

The mathematical model that we have in (1) is deterministic. It correctly reproduces experimental results [11], [26] on the timecourse of an upward shift of enzymatic activity when the initially uninduced system is placed in a high concentration of external lactose. However, even at steady state, a bimodal distribution of $\beta$-galactosidase [10]

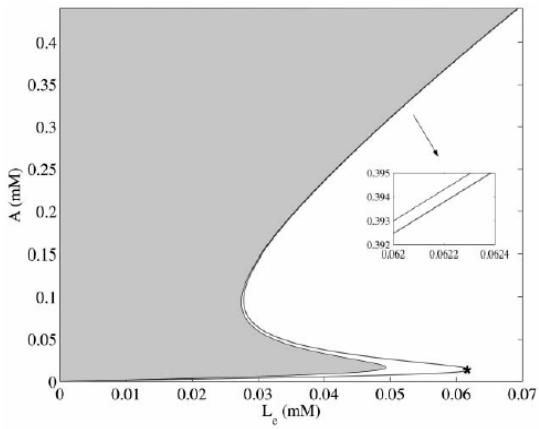


Fig. 1. The equilibria of the system given by (1), taken from [25]. The middle range of $L_e$ has three branches of equilibria.

as well as operon activity [22] have been observed. In fact, it is suggested that the bimodal distribution varies depending on the concentration of external lactose $L_e$, implying quasi-instantaneous transitions between stable equilibria.

The fact that the model of [25] is not consistent with the behavior observed at the level of individual cells is not surprising, since some of the model parameters have been explicitly chosen to reproduce the *macroscopic* behavior observed on a large number of cells. On the other hand, extensive biochemical work has identified the processes included in the model (1) as having a role in the observed phenomenology of the lac operon. In attempting to build a correct *microscopic* model, we use the structure of the Yildirim-Mackey model as a starting point, and investigate its behavior when stochastic effects are taken into account.

## III. The stochastic hybrid model

There are several sources of stochasticity in the biochemistry of individual cells [27]. In this paper we will focus on one major source, intrinsic noise generated by *low copy numbers* of molecules. The deterministic ODE model relies on continuously varying concentrations, which is a good approximation when the substances are available in huge molecule numbers. If we consider chemical reactions within a cell, whose volume is in the order of $10^{-16}\,\mathrm{l}$ [28], the number of molecules involved in the reaction may not be too large. This is especially the case if the concentration of the chemical substances is low.

Chemical reactions, at the microscopical level, amount to creation and breaking up of chemical molecules. These processes can be modeled as Poisson random processes [15], [29], whose rates depend on the state of the system, i.e. the number of molecules in the reaction. In fact, the reaction rates given by the ODE can be considered as the rates of the Poisson processes. This is not the only way to introduce stochasticity to the system. Another approach is to use an ordinary differential model perturbed by stochastic noise [30]. However, we argue that modeling the chemical reactions as Poisson processes is more physically founded. There has also been previous work where stochasticity is introduced by modeling chemical reaction as Poisson processes [31]. However, the underlying reaction model is based on empirical observation rather than physical modeling as in [25].

We develop a hybrid stochastic model for the system. The model is based on the idea that the messenger RNA $(M)$ and the $\beta$-galactosidase $(B)$ are expressed as molecule counts that evolve following some Poisson processes, while the other three substances, allolactose $(A)$, lactose $(L)$, and permease $(P)$, are expressed as chemical concentrations that evolve following deterministic ODE. A similar approach, i.e. part stochastic and part deterministic simulation for chemical processes is reported in [23]. The reason behind this idea is that a fully stochastic model is computationally expensive, while a hybrid model already demonstrates the stochastic noise that is lacking in the deterministic model.

We are interested in the phenomenology of a model with the structure of that in [25], incorporating the presence of realistic level of noise. The relative importance of stochastic fluctuations of one concentration will be the largest for the species with the lowest concentrations. We choose to discretize $M$ and $B$, whose concentrations at the uninduced steady state at a $L_e = 0.04$ mM correspond to values on the order of one molecule per cell.

We define the conversion constant $C_N$ as

$$\begin{aligned} C_N &= 10^{-16}\,\mathrm{l}\cdot 6.022\cdot 10^{23}\frac{\text{molecules}}{\text{mole}}\cdot 10^{-3}\frac{\mathrm{M}}{\mathrm{mM}}, \\ &= 6.022\cdot 10^{4}\frac{\text{molecules}}{\mathrm{mM}}. \end{aligned}$$

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

$$dM_t = d\hat{M}_t - d\tilde{M}_t, \tag{4a}$$

$$dB_t = d\hat{B}_t - d\tilde{B}_t, \tag{4b}$$

$$\frac{dA_t}{dt} = \frac{L\alpha_A}{K_L + L}\frac{B_t}{C_N} - \frac{\beta_A A_t}{K_A + A_t}\frac{B_t}{C_N} - \tilde{\gamma}_A A_t, \tag{4c}$$

$$\frac{dL_t}{dt} = \frac{L_e\alpha_L P_t}{K_{L_e} + L_e} - \frac{\beta_L P_t L_t}{K_{L_1} + L_t} - \frac{\beta_{L_2} B_t L_t}{K_{L_2} + L_t} - \tilde{\gamma}_L L_t, \tag{4d}$$

$$\frac{dP_t}{dt} = \alpha_P e^{-\mu(\tau_P+\tau_B)}\frac{M_{(t-\tau_P-\tau_B)}}{C_N} - \tilde{\gamma}_P P_t. \tag{4e}$$

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 $\beta$ - galactosidase molecules, respectively. The rates of these processes are state dependent, and are given as follows.

$$\lambda_{\hat{M}}(t) = C_N\left[\alpha_M\frac{1 + K_1(e^{-\mu\tau_M}A_{(t-\tau_M)})^n}{K + K_1(e^{-\mu\tau_M}A_{(t-\tau_M)})^n} + \Gamma_0\right], \tag{5a}$$

$$\lambda_{\tilde{M}}(t) = \tilde{\gamma}_M M_t, \tag{5b}$$

$$\lambda_{\hat{B}}(t) = \alpha_B e^{-\mu\tau_B} M_{(t-\tau_B)}, \tag{5c}$$

$$\lambda_{\tilde{B}}(t) = \tilde{\gamma}_B B_t. \tag{5d}$$

## IV. The Stochastic Simulation

We simulate the stochastic model (4) using a numerical scheme similar to the explicit tau-leaping method for Gillespie simulation [16], [32]. We pick a constant integration step $\delta$ and discretize equation (4). We use the following notation

$$M[k] := M_{t=k\delta}, B[k] := B_{t=k\delta}, P[k] := P_{t=k\delta},$$
$$A[k] := A_{t=k\delta}, L[k] := L_{t=k\delta}.$$

Since we also have to discretize the time delay, we define

$$k_M := \left\lfloor \frac{\tau_M}{\delta} \right\rfloor, k_B = \left\lfloor \frac{\tau_B}{\delta} \right\rfloor, k_P := \left\lfloor \frac{\tau_B + \tau_P}{\delta} \right\rfloor, \quad (6)$$

where

$$\forall x \in \mathbb{R}_+, \lfloor x \rfloor := \max\{n \in \mathbb{Z} \mid n \leq x\}.$$

$$M[k+1] = M[k] + \Delta\hat{M}[k] - \Delta\tilde{M}[k], \quad (7)$$
$$B[k+1] = B[k] + \Delta\hat{B}[k] - \Delta\tilde{B}[k], \quad (8)$$
$$A[k+1] = A[k] + \delta\left[\frac{L\alpha_A}{(K_L + L)}\frac{B[k]}{C_N} - \frac{\beta_A A[k]}{(K_A + A[k])}\frac{B[k]}{C_N} - \tilde{\gamma}_A A[k]\right], \quad (9)$$
$$L[k+1] = L[k] + \delta\left[\frac{\alpha_L L_e[k]P[k]}{K_{L_e} + L_e[k]} - \frac{\beta_L P[k]L[k]}{K_{L_1} + L[k]} - \frac{\beta_{L_2} B[k]L[k]}{K_{L_2} + L[k]} - \tilde{\gamma}_L L[k]\right], \quad (10)$$
$$P[k+1] = P[k] + \delta\alpha_P e^{-\mu(\tau_P + \tau_B)}\frac{M[k-k_P]}{C_N} - \delta\tilde{\gamma}_P P[k]. \quad (11)$$

The terms $\Delta\hat{M}[k]$, $\Delta\tilde{M}[k]$, $\Delta\hat{B}[k]$, $\Delta\tilde{B}[k]$ are the approximation of the increments of the Poisson processes in (4). For example,

$$\Delta\hat{M}[k] \approx \int_{k\delta}^{(k+1)\delta} d\hat{M}_t. \quad (12)$$

In the approximation, these terms are modeled as Poisson random variables with expectations

$$E\left[\Delta\hat{M}[k]\right] = \delta C_N \alpha_M \frac{1 + K_1(e^{-\mu\tau_M}A[k-k_M])^n}{K + K_1(e^{-\mu\tau_M}A[k-k_M])^n} + \delta C_N \Gamma_0, \quad (13)$$
$$E\left[\Delta\tilde{M}[k]\right] = \delta\tilde{\gamma}_M M[k], \quad (14)$$
$$E\left[\Delta\hat{B}[k]\right] = \delta\alpha_B e^{-\mu\tau_B} M[k-k_B], \quad (15)$$
$$E\left[\Delta\tilde{B}[k]\right] = \delta\tilde{\gamma}_B B[k]. \quad (16)$$

One run of the simulation is shown in Figure 2. Here we begin with the initial condition

$$M[0] = 2 \text{ molecules}, B[0] = 2 \text{ molecules},$$
$$P[0] = 2 \cdot 10^{-4} \text{ mM}, A[0] = 0.04 \text{ mM}, L[0] = 0.35 \text{ mM}.$$

Initially we set the level of external lactose concentration at $L_e = 2 \cdot 10^{-2}$mM, which will drive the system to the stable low state (see Figure 1). After 200 minutes, we set the level of external lactose concentration at $L_e = 5.5 \cdot 10^{-2}$mM. This will bring the system to the bistability zone. The deterministic model (1) predicts that the system will remain in the stable low state. Our simulation shows that indeed this is the case, however, around 500 minutes later the system is spontaneously induced to the high state.

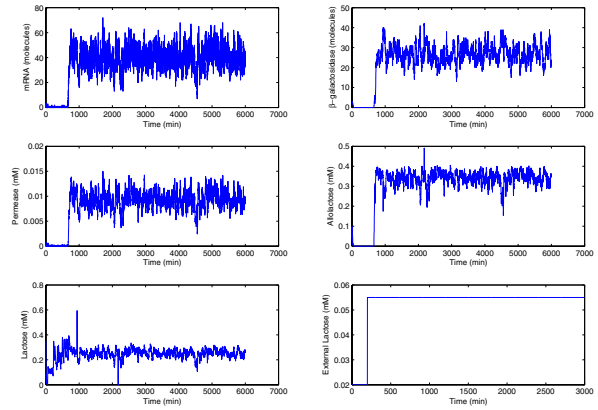


Fig. 2. One simulation result. In this plot, we observe that spontaneous induction occurs 500 minutes after the level of lactose is increased.

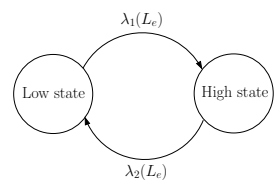


Fig. 3. The two-state continuous time Markov chain model.

## V. Finite State Abstraction of the Stochastic Model

In this section we discuss a finite state abstraction of the stochastic model (4). Our goal is to construct an abstraction of the stochastic model that is simple enough to allow for fast computation. This is particularly desirable, for example, when we want to simulate the behavior of a colony of bacteria. Without the abstraction, we would have to run multiple copies of the stochastic simulation described in the previous section which can be computationally expensive.

The abstraction that we choose is a two-state continuous time Markov chain [24]. The states of the Markov chain correspond to the low and high stable equilibria of the systems. The rates of switching between the two states are given as a function of the external lactose concentration $L_e$. See Figure 3 for a diagram of the system[1].

Although this model is seemingly very simple, we have a strong reason behind its adoption. We run the simulation of the full model (given in the previous section) 100 times, to simulate a colony of 100 cells. The macroscopic behavior of the colony, which is computed as the average across the

[1]A method for approximately abstracting stochastic hybrid systems is presented in [33]. The nonlinear dynamics in this paper makes the implementation of the method computationally challenging. However, it is noteworthy that there are more systematic ways of abstracting stochastic hybrid systems.

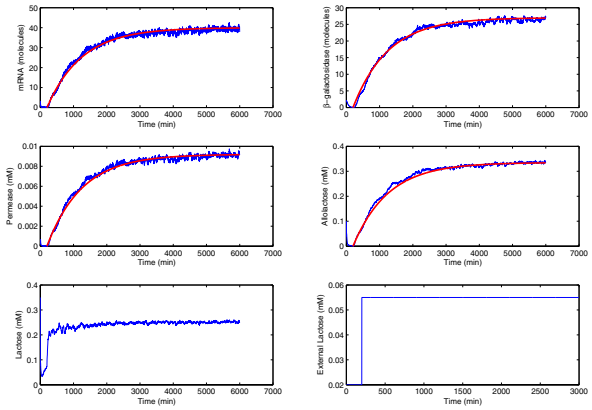


Fig. 4. The macroscopic behavior of the colony with 100 bacteria. The plotted values are taken as the average across 100 samples. The exponential curve is plotted to show that the macroscopic behavior can be fitted quite well with a first order dynamics.

100 samples are plotted in Figure 4. We observe that the behavior can be closely matched with a first order behavior (an exponential curve).

Given the continuous time Markov chain model as in Figure 3, we can compute the probability distribution of the states as follows. Define $p_{\rm lo}(t)$ and $p_{\rm hi}(t)$ as the probability of finding the system at time $t$ in the low and high state respectively. The probability distribution satisfies the following differential equation.

$$\frac{d}{dt}\begin{bmatrix} p_{\rm lo} \\ p_{\rm hi} \end{bmatrix} = \begin{bmatrix} -\lambda_1(L_e) & \lambda_2(L_e) \\ \lambda_1(L_e) & -\lambda_2(L_e) \end{bmatrix}\begin{bmatrix} p_{\rm lo} \\ p_{\rm hi} \end{bmatrix}. \quad (17)$$

For the value of $L_e$ that we use in the simulation, $L_e = 5.5 \cdot 10^{-2}\mathrm{mM}$, the transition from the low state to the high state is much more likely to happen than the opposite way. This can be explained by referring to Figure 1. At this value of $L_e$, in order to transit from high state to low state, the system has to overcome a much wider potential barrier than the opposite way. This results in $\lambda_2(L_2) \approx 0$. Thus, the solution to (17), assuming that all the cells start in the low state ($p_{\rm lo}(0) = 1$) is given by (hereafter we do not write explicitly the dependance of $\lambda_1$ on $L_e$)

$$p_{\rm lo}(t) = e^{-\lambda_1 t}, p_{\rm hi}(t) = 1 - e^{-\lambda_1 t}. \quad (18)$$

Let the concentration of one of the substances, the allolactose ($A$), in the low state and high state be denoted by $A_{\rm lo}$ and $A_{\rm hi}$ respectively. Further, let the random process $A_t^{\rm av}$ be the average of the value of the allolactose concentrations in the colony of 100 cells. Denote the concentration of allolactose in the $i$-th cell by $A_t^i$. Since the processes $\{A_t^i\}$, $1 \le i \le 100$ , are mutually independent and identically distributed, $A_t^{\rm av}$ is an unbiased estimator for $E[A_t^i]$, which is given as follows.

$$E[A_t^i] = A_{\rm lo}e^{-\lambda_1 t} + A_{\rm hi}\left(1 - e^{-\lambda_1 t}\right), \quad (19)$$

which tells us that it converges exponentially from $A_{\rm lo}$ to $A_{\rm hi}$. This is what we observe in Figure 4.

Furthermore, the rate of the exponential curve, $\lambda_1$, 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 100 samples and use its value to compute the exponent of the curves in Figure 4. We can observe that the fit is good.

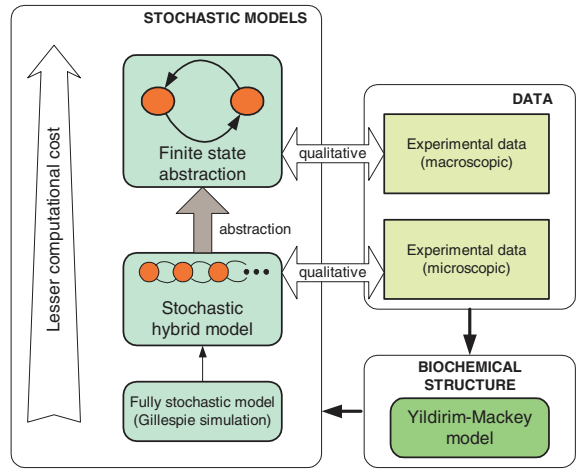


Fig. 5. A block diagram summarizing our approach. Our stochastic hybrid model and its finite state abstraction qualitatively reproduce experimental observations. These stochastic models are based on the biochemically founded deterministic model of Yildirim-Mackey.

## VI. Conclusions and Future Work

In this paper we present a hybrid stochastic model for the lactose regulation system in *E. coli* bacteria. The model is based on the assumption that when the number of molecules involved in a reaction is low, the reaction can be modeled as a Poisson random process, whose rate depends on the number of molecules. We construct a stochastic simulation for our model and show that we can reproduce some stochastic phenomena that are absent in the deterministic model.

Further, we construct a finite state abstraction of the stochastic hybrid model. The abstraction is in the form of a two-state continuous time Markov chain, with variable rate. We show that this is a good abstraction, based on the fact that the macroscopical behavior of the system is preserved.

Our contribution in this paper can be summarized as Figure 5. Our stochastic hybrid model matches experimental data qualitatively, in the sense that it reproduces the spontaneous transitions between stable equilibria in the bistable zone. This is not possible in the deterministic model [14]. The proposed finite state abstraction is faithful to the stochastic hybrid model, in the sense that it exhibits a first order behavior, which is observed in the macroscopic simulation of the stochastic hybrid model. Moreover, a similar first order behavior is also observed in experiments, as reported in [10], [11], [26].

Our model does not yet reproduce the experimental data quantitatively. The average switching time computed in the simulation is one order of magnitude higher than that consistent with experiments. We hypothesize the following explanation for this gap.

As it stands now, the kinetics of our stochastic hybrid model are identical to that of the deterministic Yildirim-Mackey model [14]. This model describes the *bulk* behavior of a population of cells. This is different from the behavior of an individual cell. Consider an experiment where a population of uninduced cells is placed in a medium with high inducer concentration. As the results of [10], [11],

[26] show, the bulk activity will gradually increase over a timescale of 1-3 hours. The Yildirim-Mackey model correctly reproduces this phenomenon, with a comparable rise time. However, this can not be correct for individual cells, which must be induced significantly faster, since the global increase of activity reflects increasing numbers of fully induced cells. An order-of-magnitude estimate of the *individual* induction time is the lag between the initial nutritional shift and the beginning of the quasi-linear increase in bulk activity observed experimentally which is close to 5-10 minutes.

Based on this hypothesis, we consider reconciling the biochemically founded structure of the Yildirim-Mackey model with the phenomenology of induction at the single cell level (as observed in [10], [11], [26] as well as recent results in [22]), as an interesting future research direction. Our strategy is to (1) adjust some of the model parameters to reduce deterministic transition time (2) identify and incorporate other sources of noise into the stochastic hybrid model.

The simple finite state model obtained in this paper can be thought of as a building brick for system design and synthesis in the higher level. Suppose that one is interested manipulating the macroscopic behavior a colony of bacteria, then each cell can be modelled as a finite state system. A control architecture that is oriented for macroscopic behavior and robust against individual variation can be derived. We are also interested in studying whether such design paradigm is relevant for networked engineering systems such as sensor networks and robotic swarms [35].

**Acknowledgements:** The authors would like to thank Mahmut Selman Sakar as well as all members of the Biocomputation group at the University of Pennsylvania.

## REFERENCES


[1] G. A. Enciso, H. L. Smith, and E. D. Sontag, "Non-monotone systems decomposable into monotone systems with negative feedback," *J. of Differential Equations*, no. 224, pp. 205–227, 2006.

[2] H. El-Samad, H. Kurata, J. Doyle, C. Gross, and M. Khammash, "Surviving heat shock: control strategies for robustness and performance," *PNAS*, vol. 102, no. 8, pp. 2736–2741, 2005.

[3] R. Ghosh and C. Tomlin, "Lateral inhibition through delta-notch signaling: A piecewise affine hybrid model," in *Hybrid Systems: Computation and Control*, ser. LNCS, vol. 2034. Springer, 2001, pp. 232–246.

[4] G. Batt, D. Ropers, H. de Jong, J. Geiselmann, M. Page, and D. Schneider, "Qualitative analysis and verification of hybrid models of genetic regulatory networks: nutritional stress response," in *Hybrid Systems: Computation and Control*, ser. LNCS. Springer, 2005, pp. 134–150.

[5] C. Belta, P. Finin, L. Habets, Á. Halász, M. Imielinski, V. Kumar, and H. Rubin, "Understanding the bacterial stringent response using reachability analysis of hybrid systems," in *Hybrid Systems: Computation and Control*, ser. LNCS, vol. 2993. Springer, 2004, pp. 111–125.

[6] A. Singh and J. Hespanha, "Models for gene regulatory networks using polynomial stochastic hybrid systems," in *Proceedings of the 44th Conference on Decision and Control*, Seville, Spain, December 2005.

[7] K. Amonlirdviman, N. A. Khare, D. R. Tree, W. S. Chen, J. D. Axelrod, and C. J. Tomlin, "Mathematical modeling of planar cell polarity to understand domineering nonautonomy," *Science*, vol. 307, pp. 423–426, 2005.

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

[9] 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.

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

[11] W. A. Knorre, "Oscillation of the rate of synthesis of beta-galactosidase in escherichia coli ml 30 and ml 308," *Biochem. Biophys. Res. Commun,*, 1968.

[12] N. P. Pih and P. Dhurjati, "Oscillatory behavior of beta-galactosidase enzyme activity in escherichia coli during perturbed batch experiments," *Biotech. Bioeng.*, vol. 29, pp. 292–296, 1987.

[13] J. M. Mahaffy and E. Simeonov, "Stability analysis for a mathematical model of the lac operon," *Q. Appl. Math.*, vol. 57, pp. 37–53, 1999.

[14] 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.

[15] 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.

[16] ——, "Approximate accelerated stochastic simulation of chemically reacting systems," *J. Chem. Phys.*, vol. 115, no. 4, p. 1716, 2001.

[17] D. Gillespie and L. Petzold, "Improved leap-size selection for accelerated stochastic simulation," *J. Chemical Physics*, vol. 119, pp. 8229–8234, 2003.

[18] T. B. Kepler and T. C. Elston, "Stochasticity in transcriptional regulation: Origins, consequences, and mathematical representations," *Biophys. J.*, vol. 81, pp. 3116–3136, 2001.

[19] A. Arkin, J. Ross, and H. H. McAdams, "Stochastic kinetic analysis of developmental pathway bifurcation on phage lambda-infected escherichia coli cells," *Genetics*, vol. 149, pp. 1633–1648, 1998.

[20] N. Q. Balaban, J. Merrin, R. Chait, L. Kowalik, and S. Leibler, "Bacterial persistence as a phenotypic switch," *Science*, vol. 305, pp. 1622–1625, 2004.

[21] E. Kussell and S. Leibler, "Phenotypic diversity, population growth, and information in fluctuating environments," *Science*, vol. 309, p. 2075, 2005.

[22] E. M. Ozbudak, M. Thattai, H. N. Lim, B. I. Shraiman, and A. van Oudenarden, "Multistability in the lactose utilization network of escherichia coli," *Nature*, vol. 427, pp. 737–740, 2004.

[23] N. A. Neogi, "Dynamic partitioning of large discrete event biological systems for hybrid simulation and analysis," in *Hybrid Systems: Computation and Control*, ser. LNCS, vol. 2993. Springer, 2004, pp. 463 – 476.

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

[25] N. Yildirim and M. C. Mackey, "Feedback regulation in the lactose operon: a mathematical modeling study and comparison with experimental data," *Biophysical Journal*, vol. 84, pp. 2841 – 2851, May 2003.

[26] S. Pestka, B. L. Daugherty, V. Jung, K. Hotta, and R. K. Pestka, "Anti-mrna: Specific inhibition of translation of single mrna molecules," *Proc. Natl. Acad. Sci. USA*, vol. 81, pp. 7525–7528, 1984.

[27] M. B. Elowitz, A. J. Levine, E. D. Siggia, and P. S. Swain, "Stochastic gene expression in a single cell," *Science*, vol. 297, pp. 1183–1186, 2002.

[28] R. A. Cox, "Quantitative relationships for specific growth rates and macromolecular compositions of mycobacterium tuberculosis, streptomyces coelicolor a3(2) and escherichia coli b/r: an intergrative theoretical approach," *Microbiology*, vol. 150, pp. 1413–1426, 2004.

[29] 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.

[30] T. Tian and K. Burrage, "Stochastic models for regulatory networks of the genetic toggle switch," *Proc. National Academy of Science*, vol. 103, no. 22, pp. 8372 – 8377, 2006.

[31] J. M. G. Vilar, C. C. Guet, and S. Leibler, "Modeling network dynamics: the lac operon, a case study," *J. Cell Biology*, vol. 161, no. 3, pp. 471–476, 2003.

[32] Y. Cao, D. Gillespie, and L.Petzold, "Avoiding negative populations in explicit Poisson tau-leaping," *Journal of Chemical Physics*, vol. 123, pp. 054 104–054 112, 2005.

[33] A. A. Julius, "Approximate abstraction of stochastic hybrid automata," in *Hybrid Systems: Computation and Control*, ser. LNCS, vol. 3927. Springer Verlag, 2006, pp. 318–332.

[34] J. T. Mettetal, D. Muzzey, J. M. Pedraza, E. M. Ozbudak, and A. van Oudenaarden, "Predicting stochastic gene expression dynamics in single cells," *Proc. National Academy of Science*, vol. 103, pp. 7304–7309, 2006.

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