Proceedings of the 9th International
Workshop on Discrete Event Systems
Göteborg, Sweden, May 28-30, 2008

From Discrete to Continuous and Back: Abstractions and Mesoscopic Phenomena in Cells

Ádám M. Halász[*#], *Member, IEEE,* A. Agung Julius[*$], *Member, IEEE,* George Pappas[*$], *Senior Member, IEEE* and Vijay Kumar[*#], *Fellow, IEEE*

***Abstract*—We discuss the interplay between stochasticity and multistability in bio-molecular networks. The resulting cell-level stochastic behavior reflects the fundamentally discrete and random nature of the underlying molecular processes. These ideas are illustrated on the well studied example of the *lac* operon. We first describe the switching behavior predicted by a differential-equation based model and then show how cell-level stochastic behavior emerges. Finally we point out that the observed macroscopic behavior may not be enough to determine both the dynamic and stochastic parameters.**

I. Introduction

Traditionally, the role of computers in molecular biology was associated with the management of large datasets that resulted from high throughput experimental methods, supporting a "breadth-first" approach to the study of living systems. More recently, the accumulated knowledge on bio-molecular processes and quantitative experimental methods have made possible a line of investigation similar to Physics, where quantitative modeling plays the central role. This "depth-first" approach aims at a *mechanistic* understanding of the emergence of cellular functions from the underlying molecular processes.

Quantitative methods are expected to help understand the functionality of complex networks involving many molecular species. However, biological complexity is fundamentally different from what we encounter in chaos or other high dimensional dynamical systems. A cell is more similar to a high-tech airplane than to a strongly coupled many-body system, in that there is a lot of structure within its several thousands of degrees of freedom. To further add to the challenge, the relevant interactions are often not well understood and the relevant parameters are not well known.

This work was supported in part by a grant from the Penn Genomics Frontiers Institute and grants from the US NSF and DARPA.

[*]The authors are with the General Robotics, Automation, Sensing and Perception Laboratory at the University of Pennsylvania, Levine Hall L403, 3330 Walnut Street, Philadelphia, PA 19104.

[#]Adam Halasz (phone: 215-573-2048; fax: 215-573-2048; email: halasz@grasp.upenn.edu) and Vijay Kumar are with the Department of Mechanical Engineering and Applied Mechanics.

[$]A. Agung Julius and George Pappas are with the Department of Electrical and Systems Engineering, University of Pennsylvania.

A central issue that mathematical theory must address is *how to simplify or abstract away elements of a complex bio-molecular network while capturing the basic functionality*. There is currently no coherent framework for the full dynamical version of this problem; (partial) solutions will likely involve hybrid dynamical systems, stochastic methods, and automated reasoning. The latter is likely the only way such large systems can be analyzed.

In this contribution we would like to illustrate the utility and the challenges of building integrative and quantitative models of cellular functions. Given the huge complexity involved, rationally derived simplified descriptions or abstractions are a central ingredient in cellular modeling. We will discuss a well known example, the lactose metabolism of the *Escherichia coli* bacteria, which is traditionally modeled as a continuum dynamical system. We will show how this picture is related to discrete event systems on two levels. On one hand the continuum model is the approximation of a discrete system described by the numbers of different molecules found in the cell; on the other hand, it is naturally abstracted to a system whose only two states correspond to those of high and low lactose metabolism.

II. Background

A. Overview of relevant processes in a bacterial cell

The functionality of a cell is ensured by a large number of molecular processes. A review of basic cell biology is far beyond the scope of this contribution and the interested reader is urged to refer to a textbook on cell biology. An excellent introduction to the field of cellular/molecular systems biology from the perspective of Control Theory is found in [8]. Brief introductions appropriate for quantitative scientists can be found in several recent Systems Biology textbooks.

978-1-4244-2593-8/08/$25.00 ©2008 IEEE

Authorized licensed use limited to: University of Pennsylvania. Downloaded on October 29,2026 at 15:28:48 UTC from IEEE Xplore. Restrictions apply.

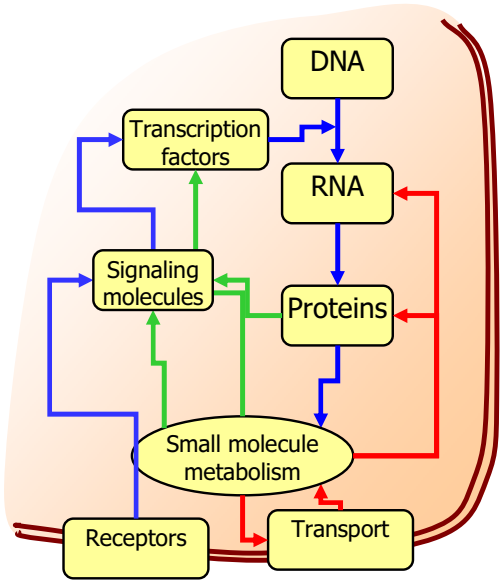


Fig.1 Schematic of important bio-molecular processes in a bacterial cell.

Below we enumerate the molecular processes relevant to our example. In addition to being the repository of genetic information, DNA also acts as a "script" that manages the functionality of cells. DNA can be seen as a long sequence of molecular characters; its different subsections have different roles. *Genes* are the smallest independently meaningful subsection; they represent blueprints of molecules with specific roles in the cell. The character sequence of a single gene is *transcribed* to a molecule of RNA by a molecular motor called *RNA polymerase*. One type of RNA, *messenger RNA*, acts as an intermediary and is *transcribed* by another molecular motor (*ribosome*) into a sequence of amino-acids. The final product, a functional *protein*, is obtained after the sequence of amino-acids folds into the appropriate spatial structure. We refer to this process as the *expression* of a gene. The DNA → RNA → protein paradigm is sometimes referred to as the "Central Dogma" of molecular biology.

The *metabolism* is the set of reactions that transform the available nutrients into energy and substances used to build new cellular material (including RNA and proteins). Some proteins act as *enzymes* (which facilitate metabolic and other chemical reactions), or are localized on the cell membrane where they act as *transporters* (allowing the uptake of substances from outside the cell). Finally, certain products of the metabolism act as *signals* for the occurrence of certain conditions (e.g. the presence or absence of a nutrient, excessive heat, etc.). They modulate the transcription process directly or indirectly, by activating or inhibiting *transcription factors*.

Thus, genes are expressed partially *in response to* the current state of the cell. This functionality is determined by the rules by which transcription factors are released and the ways they control the expression of different genes. While all of this structure is ultimately encoded in the DNA, many of these logical implications are revealed only when the relevant conditions occur. Disentangling this machinery is one of the ultimate goals of molecular systems biology.

B. *Dynamical models and their limitations*

Chemical processes are traditionally described in terms of *rate laws*. For example, consider the classic enzymatic reaction where the substrate S is transformed into the product P in the presence of enzyme E:

$$S \xrightarrow{E} P$$

The *rate* of this reaction, the quantity (concentration) of substrate consumed per unit time, is given by:

$$r = -\frac{d[S]}{dt} = \frac{d[P]}{dt} = k_{cat}\frac{[E][S]}{[S]+K_m}$$

The bracketed quantities *[S],[E],[P]* represent *concentrations* of the respective substances. The two constants k_{cat} and K_m are specific to the reaction or rather, the enzyme that makes it possible.

Metabolic reactions are described by rate laws similar to the Michaelis-Menten rate law cited above. Other processes such as transport, as well as transcription and translation, are also described by algebraic formulae that relate the reaction rate to the present (or past, in the case of transcription / translation) concentrations of the participating substances. In principle, all cellular processes could be accounted for if the rate laws and rate constants were known, and we would have a complete dynamical model of the organism in the form of a system of ordinary differential equations (ODEs). This approach is currently unrealistic for at least two reasons. There are a few tens of thousands of such processes in a single bacterial cell, and the rate laws and constants are not known for most of them. Second, the traditional methods for analyzing ODE systems are limited to either linear systems, or nonlinear systems of only a few variables.

Thus, even though most elementary cellular processes are known, a head-on approach to the entire cell as a

Authorized licensed use limited to: University of Pennsylvania. Downloaded on October 29,2026 at 15:28:48 UTC from IEEE Xplore. Restrictions apply.

dynamical system is unrealistic due to complexity and to the lack of complete kinetic information. The complexity can be reduced significantly by focusing on individual functions and by the use of abstractions. Abstractions are simplified descriptions that arise from the analysis of a model of the system.

A common abstraction is based on *switching behavior,* occurring on the level of individual genes and on that of entire cellular functions. Biologists often refer to genes being expressed or not; cellular properties linked to a [group of] gene[s] are considered either present or absent. Many cellular behaviors are well described by such *discrete abstractions* where continuous variables are replaced by discrete ones. These abstractions better capture the functionality of cells, and also reduce the complexity of the mathematical description [2].

C. *The discrete nature of molecular processes*

The number of molecules in a *macroscopic* amount of a substance is huge; thus the amount of substance is normally regarded as a continuous quantity. However, bacterial cells are extremely small, and the number of molecules of some species in a cell may be too small to justify the continuum description. Continuous chemical rate laws break down for small molecule numbers. Sometimes we refer to this as the *mesoscopic regime*.

A correct description on this level must take into account the fundamentally stochastic nature of molecular processes. The appropriate computational framework was initiated by the work of Gillespie [3]. The system is described as a continuous time Markov chain, whose states are labeled by the numbers of molecules of each species. Possible transitions correspond to individual occurrences of one of the reactions. First-principles stochastic simulation methods are well studied, but are significantly more expensive than differential equation simulators [3, 4]. On the other hand the stochastic description of all variables in the system is typically not necessary, since many species have large molecule numbers. The proper framework and approximation methods are currently being developed.

We can summarize this cursory presentation of molecular processes as follows. The relevant molecular processes are traditionally described by ordinary differential equations for the continuous concentrations of the respective substances. The continuous picture needs to be (partially) replaced with discrete elements for two apparently distinct reasons. These are: the emergence of switching behavior on the level of the entire cell, which justifies discrete abstractions; and the fact that on the molecular level, the correct description of chemical reactions involves *individual* molecules.

We will now discuss an example where the two discrete 'ends' of the continuous rate-law based description overlap. We will see the underlying stochasticity reflected in a high level description of a cellular system. This system controls the lactose metabolism of *E.coli* bacteria by regulating the expression level of a group of genes that form the *lac* operon [4].

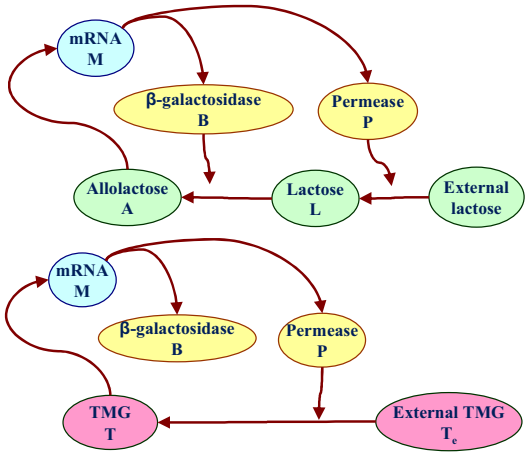


Fig.2 The lactose utilization network and the network induced by TMG

III. BISTABILITY IN THE *LAC* OPERON

Lactose is a sugar, a carbon source which bacteria (like humans) can use in their metabolism. There are other sugars, for example glucose, which are better nutrients and are thus preferred over lactose. In order to metabolize lactose, the cell requires a number of enzymes, which are encoded by two of the three genes of the *lac* operon. However, their synthesis represents an expense which is to be avoided if either there is no lactose in the environment or if a better alternative is available. This is a likely motivation for the bistable behavior of the dynamical system formed by the *lac* promoter, its gene products, lactose and its metabolized versions. On a technical level, the bistability, the existence of two dynamic equilibria, is made possible by a *positive feedback* mechanism (Figure 2) which can be understood using a simplified model involving five substances: the mRNA of the

Authorized licensed use limited to: University of Pennsylvania. Downloaded on October 29,2026 at 15:28:48 UTC from IEEE Xplore. Restrictions apply.

operon (M); two of its products: permease (P), which allows the uptake of external lactose and beta-galactosidase (B) which metabolizes lactose; intracellular lactose (L); and finally, allolactose (A). Allolactose has an enhancing effect on the expression of the *lac* genes, thereby increasing the production of the enzymes that bring in and metabolize lactose.

Thus, the *lac* operon has two modes, *induced*, where lactose is actively metabolized, and *uninduced*, where the system is in a low, conservation mode where just enough enzyme is maintained that can serve as a seed to build up the machinery if a large enough concentration of lactose is encountered (Figure 2, top). This is an example of a system that can be abstracted

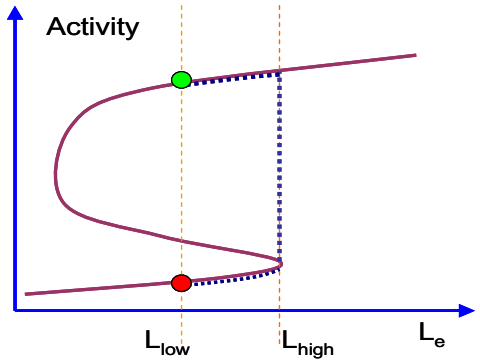


Fig. 3. Steady state diagram.

to a deterministic automaton (Figure 3), where transitions between the induced and uninduced states follow simple rules: induction if the external lactose concentration L_e exceeds an upper threshold L_{high} and de-induction if L_e drops below a lower threshold L_{low}.

This system is studied experimentally using TMG instead of lactose. TMG is not metabolized, but it has the same effect on transcription as allolactose. This simplifies both the experimental conditions and the network (Figure 2, bottom). This system is modeled using a set of four ordinary differential equations.

$$\frac{dM}{dt} = \alpha_M \frac{1 + K_1 T^2}{K + K_1 T^2} + \Gamma_0 - (\gamma_M + \mu) M$$

$$\frac{dB}{dt} = \alpha_B M - (\gamma_B + \mu) B$$

$$\frac{dT}{dt} = \alpha_L P \frac{T_e}{K_{T_e} + T_e} - \beta_L P \frac{T}{K_L + T} - (\gamma_T + \mu) T$$

$$\frac{dP}{dt} = \alpha_P M - (\gamma_P + \mu) P$$

Here, *M,B,T,P* are is the concentrations of mRNA, beta-galactosidase, (internal) TMG, and permease, respectively. The remaining symbols are model constants taken from Yildirim and Mackey [1].

IV. Stochasticity

Significant mathematical biological work has been devoted to dynamical models of the *lac* operon (see for example [1] and references therein). However, as early as the 1950's, it was known that the deterministic picture could not possibly be rigorously correct. While studying the enzymatic activity of a culture of bacteria, Novick and Weiner [5] observed that a continuous variation of *total* enzymatic activity is actually generated by individual cells that have high or very low activity. They termed this the *all-or-none* phenomenon and used it as a starting point to develop the bistable model discussed in the previous section.

If the bistable ODE model were rigorously correct, the *entire population* should be *simultaneously* induced or uninduced. All bacteria should behave identically and follow the time dependence predicted by the equations

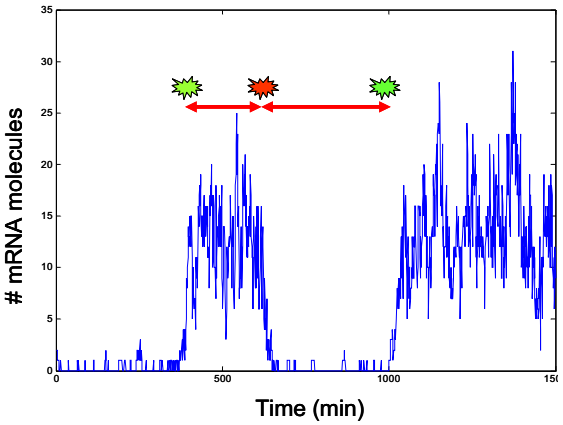


Fig. 4. Spontaneous transitions

of motion. In particular, in an experiment where an initially uninduced population is suddenly exposed to a high lactose concentration, all cells should induce simultaneously. This has been found not to be the case In addition to the classic results of Novick and Weiner, modern single cell experiments [6] have demonstrated *directly* that this is not the case. In hysteresis experiments where the external inducer concentration was varied gradually, Ozbudak and coworkers found that individual cells switched on or off faster than the observed timescale of the variation of the population-wide enzyme activity. This indicates that there is a source of noise or fluctuations which make each cell evolve differently.

Authorized licensed use limited to: University of Pennsylvania. Downloaded on October 29,2026 at 15:28:48 UTC from IEEE Xplore. Restrictions apply.

Mesoscopic effects due to *small molecule copy numbers* are a ubiquitous source of such cellular noise, albeit not the only one. There are standard methods to simulate chemical reactions under these circumstances [3, 4]. We performed stochastic simulations using the deterministic model discussed above as a starting point; not surprisingly, they revealed the existence of "tunneling" transitions between the classical regions of attraction of the two steady states (Figure 4). Thus the continuous differential equation model turns out to be an idealization of a Markovian stochastic process.

We should point out that first-principle molecular simulations are orders of magnitude more expensive than the corresponding ODE simulations. Often one has to deal with systems where some molecular species are present in low copy numbers, requiring a stochastic simulation, while other species are very abundant and have consistently high molecule copy numbers. Stochastic simulation of these abundant species is very costly and does not provide useful information (since the intrinsic stochastic fluctuations of these species are negligible). The proper approach to mixed stochastic-deterministic simulations for these cases is somewhat subtle. There is ongoing work on building reliable algorithms that can automatically switch to the most economical mode that gives an acceptable precision [4].

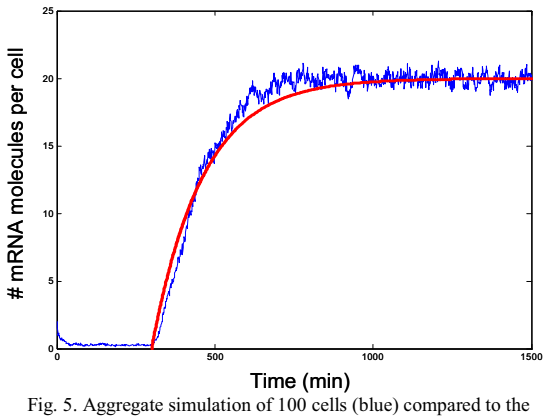


Fig. 5. Aggregate simulation of 100 cells (blue) compared to the analytical prediction derived from the discrete stochastic model.

For more details on the simulations discussed here we refer to [7]. Our mixed simulations allowed us to perform many individual runs, equivalent to simulations of small colonies of 100-1000 cells. Thus we were able to make "aggregate" predictions of how the average observed activity would evolve during an induction experiment. The blue line in Figure 5 is the result of this type of bulk simulation. We observed that the total activity approached a smooth curve, which was well approximated by a simple analytical formula. This was the starting point for building a simplified model, where the behavior of a single cell, as predicted by the stochastic version of the four-dimensional model discussed in Section III, was approximated by a two-state stochastic switch.

V. Stochastic Automaton

A. Two-level stochastic automaton

The ODE model exhibited switching behavior and could be abstracted to a deterministic switch [2]. In the stochastic case, one would still like to be able to think of the *lac* system as bistable, but with appropriately modified transition rules. This is exactly what we did in the case of the *lac* operon, connecting the high level discrete abstraction with the underlying stochastic phenomenon [7]. By performing first-principles

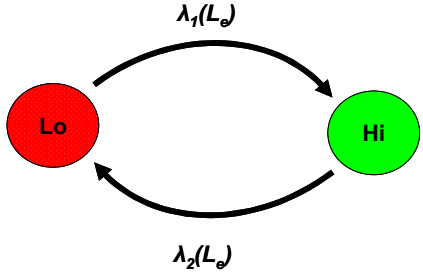


Fig. 6. Discrete stochastic abstraction.

simulations we were able to derive estimates for the *spontaneous transition rate* between the induced and uninduced states. Thus, the identities and parameters of the steady states remain essentially the same, including the dependence of the various equilibrium concentrations on external lactose. The transitions, however, become stochastic, characterized by rates that are a function of the same external lactose concentration (Figure 6). This model can now be used predict the aggregate behavior of a large population of cells.

$$\frac{dN_{Hi}}{dt} = -\frac{dN_{Lo}}{dt} = \lambda_1(L_e)N_{Lo} - \lambda_2(L_e)N_{Hi}$$

In Figure 5 we show the aggregate simulation of 100 cells compared with an analytical prediction based on the transition rates derived from individual simulations.

B. Stochastic versus dynamic parameters

Often, only aggregate quantities are measurable in experiments (see Figure 5). In fact, the parameters

Authorized licensed use limited to: University of Pennsylvania. Downloaded on October 29,2026 at 15:28:48 UTC from IEEE Xplore. Restrictions apply.

used in the original ODE model [1] were mostly derived based on this type of experiment, with no regard to the stochastic nature of the underlying processes. The time constant of the aggregate behavior in Figure 5, is the result of the interplay of two characteristic times: the average escape time, a stochastically defined quantity, and the ballistic equilibration time [6], which describes the deterministic convergence to the equilibrium states. It is conceivable that different configurations of stochasticity and individual deterministic dynamics may lead to very similar bulk dynamics.

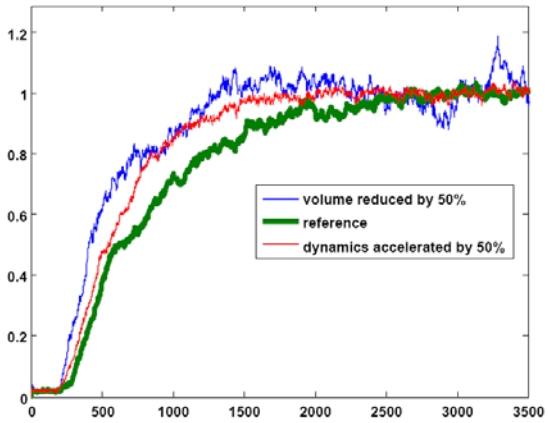


Fig. 7: Comparison of the effects of speeding up the dynamics (red) and reducing the cell volume (blue)

We investigated this question. Our preliminary results indicate that bulk measurements may not directly distinguish between stochastic and dynamic properties of the system. In Figure 7 we show the results of three simulations of an induction experiment of an aggregate of 100 cells. The reference simulation is shown in green. We performed the same simulation once with the cell volume reduced by a factor of 50%, and once with the dynamics accelerated by 50%. While the two results do not coincide precisely, we see that the effects of the two changes are similar. Thus, a variation of the cell volume, which only influences the stochastic aspects (since it does not change the corresponding ODE model), can have the same apparent effect as changes to the time constants of the underlying molecular processes (which do not change the equilibria but scale the time dependence of the solutions).

This result highlights the practical problem of indirectly known parameters. The kinetic parameters of the ODE models in the literature are fitted to match macroscopic observations of cell populations with the ODE prediction. The stochastic version of such a model will lead to a qualitatively different aggregate behavior. The distinction can be seen by comparing Figure 5 and Figure 4. The time for the induction of an individual cell is approximately 20 times shorter than the time for the induction of a colony of 100 cells. The correct microscopic model should match the individual distributions of cells as well as correctly predict the aggregate behavior of cell colonies.

In summary, we propose a "depth-first" approach to molecular systems biology. The aim is to understand the emergence of cellular functions by integrating the quantitative knowledge about individual molecular processes. Traditionally these are described in terms of differential equations. Two major challenges are incomplete dynamical information and the complexity of the systems involved. Both issues can be alleviated by identifying good abstractions. Abstractions based on switching behavior are one source of discrete models in genetic networks. Another source of discrete features is due to the underlying discrete and stochastic nature of biochemical processes. Consistent simulation methods for this type of systems are readily available and can be a source of significant insight. We have shown that these three features: continuous dynamics, emerging discrete behavior and underlying stochasticity can be intertwined in subtle ways which are, fortunately, amenable to our analysis.

REFERENCES

[1] N. Yildirim and M. C. Mackey, "Feedback regulation in the lactose operon: A mathematical modeling study and comparison with experimental data," *Biophys. J.*, vol. 84, pp. 2841–2851, 2003.

[2] A. Halasz, V. Kumar, M. Imielinski, C. Belta, O. Sokolsky, and S. Pathak, "Analysis of lactose metabolism in *E. coli* using reachability analysis of hybrid systems," *IET Syst. Biol.*, vol. 1, no. 2, pp. 130–148, 2007.

[3] D. T. Gillespie, "A general method for numerically simulating the stochastic time evolution of coupled chemical reactions," *J. Comput. Phys.*, vol. 22, pp. 403–434, 1976.

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

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

[6] 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, Feb. 2004

[7] A. A. Julius, A. M. Halasz, M. S. Sakar, H. Rubin, V. Kumar, and G. J. Pappas, "Stochastic Modeling and Control of Biological Systems: The Lactose Regulation System of *Escherichia coli*", IEEE Transactions on Automatic Control, Special Issue on Systems Biology, pp. 51-65, 2008.

[8] E.D. Sontag. "Molecular systems biology and control**"**, *Eur. J. Control*, 11(4-5):396-435, 2005

Authorized licensed use limited to: University of Pennsylvania. Downloaded on October 29,2026 at 15:28:48 UTC from IEEE Xplore. Restrictions apply.