

# Analysis and control of gene regulation network models using kinetic semi-discretization

Mihály A. Vághy<sup>1</sup>, Irene Otero-Muras<sup>2</sup>, Gábor Szederkényi<sup>1,3</sup>

**Abstract**—In this paper the exogenous control of gene regulatory networks is investigated through the semi-discretized partial integro-differential equation (PIDE) describing the time-evolution of the network's probability density function. With an appropriate finite volume method the semi-discretized system is a mass-conservative linear compartmental model, and thus it preserves most qualitative properties of the solution of the PIDE, namely, it is nonnegative and mass conservative. These advantages combined with the newly investigated mesh-invariance of control allows us to efficiently determine the reachability set. The possibilities of this framework are demonstrated through an illustrative example from literature.

## I. INTRODUCTION

Gene expression is one of the most fundamental biological processes, where DNA information is realized to proteins in living organisms. Gene regulatory networks (GRNs) are complex mechanisms through which cells can process internal and external information (signals) [1].

The control of gene regulatory networks is an active area of research in synthetic biology, where the design and implementation of suitable controllers is needed to improve the reliability and performance of synthetic gene circuits for various applications [2]. One of the challenges in controlling gene regulatory processes is the presence of molecular noise, which is particularly relevant in bacteria, due to the low copy number of the species involved. Two important advances in control in synthetic biology over the last years are the anti-thetic control by Briat et al [3] allowing perfect adaptation in the context of noise, and the stabilization of the toggle switch in *E. coli* by Lugagne et al [4]. A linear feedback control algorithm is employed in both cases.

The dynamical model applied in this work is rooted in [5], where the evolution of the probability density function of protein concentration is approximated with a partial integro-differential equation (PIDE), derived from the master equation. This model is able to capture the characteristic random bursts of protein production as it has been observed and described in the literature [6]. An important step in the PIDE-based modeling of GRNs was [7], where the multi-dimensional (also called generalized) Friedman model was introduced, which is able to describe the dynamics of several genes expressing different protein types. The numerical solution of such PIDEs is a difficult technical challenge for which

a semi-Lagrangian computation framework was proposed in [8]. It has been shown that such an approach is already suitable for simulation based control [9], [10]. However, an ODE-based description of the process is also preferred both for dynamical analysis and controller design. Therefore, a semi-discretization of the multidimensional PIDE model was proposed in [11] resulting in a kinetic compartmental description of the system in possibly time-varying linear ODE form.

Based on the above mentioned results, the goals of this paper are firstly to further develop the kinetic discretization scheme in [11], and secondly to propose a control design approach to reach a stationary protein distribution with prescribed properties.

## II. NOTATIONS AND BACKGROUND

In this section we briefly introduce multidimensional gene regulatory networks and the relevant theory of compartmental and kinetic systems.

### A. PIDE model for gene regulatory networks

The following short introduction is based on [5], [7], [11]. We consider a gene regulatory network consisting of  $n$  genes  $G = \{DNA_1, DNA_2, \dots, DNA_n\}$  that express  $n$  proteins  $X = \{X_1, X_2, \dots, X_n\}$  via the corresponding messenger RNAs  $M = \{mRNA_1, mRNA_2, \dots, mRNA_n\}$ . We follow the central dogma of molecular biology, which asserts that the gene instructions are transcribed into mRNAs, that are translated into proteins. The continuous number of mRNA molecules and proteins are denoted by  $m, x \in \mathbb{R}^n$ , respectively.

The promoters corresponding to each gene are assumed to switch between active and inactive states, denoted by  $DNA_{i,on}$  and  $DNA_{i,off}$ , respectively. The transition is controlled by a feedback mechanism of protein binding. In general, this mechanism may require the binding of multiple types of proteins besides the one expressed by the given gene, and thus for the sake of generality, we assume that any protein can repress or activate any gene in the network. This mechanism is typically modeled by multivariate Hill functions. We define the matrix  $H \in \mathbb{Z}^{n \times n}$ , where  $H_{ij}$  represents the Hill coefficient of the cross-regulation. If  $H_{ij}$  is positive (respectively, negative), then  $X_j$  inhibits (respectively, promotes) the expression of  $X_i$ .

The transcription of  $DNA_i$  into  $mRNA_i$  is assumed to be a first order processes occurring with rate  $k_m^i$  per unit time and with transcriptional leakage  $\epsilon_i \in (0, 1)$ . Then the

<sup>1</sup>Faculty of Information Technology and Bionics, Pázmány Péter Catholic University, Práter u. 50/a, Budapest, H-1083, Hungary.

<sup>2</sup>Institute for Integrative Systems Biology, Spanish Council for Scientific Research, Carrer del Catedràtic Agustín Escardino Benlloch, Valencia, 46980, Spain.

<sup>3</sup>Systems and Control Laboratory, ELKH Institute for Computer Science and Control (SZTAKI), Kende u. 13-17, H-1111, Budapest.

transcription can be written as

$$R_T^i(\mathbf{x}) = k_m^i c_i(\mathbf{x}),$$

where  $c_i : \mathbb{R}_+^n \rightarrow [\epsilon_i, 1]$  is an appropriate Hill function depending on the feedback regulation. Finally, the translation rate of protein  $X_i$  is defined as

$$R_X^i(m_i) = k_x^i m_i.$$

The messenger RNA and protein degradation is assumed to take the form

$$G_m^i(m_i) = -\gamma_m^i m_i \quad G_X^i(\mathbf{x}) = -\gamma_x^i(\mathbf{x}) x_i,$$

where  $\gamma_m^i > 0$  and  $\gamma_x^i : \mathbb{R}_+^n \mapsto \mathbb{R}_+$ . Following [7] it is assumed that  $\frac{\gamma_m^i}{\gamma_x^i(\mathbf{x})} \gg 1$  in order to ensure the validity of the subsequent model.

We use the standard exponential distribution to model protein bursting; that is, the conditional probability of the protein level jumping from  $y_i > 0$  to  $x_i \geq y_i$  is

$$\omega_i(x_i - y_i) = \frac{1}{b_i} \exp\left[-\frac{x_i - y_i}{b_i}\right],$$

where  $b_i = \frac{k_x^i}{\gamma_m^i}$ .

With the above assumptions the probability density function (PDF) of the protein level, denoted by  $p(t, \mathbf{x})$ , can be modeled with the following PIDE:

$$\begin{aligned} \frac{\partial p(t, \mathbf{x})}{\partial t} &= \sum_{i=1}^n \frac{\partial}{\partial x_i} [\gamma_x^i(\mathbf{x}) x_i p(t, \mathbf{x})] \\ &+ \sum_{i=1}^n k_m^i \int_0^{x_i} \beta_i(x_i - y_i) c_i(\mathbf{y}_i) p(t, \mathbf{y}_i) dy_i, \end{aligned} \quad (1)$$

where  $\mathbf{y}_i = \mathbf{x} + (y_i - x_i)e_i$  and the  $\beta_i$  functions have the following form:

$$\beta_i(x) = \omega_i(x) - \delta(x),$$

where  $\delta$  is the Dirac delta function. In [12] the authors show the well-posedness of (1) in the generalized (mild) sense; that is, for  $p_0 \in \mathcal{L}^1(\mathbb{R}^n)$  there exists a unique mild solution  $p \in \mathcal{C}(\mathbb{R}_+; \mathcal{L}^1(\mathbb{R}^n))$  with the following properties:

- 1) nonnegativity: if  $p_0$  is nonnegative, then so is the solution  $p(t, \cdot)$  for all  $t \geq 0$ ,
- 2) mass conservation:

$$\int_{\mathbb{R}_+^n} p(t, \mathbf{x}) d\mathbf{x} = \int_{\mathbb{R}_+^n} p_0(\mathbf{x}) d\mathbf{x}.$$

In fact, if  $p_0 \in \mathcal{C}^{1,b}(\mathbb{R}_+^n)$  for some appropriate  $b > 0$  (e.g., in one dimension  $b = b_1$ ), then there exists a unique classical solution  $p \in \mathcal{C}^1(\mathbb{R}_+; \mathcal{L}^1(\mathbb{R}_+^n))$ . Note, that in the probabilistic setting in applications we usually assume that  $p_0$  is nonnegative and its integral is one.

## B. Compartmental and kinetic systems

We briefly introduce compartmental systems based on [11], [13]. Compartmental differential equations are often used to model physical phenomena governed by a conservation law, such as conservation of mass. A compartment can store a certain amount of a material that is kinetically homogeneous; that is, the entering material is instantly mixed with that of the compartment. As long as we can interpret the conservation law, a compartment can even describe abstract quantities, such as probabilities in our case. Nevertheless, we will usually refer to the amount of the modeled quantity in the compartment as the mass in the compartment, and to the conservation law as conservation of mass.

Let us consider a system with  $m$  compartments and let  $q_i$  represent the mass in the  $i$ th compartment. In general, any compartment can be connected to any other compartment and to the environment in both directions. We denote with  $F_{ij}$  the flow from the compartment  $q_j$  to the compartment  $q_i$ , with  $I_i$  the material inflow from the environment to compartment  $q_i$  and with  $F_{0i}$  the material outflow from compartment  $q_i$  to the environment. We assume that there are no loops in the system as they do not introduce additional dynamics in this setting. Then the time-evolution of the system is given by the following system of differential equations:

$$\dot{q}_i = \sum_{j \neq i} (-F_{ji} + F_{ij}) + I_i - F_{0i}. \quad (2)$$

We impose the following physical assumptions to the system:

- 1) for any  $i, j, t \geq 0$ ,  $i \neq j$  we have that  $F_{ij} \geq 0$ ,  $I_i \geq 0$  and  $F_{0i} \geq 0$ ,
- 2) for any  $i, t \geq 0$  if  $q_i(t) = 0$ , then  $F_{0i} = F_{ji} = 0$  for each  $j$ .

These properties ensure the invariance of the nonnegative orthant; that is, assuming a nonnegative initial condition, our solution is guaranteed to be nonnegative. In general, the above functions can depend on the mass of any compartment and possibly on  $t$  as well. Then it can be shown that if each  $F_{ij}$  and  $F_{0i}$  is at least  $\mathcal{C}^k$ , then we can rewrite (2) as

$$\dot{q}_i = -\left(f_{0i} + \sum_{j \neq i} f_{ji}\right) q_i + \sum_{j \neq i} f_{ij} q_j + I_i, \quad (3)$$

where  $F_{ij} = f_{ij} q_j$  and the so-called fractional transfer coefficients  $f_{ij}$  are at least  $\mathcal{C}^{k-1}$ . We can then naturally rewrite (3) in matrix form as

$$\dot{q} = fq + I.$$

If each fractional transfer coefficient  $f_{ij}$  only depends on  $q_j$ , then the system is called a donor controlled system. If each coefficient is constant, then the system is called a linear donor controlled system.

Linear donor controlled systems can naturally be represented as chemical reaction networks, or kinetic systems. For a brief introduction, we refer to [14]. For each compartment

with index  $i$ ,  $q_i$  represents the mass (or alternatively, the concentration) of the one-specie complex  $Q_i$ , and for each transition from compartment  $i$  to  $j$ , we assign the reaction  $Q_i \rightarrow Q_j$ . Using this construction, we can not only rely on the comprehensive theory of compartmental models but on that of kinetic systems as well. While most qualitative properties of linear donor controlled systems we consider can be derived from both modeling approaches, a notable piece of additional information in chemical reaction network theory is the stability analysis using a logarithmic Lyapunov function, discussed in more detail in III-B.

### III. MODIFIED KINETIC FINITE VOLUME METHOD

In this section we introduce an extended version of the kinetic finite volume method described in [11], modified to be suitable for control design. We wish to employ an exogenous control on the population level through appropriate inducers affecting protein bursts; that is, we assume that  $c_i(\mathbf{x}) = c_i(\mathbf{x}, \mathbf{I})$  in (1), where  $\mathbf{I}$  denotes the concentration vector of the inducers. In order to adhere certain biological constraints we assume that the range of  $c_i$  remains in  $\in (0, 1)$ . For the sake of simplicity we assume that  $\mathbf{I} \in \mathbb{R}^n$  and note that we set  $I_i \equiv 0$  if we do not control the production of protein  $P_i$ .

Note that in practical applications we may assume that there can only be a finite number of proteins of each kind. This consideration is naturally backed by the fact that the solution of (1) is integrable so that  $p(t, \mathbf{x})$  vanishes as  $\|\mathbf{x}\|_{\mathbb{R}^n} \rightarrow \infty$ , for any  $t \geq 0$ . Thus, we discretize over the finite domain  $\Omega = \times_{i=1}^n (0, L_i)$  for appropriately large  $L_i > 0$  values. According to these considerations we also assume that  $\int_{\Omega} p_0(\mathbf{x}) d\mathbf{x} = 1$ .

For each  $i = 1, 2, \dots, n$  we divide the interval  $(0, L_i)$  into  $N_i$  equal subintervals, and thus define the positive step sizes  $h_i = \frac{L_i}{N_i}$ . Let

$$K_{\alpha} = \times_{i=1}^n K_{\alpha_i}(h_i) = \times_{i=1}^n [(\alpha_i - 1)h_i, \alpha_i h_i],$$

where  $\alpha \in \mathbb{N}^n$  is a multi-index and  $\bigcup_{\alpha} K_{\alpha} = \Omega$ . Let us note that each cell has the same size and define  $h = |K_{\alpha}| = \prod_{i=1}^n h_i$ . For each cell  $K_{\alpha}$  we introduce the function  $p_{\alpha}(t)$  assumed to approximate the cell average as

$$p_{\alpha}(t) \approx \frac{1}{h} \int_{K_{\alpha}} p(t, \mathbf{y}) d\mathbf{y}.$$

Let  $x_{\alpha} = [x_{\alpha}^1 \ x_{\alpha}^2 \ \dots \ x_{\alpha}^n]^{\top}$  be the midpoint (w.r.t. each dimension) of  $K_{\alpha}$  and  $x_{\alpha}^{i \pm \frac{1}{2}} = x_{\alpha}^i \pm \frac{h_i}{2}$ ; that is, the variables  $x_{\alpha}^{i \pm \frac{1}{2}}$  correspond to the coordinates of the boundaries of  $K_{\alpha}$ . Similarly to a classical finite volume setting, we introduce coefficients as cell averages of the coefficient functions  $\gamma_x^i$  as

$$\gamma_{\alpha}^i = \frac{1}{h} \int_{K_{\alpha}} \gamma_x^i(\mathbf{y}) d\mathbf{y},$$

We slightly modify this in the case of the functions  $c_i$  and instead use their midpoint values so that they are not

integrated; that is, we set  $c_{\alpha}^i(\mathbf{I}) = c_i(x_{\alpha}, \mathbf{I})$ . Hence, for  $i = 1, 2, \dots, n$  we compute the coefficients

$$\begin{aligned} b_{\alpha, \alpha_i}^i &= \frac{1}{h_i/2} \int_{[(i-1)h_i, (i-1/2)h_i]} \beta_i(x_{\alpha}^i - y) dy \\ &= \frac{1}{h_i/2} \int_{[x_{\alpha}^i - h_i/2, x_{\alpha}^i]} \beta_i(x_{\alpha}^i - y) dy, \\ b_{\alpha, j}^i &= \frac{1}{h_i} \int_{K_j(h_i)} \beta_i(x_{\alpha}^i - y) dy, \quad j = 1, 2, \dots, \alpha_i - 1. \end{aligned}$$

The derivative terms describing protein degradation are approximated with difference quotients of the form

$$\begin{aligned} &\left. \frac{\partial}{\partial x_i} [\gamma_x^i(\mathbf{x}) x_i p(t, \mathbf{x})] \right|_{K_{\alpha}} \\ &\approx \frac{1}{h_i} (\gamma_{\alpha+e_i}^i x_{\alpha}^{i+\frac{1}{2}} p_{\alpha+e_i}(t) - \gamma_{\alpha}^i x_{\alpha}^{i-\frac{1}{2}} p_{\alpha}(t)), \end{aligned}$$

and the integrals corresponding to protein bursts in (1) are approximated with weighted sums, yielding the system

$$\begin{aligned} \dot{p}_{\alpha}(t) &= \sum_{i=1}^n \frac{1}{h_i} (\gamma_{\alpha+e_i}^i x_{\alpha}^{i+\frac{1}{2}} p_{\alpha+e_i}(t) - \gamma_{\alpha}^i x_{\alpha}^{i-\frac{1}{2}} p_{\alpha}(t)) \\ &\quad + \sum_{i=1}^n k_m^i \sum_{j=1}^{\alpha_i} h_{\alpha, j}^i b_{\alpha, j}^i c_{\alpha, j}^i(\mathbf{I}) p_{\alpha_{i, j}}(t); \quad (4) \\ p_{\alpha}(0) &= \frac{1}{h} \int_{K_{\alpha}} p_0(\mathbf{y}) d\mathbf{y}, \end{aligned}$$

where  $\alpha_{i, j} = \alpha + (j - \alpha_i)e_i$  and

$$h_{\alpha, j}^i = \begin{cases} h_i/2, & j = \alpha_i, \\ h_i, & j \neq \alpha_i. \end{cases}$$

A calculation identical to that of [11] shows the system (4) is a mass-conservative linear compartmental/kinetic system with a strongly connected compartmental structure even if the  $c_{\alpha}^i$  variables are just point-values and not cell averages. This shows via standard results on compartmental systems that if  $\mathbf{I}$  is constant, then there exists a unique positive equilibrium that attracts every admissible initial value [15, Theorem 6]. Note that while the bursts and degradations inherently define some ‘‘spatial’’ structure between the  $p_{\alpha}$  variables, it might be more useful to think of the truncated semi-discretized model as a flattened  $N := \prod_{i=1}^n N_i$ -dimensional system of the form

$$\dot{p}(t) = \Gamma(\mathbf{I})p(t). \quad (5)$$

Here  $\Gamma(\mathbf{I})$  corresponds to a compartmental strongly connected graph structure; that is, according to Section II-B it has nonnegative off-diagonal elements and zero column sums and it describes a trap containing all compartments. Let us collect the degradation coefficients  $\gamma_{\alpha}^i$ , burst coefficients  $b_{\alpha, j}^i$  and controlled coefficients  $c_{\alpha}^i(\mathbf{I})$  into the matrices  $G$ ,  $B$  and  $C(\mathbf{I})$ , respectively. Then (5) can be rewritten as

$$\dot{p}(t) = Gp(t) + (B \odot C(\mathbf{I}))p(t),$$

where  $\odot$  denotes the Hadamard (or elementwise) product.

### A. Explicit equilibrium computation for constant control

When we seek the equilibrium with a constant  $\mathbf{I}$  control value we wish to compute  $p^* \in \mathbb{R}^N$  such that  $\Gamma p^* = 0$  and  $h \sum_{j=1}^N p_j^* = 1$ , where  $\Gamma = \Gamma(\mathbf{I})$ . Following [11] we know that such a vector uniquely exists and is strictly positive. Of course one could simulate (5) for an appropriately large final time using some numerical solver to approximate  $p^*$ . However, an essential observation of [11] is that by incorporating the mass-conservation into  $\Gamma$  we may approximate the equilibrium by solving a system of linear equations. In particular, it follows from compartmental theory that  $\text{rank } \Gamma = N - 1$ , but replacing the first (or in fact, any) row of  $\Gamma$  with  $h \mathbf{1}_N^\top \in \mathbb{R}^N$  yields a matrix with full rank. Let  $\tilde{\Gamma}$  denote this modified matrix. To compute the equilibrium we need to solve  $\tilde{\Gamma} p^* = e_1$ , where  $e_1 \in \mathbb{R}^N$  denotes the first unit vector.

### B. Asymptotic behaviour and performance of control

An important observation is that the coefficients of the system converge given that the controller converges to some  $I^* \in \mathbb{R}^n$ , and thus the steady-state (and the steady-state error) is uniquely determined by the limit of the controller dynamics. This shows that the exact control trajectory may be omitted and we can apply a constant control to yield the same closed-loop equilibrium. Of course it might be important to compute control trajectories that are optimal in some sense, for example w.r.t. convergence speed. As shown in [11] there is an entropy-like Lyapunov function for the system that simplifies to the well-known Kullback-Leibler divergence in a mass-conservative setting, given as follows

$$\begin{aligned} V(p, p^*) &= \sum_{j=1}^N \left( p_j \log \frac{p_j}{p_j^*} + p_j^* - p_j \right) \\ &= \sum_{j=1}^N p_j \log \frac{p_j}{p_j^*} = D_{KL}(p || p^*). \end{aligned}$$

We note that while the Kullback-Leibler divergence is not a metric as it is not symmetric and fails to satisfy the triangle inequality, it is a nonnegative measure and it is often used to estimate the difference of discrete probability distributions [16].

### C. Admissible meshes and mesh-independent control

As described before, the mesh size directly determines the number of variables of the system (5). We consider an explicit Euler scheme on (4) and denote the approximation of  $p_\alpha(t_k)$  as  $p_\alpha^k$ . Clearly we have that  $\sum_\alpha p_\alpha^{k+1} = \sum_\alpha p_\alpha^k$  for each  $k \geq 0$  since (4) is governed by a linear conservation law. An elementary computation shows that if the step sizes satisfy the following Courant-Friedrichs-Lewy (CFL) condition, then  $p_\alpha^k \geq 0$  holds for any  $k \geq 0$  and  $\alpha$ :

$$\Delta t \sum_{i=1}^n \max_{\mathbf{x} \in \Omega} \left( \frac{1}{h_i} \gamma_x^i(\mathbf{x}) x_i + k_m^i \exp\left(-\frac{h_i}{2b_i}\right) c_i(\mathbf{x}, \mathbf{I}) \right) \leq 1.$$

We note that the  $c_i$  functions are usually Hill-type saturating functions with the property  $c_i(\mathbf{x}, \mathbf{I}) \leq 1$  and that

$\exp\left(-\frac{h_i}{2b_i}\right) \leq 1$ , thus the second term is bounded by  $k_m^i$ . This shows that the degradation terms are often more dominant, hence in applications of biological relevance the CFL condition can be estimated as

$$\Delta t \sum_{i=1}^n \frac{1}{h_i} \max_{\mathbf{x} \in \Omega} \gamma_x^i(\mathbf{x}) x_i \leq 1.$$

Of course we can normally set larger  $\Delta t$  values when applying a more sophisticated time discretization method. However, this demonstrates a further benefit of the FVM-based population level control, since our investigation shows that usually one can resort to very coarse grids leading to smaller systems and larger admissible temporal steps. The computed control trajectory (or the steady-state constant control) can then be applied to a system with a finer mesh.

### D. Reachability

A crucial question is what kinds of probability distributions can be reached from an initial one. The considered control structure is strictly positive and bounded from below and above, thus it is anticipated that we cannot reach arbitrarily low and high expected values. However, relying on the above observations we can estimate the reachability set of the system numerically by computing the considered statistical measures of the unique equilibrium for a simple scan of control configurations. The continuous dependence on parameters (see, [17, Chapter V]) shows that the reachability set should be a connected set in  $\mathbb{R}^n$ , thus we could even interpolate control values based on an appropriately fine scan.

### E. PI control terms

A natural design principle of PID controllers can be to use as few control terms as possible. In many applications a well tuned proportional controller may suffice. This is not the case for semi-discretized gene regulatory networks as the above discussions show that in general we need nonzero steady-state control; that is, the steady-state error of the controlled system will be proportional to the required control value. While integral control has proven to be reliable for biomolecular networks [3], we found that its performance can be inferior to proportional-integral control. In certain cases introducing a derivative term could further increase the convergence speed or reduce overshoots and oscillations, but it does not seem to be necessary.

## IV. NUMERICAL EXPERIMENT: GENETIC TOGGLE SWITCH

We consider the classical toggle switch configuration consisting of two repressible promoters in a mutually inhibitory network. We introduce two corresponding inducers, each affecting one of the promoters. Our goal is to shift the expected values of the stationary probability density to some prescribed values. Figure 1 shows the structure of the controlled gene regulatory network.

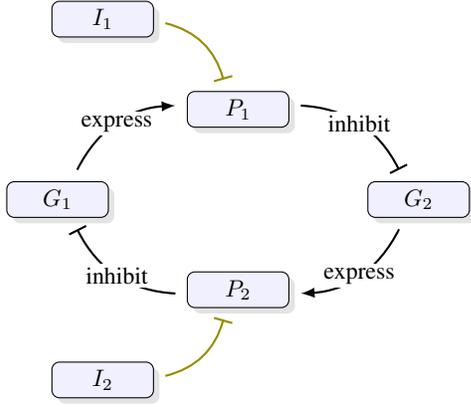


Fig. 1: Structure of the gene toggle switch.

Following [4] we introduce the parameters  $\theta_{I_i}$ ,  $\theta_{X_i}$  and  $\mu_{I_i}$  associated with the inducers' effects in the protein regulation. The burst coefficients are given by the following Hill-type functions

$$c_1(\mathbf{x}, \mathbf{I}) = c_1(x_2, I_1) = \frac{K_{12}(I_1)^H + \epsilon_1 x_2^H}{K_{12}(I_1)^H + x_2^H},$$

$$c_2(\mathbf{x}, \mathbf{I}) = c_2(x_1, I_2) = \frac{K_{21}(I_2)^H + \epsilon_2 x_1^H}{K_{21}(I_2)^H + x_1^H},$$

with

$$K_{12}(I_1) = \theta_{X_2} \left( 1 + \left( \frac{I_1}{\theta_{I_1}} \right)^{\mu_{I_1}} \right),$$

$$K_{21}(I_2) = \theta_{X_1} \left( 1 + \left( \frac{I_2}{\theta_{I_2}} \right)^{\mu_{I_2}} \right).$$

We consider  $H = 4$  and Table I shows the rest of the parameters of the system.

$\gamma_x^i$	$k_m^i$	$b_i$	$\epsilon_i$	$\theta_{X_i}$	$\theta_{I_i}$	$\mu_{I_i}$
1	12	6	0.1	31.94	11.65	2
1	7	$\frac{78}{7}$	0.1	30	$9.06 \times 10^{-2}$	2

TABLE I: PIDE parameters of the gene toggle switch.

First, we compute the equilibrium of the open-loop system (that is, when  $I_1 = I_2 = 0$ ) and then apply a PI controller to shape the protein density function. We consider a simple population level controller based on the expected values of the number of proteins. The desired and actual expected values are denoted as  $m_1^*$ ,  $m_2^*$  and  $m_1(t)$ ,  $m_2(t)$ , respectively. We note, that we may use other statistical measures, for example the modes of the marginal probability density functions as in [10]. Defining the errors  $e_1(t) = m_1^* - m_1(t)$  and  $e_2(t) = m_2^* - m_2(t)$  the dynamics of the PI controller is of the form

$$I_1(t) = I_1^0 + K_P^1 e_1(t) + K_I^1 \int_0^t e_1(s) ds,$$

$$I_2(t) = I_2^0 + K_P^2 e_2(t) + K_I^2 \int_0^t e_2(s) ds,$$

where we assume based on biological constraints that  $I_1 \in [0, 50]$  and  $I_2 \in [0, 1]$ . The initial values are set as  $I_1^0 = 20$ ,

$I_2^0 = 0.25$  and the feedback gains, based on [4], [10], as  $K_P^1 = 60$ ,  $K_I^1 = 20$ ,  $K_P^2 = 2.5$  and  $K_I^2 = 6.94 \cdot 10^{-1}$ . We note that for a new model these values could be obtained through the linearization of a coarse discretization. Figure 2a shows the open-loop equilibrium, while Figures 2b and 2c show the closed-loop equilibrium for  $m_1^* = 41$  and  $m_2^* = 55$  on a  $300 \times 300$  and a  $50 \times 50$  mesh, respectively. Table II shows the performance of the FVM with an explicit Euler discretization on different mesh-sizes with the same CFL ratio.

$50 \times 50$	$100 \times 100$	$200 \times 200$	$300 \times 300$
0.2087 s	2.4426 s	20.7634 s	90.4794 s

TABLE II: Average runtime of 100 simulations, performed on a computer with Intel(R) Core(TM) i7-8565U CPU @ 1.80GHz and 16 GB of RAM in MATLAB R2022b.

Figure 3 shows the performance of the PI control and the constant control measured as the time-evolution of the Kullback-Leibler divergence of the state and the equilibrium. While in this case the PI control outperforms the constant control, it is clear that the monotonicity cannot be guaranteed, while in a constant control setting  $D_{KL}(\cdot \| p^*)$  is known to be a Lyapunov function, thus it is strictly decreasing. We emphasize that the control is based on the error of the expected values, not on the Kullback-Leibler divergence.

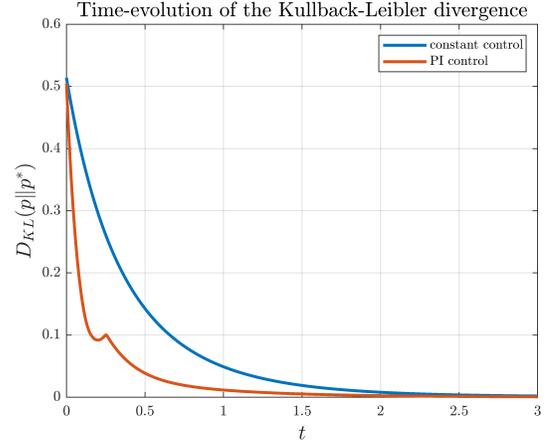
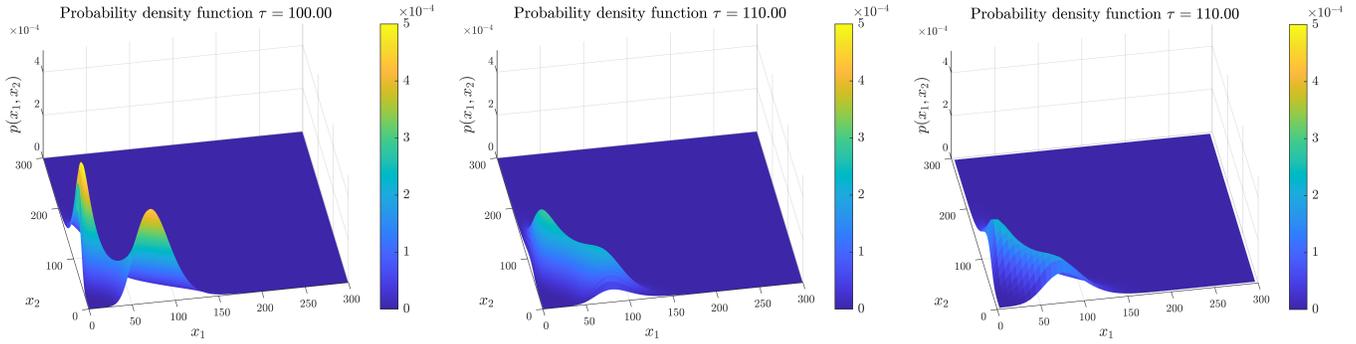


Fig. 3: Time-evolution of the Kullback-Leibler divergence of the PI control and the constant control.

Figure 4 shows the estimated reachability set of the system discretized on a  $50 \times 50$  mesh. For 200 evenly spaced control values  $I_1 \in [0, 50]$  and  $I_2 \in [0, 1]$  we compute and plot the expected values of the protein molecules. Each point has color represented with an RGB triplet, where the green channel is constant and the red and blue channels correspond to  $I_1$  and  $I_2$ , respectively. The black polygon in the background is the filled boundary polygon of the computed points.

## V. CONCLUSIONS

We introduced a modified finite volume method for the semi-discretization of the PIDE model of gene regulatory



(a) Open-loop system on  $300 \times 300$  mesh. (b) Closed-loop system on  $300 \times 300$  mesh. (c) Closed-loop system on  $50 \times 50$  mesh.

Fig. 2: PI control of the genetic toggle switch on various mesh sizes with prescribed expected values  $m_1^* = 41$  and  $m_2^* = 55$ .

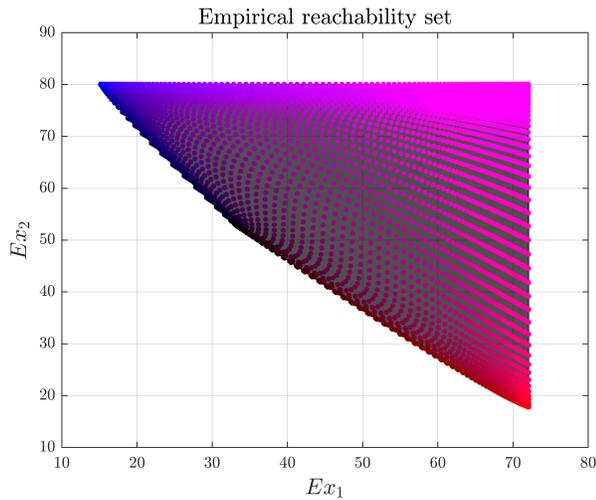


Fig. 4: Empirical reachability set computed on a mesh of size  $50 \times 50$  for 200 equidistant control values  $I_1 \in [0, 50]$  and  $I_2 \in [0, 1]$ .

networks that is suitable for population level control based on certain statistical measures. We computed a CFL condition for admissible meshes and made crucial observations regarding the system, such as the possibility of applying constant control-measuring control performance using Kullback-Leibler divergence and mesh-independent control. These facts allow us to efficiently estimate the reachability of the system. The results were demonstrated through a PI control of the classical gene toggle switch. While the feedback gains were set manually the kinetic semi-discretization can be used to algorithmically tune a PI(D) controller through linearization.

#### ACKNOWLEDGEMENTS

I.O.M. acknowledges MCIN/AEI/10.13039/501100011033 funding through grant PID2021-127888NA-I00 (COMPSYNBIO). G.S. acknowledges the support of the Hungarian National Research, Development and Innovation Office (NKFIH) through the grant 131545. M.A.V. acknowledges the support of the ÚNKP-23-3-II-PPKE-81 National Excellence Program of the Hungarian

Ministry for Innovation and Technology.

#### REFERENCES

- [1] P. Smolen, D. A. Baxter, and J. H. Byrne, “Mathematical modeling of gene networks,” *Neuron*, vol. 26, no. 3, pp. 567–580, 2000.
- [2] D. del Vecchio, J. Dy, and Y. Qian, “Control theory meets synthetic biology,” *The Royal Society Interface*, vol. 13, no. 120, p. 20160380, 2016.
- [3] C. Briat, A. Gupta, and M. Khammash, “Antithetic Integral Feedback Ensures Robust Perfect Adaptation in Noisy Biomolecular Networks,” *Cell Systems*, vol. 2, no. 1, pp. 14–26, 2016.
- [4] J.-B. Lugagne, S. S. Carrillo, M. Kirch, A. Köhler, G. Batt, and P. Hersen, “Balancing a genetic toggle switch by real-time feedback control and periodic forcing,” *Nature Communications*, vol. 8, no. 1, 2017.
- [5] N. Friedman, L. Cai, and X. S. Xie, “Linking Stochastic Dynamics to Population Distribution: An Analytical Framework of Gene Expression,” *Physical Review Letters*, vol. 97, no. 16, 2006.
- [6] V. Elgart, T. Jia, A. T. Fenley, and R. Kulkarni, “Connecting protein and mRNA burst distributions for stochastic models of gene expression,” *Physical Biology*, vol. 8, no. 4, p. 046001, 2011.
- [7] M. Pájaro, A. A. Alonso, I. Otero-Muras, and C. Vázquez, “Stochastic modeling and numerical simulation of gene regulatory networks with protein bursting,” *Journal of Theoretical Biology*, vol. 421, pp. 51–70, 2017.
- [8] M. Pájaro, I. Otero-Muras, C. Vázquez, and A. A. Alonso, “SELANSI: a toolbox for simulation of stochastic gene regulatory networks,” *Bioinformatics*, vol. 34, no. 5, pp. 893–895, 2017.
- [9] P. Bokes and A. Singh, “Controlling Noisy Expression Through Auto Regulation of Burst Frequency and Protein Stability,” in *Hybrid Systems Biology*, M. Češka and N. Paoletti, Eds. Cham: Springer International Publishing, 2019, pp. 80–97.
- [10] C. Fernández, H. Faquir, M. Pájaro, and I. Otero-Muras, “Feedback control of stochastic gene switches using PIDE models,” *IFAC-PapersOnLine*, vol. 55, no. 18, pp. 62–67, 2022.
- [11] M. A. Vághy, I. Otero-Muras, M. Pájaro, and G. Szederkényi, “A Kinetic Finite Volume Discretization of the Multidimensional PIDE Model for Gene Regulatory Networks,” *Bulletin of Mathematical Biology*, vol. 86, no. 2, 2024.
- [12] J. A. Cañizo, J. A. Carrillo, and M. Pájaro, “Exponential equilibration of genetic circuits using entropy methods,” *Journal of Mathematical Biology*, vol. 78, pp. 373–411, 1 2019.
- [13] J. A. Jacquez and C. P. Simon, “Qualitative theory of compartmental systems,” *SIAM Review*, vol. 35, no. 1, pp. 43–79, 1993.
- [14] D. Angeli, “A Tutorial on Chemical Reaction Network Dynamics,” *European Journal of Control*, vol. 15, no. 3-4, pp. 398–406, 2009.
- [15] H. Maeda, S. Kodama, and Y. Ohta, “Asymptotic behavior of nonlinear compartmental systems: Nonoscillation and stability,” *IEEE Transactions on Circuits and Systems*, vol. 25, no. 6, pp. 372–378, 1978.
- [16] D. J. C. MacKay, *Information Theory, Inference and Learning Algorithms*. Cambridge University Press, 2003.
- [17] P. Hartman, *Ordinary Differential Equations*. Society for Industrial and Applied Mathematics, 2002.