




A Kinetic Finite Volume Discretization of the Multidimensional PIDE Model for Gene Regulatory Networks

Mihály A. Vághy¹  · Irene Otero-Muras² · Manuel Pájaro³ · Gábor Szederkényi^{1,4}

Received: 21 April 2023 / Accepted: 22 December 2023 / Published online: 22 January 2024
© The Author(s) 2024

Abstract

In this paper, a finite volume discretization scheme for partial integro-differential equations (PIDEs) describing the temporal evolution of protein distribution in gene regulatory networks is proposed. It is shown that the obtained set of ODEs can be formally represented as a compartmental kinetic system with a strongly connected reaction graph. This allows the application of the theory of nonnegative and compartmental systems for the qualitative analysis of the approximating dynamics. In this framework, it is straightforward to show the existence, uniqueness and stability of equilibria. Moreover, the computation of the stationary probability distribution can be traced back to the solution of linear equations. The discretization scheme is presented for one and multiple dimensional models separately. Illustrative computational examples show the precision of the approach, and good agreement with previous results in the literature.

Keywords Gene regulatory networks · Partial integro-differential equations · Compartmental systems · Kinetic systems · Numerical methods · Finite volume methods

✉ Mihály A. Vághy
vaghymihaly.andras@itk.ppke.hu

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

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

³ Department of Mathematics, Escola Superior de Enxeñaría Informática, University of Vigo, Campus Ourense, 32004 Ourense, Spain

⁴ Systems and Control Laboratory, ELKH Institute for Computer Science and Control (SZTAKI), Kende u. 13-17, Budapest 1111, Hungary

1 Introduction

Gene expression is a fundamental biological process of actually realizing DNA information in the form of proteins in living organisms. Therefore, the (quantitative) modeling of gene expression has been in the focus of research during the last decades (Smolen et al. 2000; Ay and Arnosti 2011). Gene regulatory networks (GRNs) are complex mechanisms through which cells are able to react to internal and external signals in a controlled way (Peter and Davidson 2015). The set of techniques successfully applied for the modeling of GRNs is really wide (De Jong 2002; Karlebach and Shamir 2008; Barbuti et al. 2020). It was pointed out already in the 1970s that the stochastic nature of gene expression has to be taken into consideration during modeling (Berg 1978). Experimental results and model analysis clearly show that both translational and transcriptional bursting contribute to stochasticity in prokaryote and eukaryote gene expression (Kaern et al. 2005; Dar et al. 2015). It is also known that in many cases, stochasticity in gene expression is functionally advantageous, and it can even result in robust phenotypes (MacNeil and Walhout 2011). In Mackey and Tyranskińska (2015), the dynamics of a bistable genetic switch was analyzed, and it was shown that the qualitative properties of the stationary distribution are not significantly influenced by stochastic terms describing bursting and degradation. A theoretically sound hybrid modelling and simulation framework was proposed and applied to study the transcriptional dynamics of an autoregulation loop and a toggle switch in Bokes et al. (2013). An interesting computation result is that there exist two qualitatively different scenarios called "fluctuating" and "simultaneous" in the behaviour of the toggle switch in the presence of gene expression noise.

The dynamical model studied in this paper is originated in Friedman et al. (2006), where a symbolic approach is proposed for describing the stationary distribution of protein concentration in living cells in the form of partial integro-differential equations (PIDEs). The model is based on the master equation, and considers protein production in random bursts (see, also Elgart et al. 2011; Dar 2012) extended by transcription autoregulation. Feasible stationary distributions for this PIDE model with a slightly modified transcription rate were derived and classified in Pájaro et al. (2015). The generalized Friedman (or multidimensional PIDE) model was later introduced in Pájaro et al. (2017) which describes the operation of a genetic circuit of n genes expressed into n different protein types. Since finding symbolic solutions for the stationary distributions of the generalized Friedman model is not straightforward due to its generality, Pájaro et al. (2017) proposed a numerical procedure for the computation. The approach is based on a semilagrangian method for the discretization of the PIDE, and the computational results show that it is suitable to describe the behaviour of a wide class of GRNs with several different regulatory interactions and protein degradation rates. The generalized Friedman model and the subsequently developed simulation framework SELANSI (Pájaro et al. 2017) has since been widely used for design (Sequeiros et al. 2022), identification (Sequeiros et al. 2022) and control (Bokes and Singh 2019; Fernández et al. 2022) of GRNs. In Pájaro et al. (2019) a truncated version of the master equation corresponding to a special version of the one-dimensional Friedman model was proposed. As we will show later, this can be formally seen as a semi-

discretization of the PIDE and can be generalized to both variable degradation rates and multidimensional GRNs.

Compartmental models are used to describe the transmission of modeled entities (material, particles, vehicles, people, information, etc.) between physical or abstract storage units, called compartments (Haddad et al. 2010). Accordingly, the application field of compartmental systems is really wide including (bio)chemistry, pharmacokinetics, ecological, epidemiological and transportation modeling (Godfrey 1983). Mathematically, the descriptive power of compartmental models is quite good as they are capable of representing numerous complex dynamical phenomena (Rap 1986; Brown 1980). Additionally, one of their main benefits is the associated directed graph structure (called the compartmental graph) representing possible transitions, and the related strong results on qualitative dynamical properties (Jacquez and Simon 1993; Cobelli and Romanin-Jacur 1976). Naturally, there is a vast literature on the dynamics of linear compartmental systems and the algebraic properties of the corresponding compartmental matrices (see, e.g. Anderson 2013; Berman and Plemmons 1994; Meshkat et al. 2015; Bortner and Meshkat 2022). Another important related family of models is the class of chemical reaction networks (CRNs) or kinetic systems which includes dynamical models that can be formally represented by a set of transformations (reactions) between abstract chemical complexes. The theory of CRNs is an intensively studied area, where deep results have been achieved on the properties of kinetic systems such as the structure and stability of equilibria (Feinberg 2019; Angeli 2009; Chellaboina et al. 2009). Linear kinetic systems with one-species complexes are trivially compartmental, where the compartments correspond to chemical species, and the reaction graph is identical to the compartmental graph (Érdi and Tóth 1989).

Many hyperbolic conservation laws are derived in an integral form, which, in the case of sufficiently smooth solutions and fluxes, can be rewritten in their usual differential form (LeVeque 2002). However, many practical problems involve discontinuous solutions, where shocks or singularities can develop quickly even from smooth initial data. We highlight that the (generalized) Friedman model can have this problem too, see Sect. 5 for a special single gene example where the symbolic solution is known and has a singularity in the origin. Thus, numerical methods derived from the differential form, such as finite differences, are expected to lose accuracy near discontinuities. This problem can for example be mitigated by an appropriate Finite Volume Method (FVM) based on the integral form of the PDE. Instead of computing possibly unreliable pointwise approximations we define grid cells and approximate the cell averages of the solution. This approach introduces a clear compartmental interpretation of semi-discretized PDEs and can naturally capture the underlying conservation law, too (Eymard and Gallouët 2000). As a consequence, in many applications the result of an FVM is guaranteed to be nonnegative and conservative, thus it is bounded; that is, FVM can be robust to the singularities of the solution.

Motivated by the above mentioned preliminary results, the aim of this paper is to propose an efficient novel computational approach based on compartmental discretization for the numerical solution of the multidimensional PIDE model introduced in Pájaro et al. (2017). It is shown that the dynamics of the approximating model can be written as a nonnegative linear time-invariant ODE. Therefore, the stationary distribution can be computed by solving a set of linear equations. The structure of

the paper is the following. In the next section we introduce the PIDE model and the compartmental and kinetic systems. Next, in Sect. 3, we describe the kinetic finite volume discretization proposed to solve the PIDE model. A qualitative analysis of the proposed method is addressed in Sect. 4 and several illustrative numerical examples are shown in Sect. 5. Finally, in Sect. 6, we end up with a summary of the work.

2 Notations and Background

In this section we give a brief introduction of multidimensional GRNs and (linear) compartmental and kinetic systems.

2.1 Multidimensional Gene Regulatory Networks

The following short introduction is based on Friedman et al. (2006), Pájaro et al. (2017). 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 still relevant central dogma of molecular biology, which asserts that the gene instructions are transcribed into messenger RNAs, that are translated into proteins. The continuous number of mRNA molecules and proteins are denoted by $\mathbf{m}, \mathbf{x} \in \mathbb{R}^n$, 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 the binding of proteins. Note that in general, the feedback mechanism may require the binding of multiple types of proteins besides the one expressed by the given gene. For the sake of generality, we assume that any protein can repress or activate any gene in the network. This mechanism is typically modelled 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 process occurring with rate k_m^i per unit time and with transcriptional leakage $\varepsilon_i \in (0, 1)$. Then the transcription can be written as

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

where $c_i : \mathbb{R}_+^n \rightarrow [\varepsilon_i, 1]$ depend on the feedback regulation mechanism. In a single gene setting we could consider

$$c_1(x_1) = \frac{K^H + \varepsilon_1 x_1^H}{K^H + x_1^H},$$

for some appropriate K and H constants; see Sect. 5 for more examples of c_i Hill functions. We emphasize that the transcriptional leakage directly affects the range of these coefficient functions. 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 $\gamma_x^i : \mathbb{R}_+^n \rightarrow \mathbb{R}_+$. Following Pájaro et al. (2017) 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 $x_i > y_i$ 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, $p(t, \mathbf{x})$, can be modelled with the following PIDE:

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

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

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

In Cañizo et al. (2019) 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:

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

$$\int_{\mathbb{R}_+^n} p(t, \mathbf{x}) d\mathbf{x} = \int_{\mathbb{R}_+^n} p_0(\mathbf{x}) d\mathbf{x}, \quad \forall t \geq 0.$$

In fact, if the initial data is in the space $\mathcal{C}^{1,b}(\mathbb{R}_+^n)$ of Hölder-continuous functions 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.

2.2 Compartmental and Kinetic Systems

We briefly introduce compartmental systems based on Jacquez and Simon (1993). Compartmental differential equations are often used to model physical phenomena

governed by a conservation law such as conservation of mass. A compartment can represent 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. Despite this, 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 of 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. Loop flows are not allowed, i.e. $i \neq j$ in F_{ij} . Then the time-evolution of the system is given by the following system of differential equations:

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

We impose the following physical assumptions to the system:

1. for any $i, j, t \geq 0, i \neq j$ we have that $F_{ij}(q(t)) \geq 0, I_i(t) \geq 0$ and $F_{0i}(q(t)) \geq 0$,
2. for any $i, t \geq 0$ if $q_i(t) = 0$, then $F_{0i}(q(t)) = F_{ji}(q(t)) = 0$ 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} belong to the class \mathcal{C}^k , then we can rewrite (2) as

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

where $F_{ij} = f_{ij} q_j$ and the fractional transfer coefficients f_{ij} belong to the class \mathcal{C}^{k-1} , for $k \geq 1$. We can then naturally rewrite (3) in matrix form as

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

where $\{f\}_{ij} = f_{ij}$ for $i = 0, 1, \dots$ and $j = 1, 2, \dots$. 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 Angeli (2009). For each compartment with index i , q_i represents the mass (or alternatively, the concentration) of the one-species 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 4.2.

3 Kinetic Finite Volume Discretization

In this section we formulate a finite volume discretization of (1), the result of which is a mass conservative kinetic system. We also note that since (1) is linear (that is, if p and q are solutions, then so is $p + q$), the result of the semi-discretization is anticipated to also be linear.

3.1 One-Dimensional Case

Let us first consider the one-dimensional Friedman model describing the temporal evolution of protein distribution given by

$$\frac{\partial p(t, x)}{\partial t} = \frac{\partial}{\partial x} [\gamma_x^1(x)xp(t, x)] + k_m^1 \int_0^x \beta_1(x - y)c_1(y)p(t, y)dy, \tag{4}$$

with initial condition $p(0, x) = p_0(x)$ that has integral one. The mass conservation of (4) is well-known but the subsequent informal investigation provides further insight that can be transferred to the design of the numerical scheme. Since ω_1 is the probability density function of an exponential distribution its integral is one, and thus

$$\int_0^\infty \beta_1(x)dx = 0.$$

Integrating over \mathbb{R}_+ shows after a change of variables that

$$\begin{aligned} \int_0^\infty \frac{\partial p(t, x)}{\partial t} dx &= \frac{\partial}{\partial t} \int_0^\infty p(t, x)dx = \int_0^\infty \frac{\partial}{\partial x} [\gamma_x^1(x)xp(t, x)] dx \\ &+ k_m^1 \int_0^\infty \int_0^x \beta_1(x - y)c_1(y)p(t, y)dydx = \underbrace{\lim_{x \rightarrow \infty} \gamma_x^1(x)xp(t, x) - \gamma_x^1 \cdot 0 \cdot p(t, 0)}_{=0} \\ &+ k_m^1 \int_0^\infty \int_y^\infty \beta_1(x - y)c_1(y)p(t, y)dx dy \\ &= k_m^1 \int_0^\infty c_1(y)p(t, y) \int_y^\infty \beta_1(x - y)dx dy = 0, \end{aligned} \tag{5}$$

so that the equality

$$\int_0^\infty p(t, x)dx = \int_0^\infty p_0(x)dx = 1$$

holds for any $t \geq 0$.

In a finite volume setting the coefficients are calculated as averages (that is, integrals) over appropriate subdomains. Hence, as an intuition we should note that the mass conservation property of the novel scheme should be the result of a calculation very similar to (5).

Our main goal is to perform a spatial discretization (with resolution h) to obtain an infinite dimensional dynamical system describing the temporal evolution of the functions $p_i(t)_{i \in \mathbb{N}}$ with the usual properties of a PDF; that is, we should have that:

1. $0 \leq p_i(t)$ for all $i \in \mathbb{N}$ and $t \geq 0$,
2. $\sum_{i=1}^{\infty} h p_i(t) = 1$ for all $t \geq 0$.

In order to do so, consider the set of intervals

$$K_i = \left[x_{i-\frac{1}{2}}, x_{i+\frac{1}{2}} \right] = [(i-1)h, ih], \quad i = 1, 2, \dots$$

for some $h > 0$ and introduce the set of variables $p_i(t)$, where

$$p_i(t) \approx \frac{1}{|K_i|} \int_{K_i} p(t, y) dy = \frac{1}{h} \int_{K_i} p(t, y) dy;$$

that is, the value $p_i(t)$ is assumed to approximate the average in the cell K_i and we set the initial values accordingly. Further introduce the cell averages of the functions γ_x^1 and c_1 given as

$$\gamma_i^1 = \frac{1}{|K_i|} \int_{K_i} \gamma_x^1(y) dy, \quad c_i^1 = \frac{1}{|K_i|} \int_{K_i} c_1(y) dy.$$

Let x_i be the midpoint of K_i for $i = 1, 2, \dots$ and define

$$b_{i,i}^1 = \frac{1}{h/2} \int_{[(i-1)h, (i-1/2)h]} \beta_1(x_i - y) dy = \frac{1}{h/2} \int_{[x_i-h/2, x_i]} \beta_1(x_i - y) dy,$$

$$b_{i,j}^1 = \frac{1}{|K_j|} \int_{K_j} \beta_1(x_i - y) dy, \quad j = 1, 2, \dots, i-1.$$

As the $b_{i,j}^1$ coefficients come from a discretization of an exponential function, the series is geometric (apart from $b_{i,i}^1$, that takes the Dirac delta into account), see Berg (1978), Bokes et al. (2011), McAdams and Arkin (1997). As the derivative on the right-hand side of (4) describes protein degradation (that is, a vector field pointing towards the origin) we will approximate it with a difference quotient of the form

$$\frac{\partial}{\partial x} \left[\gamma_x^1(x) x p(t, x) \right] \Big|_{K_i} \approx \frac{1}{h} \left(\gamma_{i+1}^1 x_{i+\frac{1}{2}} p_{i+1}(t) - \gamma_i^1 x_{i-\frac{1}{2}} p_i(t) \right).$$

Then approximating the integral in (4) with a sum yields the system

$$\begin{aligned} \dot{p}_i(t) &= \frac{1}{h} \left(\gamma_{i+1}^1 x_{i+\frac{1}{2}} p_{i+1}(t) - \gamma_i^1 x_{i-\frac{1}{2}} p_i(t) \right) + k_m^1 \sum_{j=1}^i h_{i,j}^1 b_{i,j}^1 c_j^1 p_j(t); \\ p_i(0) &= \frac{1}{|K_i|} \int_{K_i} p_0(y) dy, \end{aligned} \tag{6}$$

where

$$h_{i,j}^1 = \begin{cases} h/2, & i = j, \\ h, & i \neq j. \end{cases}$$

Observe, that the resulting infinite dimensional system (6) is clearly a linear donor controlled compartmental system of the form

$$\dot{p}(t) = \Gamma p(t),$$

where the infinite matrix defined element-wise as

$$\Gamma_{ij} = \begin{cases} k_m^1 h_{i,j}^1 b_{i,j}^1 c_j^1, & j < i, \\ -\frac{1}{h} \gamma_i^1 x_{i-\frac{1}{2}} + k_m^1 h_{i,i}^1 b_{i,i}^1 c_i^1, & j = i, \\ \frac{1}{h} \gamma_{i+1}^1 x_{i+\frac{1}{2}}, & j = i + 1, \\ 0, & j > i + 1 \end{cases}$$

is an infinite Kirchoff matrix; that is, it is a Metzler matrix with zero column-sums. While the superdiagonal elements of Γ can be arbitrarily large, its sign structure guarantees the well-posedness, in the ℓ^1 space of absolutely summable sequences, of the underlying continuous-time infinite Markov chain of (6), see Sects. 4 and 6 of Reuter (1957). This further implies the well-posedness of (6), and thus relying on the fact that β_1 integrates to zero we have that

$$\begin{aligned} \sum_{i=1}^{\infty} \dot{p}_i(t) &= \sum_{i=1}^{\infty} \frac{1}{h} \left(\gamma_{i+1}^1 x_{i+\frac{1}{2}} p_{i+1}(t) - \gamma_i^1 x_{i-\frac{1}{2}} p_i(t) \right) + k_m^1 \sum_{i=1}^{\infty} \sum_{j=1}^i h_{i,j}^1 b_{i,j}^1 c_j^1 p_j(t) \\ &= \lim_{l \rightarrow \infty} \frac{1}{h} \gamma_{l+1}^1 x_{l+\frac{1}{2}} p_{l+1}(t) - \frac{1}{h} \gamma_1^1 \cdot 0 \cdot p_1(t) + k_m^1 \sum_{j=1}^{\infty} \sum_{i=j}^{\infty} h_{i,j}^1 b_{i,j}^1 c_j^1 p_j(t) \\ &= k_m^1 \sum_{j=1}^{\infty} c_j^1 p_j(t) \sum_{i=j}^{\infty} h_{i,j}^1 b_{i,j}^1 \\ &= k_m^1 \sum_{j=1}^{\infty} c_j^1 p_j(t) \left(\int_{[x_j-h/2, x_j]} \beta_1(x_j - y) dy + \sum_{i=j+1}^{\infty} \int_{K_j} \beta_1(x_i - y) dy \right) \end{aligned}$$

$$\begin{aligned}
 &= k_m^1 \sum_{j=1}^{\infty} c_j^1 p_j(t) \left(\int_0^{h/2} \beta_1(y) dy + \sum_{i=j+1}^{\infty} \int_{[(i-j-1/2)h, (i-j+1/2)h]} \beta_1(y) dy \right) \\
 &= k_m^1 \sum_{j=1}^{\infty} c_j^1 p_j(t) \int_0^{\infty} \beta_1(y) dy = 0,
 \end{aligned}$$

so that the equality

$$\sum_{i=1}^{\infty} h p_i(t) = \sum_{i=1}^{\infty} h p_i(0) = 1$$

holds for any $t \geq 0$. The above facts combined also show that $p_i(t) \leq \frac{1}{h}$ for any $t \geq 0$.

Remark 1 The above calculation shows that the conservativity of the scheme only requires that the $b_{i,j}^1$ coefficients are averages of the β_1 function; that is, the γ_i^1 and c_i^1 coefficients could be just point values of the underlying coefficient functions on the mesh points. Nevertheless, taking the averages might be more precise if those functions change rapidly.

3.2 Multidimensional Case

Let us consider the multidimensional model (1) with $n > 1$. Define the positive step sizes h_1, h_2, \dots, h_n and sets

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

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

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

For $i = 1, 2, \dots, n$ we also compute the variables

$$\begin{aligned}
 \gamma_{\alpha}^i &= \frac{1}{h} \int_{K_{\alpha}} \gamma_x^i(\mathbf{y}) dy, \\
 c_{\alpha}^i &= \frac{1}{h} \int_{K_{\alpha}} c_i(\mathbf{y}) dy.
 \end{aligned}$$

Let $x_{\alpha} = [x_{\alpha}^1 \ x_{\alpha}^2 \ \dots \ x_{\alpha}^n]^T$ be the midpoint (w.r.t. each dimension) of K_{α} 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_{α} . For $i = 1, 2, \dots, n$ define

$$b_{\alpha,\alpha_i}^i = \frac{1}{h_i/2} \int_{[(i-1)h_i, (i-1/2)h_i]} \beta_i(x_\alpha^i - y) dy = \frac{1}{h_i/2} \int_{[x_\alpha^i - h_i/2, x_\alpha^i]} \beta_i(x_\alpha^i - y) dy,$$

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

Similarly to the one-dimensional case the derivatives are approximated with difference quotients of the form

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

Approximating the integrals in (1) with sums as before, yields the system

$$\begin{aligned} \dot{p}_\alpha(t) &= \sum_{i=1}^n \frac{1}{h_i} \left(\gamma_{\alpha+e_i}^i x_\alpha^{i+\frac{1}{2}} p_{\alpha+e_i}(t) - \gamma_\alpha^i x_\alpha^{i-\frac{1}{2}} p_\alpha(t) \right) \\ &\quad + \sum_{i=1}^n k_m^i \sum_{j=1}^{\alpha_i} h_{\alpha,j}^i b_{\alpha,j}^i c_{\alpha_i,j}^i p_{\alpha_i,j}(t); \end{aligned} \tag{7}$$

$$p_\alpha(0) = \frac{1}{h} \int_{K_\alpha} p_0(\mathbf{y}) d\mathbf{y},$$

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

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

Again, the system is clearly kinetic and the mass conservation follows from a calculation very similar to the one-dimensional case:

$$\begin{aligned} \sum_\alpha \dot{p}_\alpha(t) &= \sum_\alpha \sum_{i=1}^n \frac{1}{h_i} \left(\gamma_{\alpha+e_i}^i x_\alpha^{i+\frac{1}{2}} p_{\alpha+e_i}(t) - \gamma_\alpha^i x_\alpha^{i-\frac{1}{2}} p_\alpha(t) \right) \\ &\quad + \sum_\alpha \sum_{i=1}^n k_m^i \sum_{j=1}^{\alpha_i} h_{\alpha,j}^i b_{\alpha,j}^i c_{\alpha_i,j}^i p_{\alpha_i,j}(t) \\ &= \sum_\alpha \sum_{i=1}^n k_m^i \sum_{j=\alpha_i}^\infty h_{\alpha_i,j}^i b_{\alpha_i,j}^i c_{\alpha_i,j}^i p_\alpha(t) = \sum_{i=1}^n k_m^i \sum_\alpha c_{\alpha_i}^i p_\alpha(t) \underbrace{\sum_{j=\alpha_i}^\infty h_{\alpha_i,j}^i b_{\alpha_i,j}^i}_{=0} = 0. \end{aligned}$$

This shows for any $t \geq 0$ that

$$\sum_\alpha h p_\alpha(t) = \sum_\alpha h p_\alpha(0) = 1,$$

further implying that $p_\alpha(t) \leq \frac{1}{h}$ for each α .

3.3 Discretization on a Truncated Domain

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 $\lim_{\|x\|_{\mathbb{R}^n} \rightarrow \infty} p(t, \mathbf{x}) = 0$ 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 $\int_{\Omega} p_0(\mathbf{x}) d\mathbf{x} = 1$.

We divide the $(0, L_i)$ intervals into N_i equal subintervals and proceed to calculate the variables $p_{\alpha}(0)$ and the coefficients γ_{α}^i and c_j^i as before. We similarly compute $b_{\alpha,j}^i$ for $j = 1, 2, \dots, \alpha_i - 1$, but modify b_{α,α_i}^i to capture the fact that the number of i th kind of protein is maximalized in L_i .

Note, that the resulting system can still be given by (7) with the difference that the set of variables $\{p_{\alpha}\}$ is finite. While the bursts and degradations inherently define some “spatial” structure between the p_{α} variables (discussed in detail later), it might be more useful to think of the truncated semi-discretized model as a flattened N -dimensional system of the form

$$\dot{p}(t) = \tilde{\Gamma}^{(N)} p(t) \quad \text{with } N := \prod_{i=1}^n N_i. \tag{8}$$

4 Qualitative Analysis

In this section we show that the result of the truncated kinetic finite volume discretization is not only a mass conservative nonnegative system but it has several advantageous qualitative properties.

4.1 Structural Descriptions

While we could rely on the linearity of (8) to investigate its dynamical behaviour, the large number of variables and the complexity of the coefficient matrix $\tilde{\Gamma}^{(N)}$ renders this approach futile. Instead, let us focus on the inner structure of the system through its compartmental and CRN representations. These observations will immediately imply most qualitative properties of interest.

Compartmental representation

Consider the N -dimensional truncated system of the form (7). Based on the burst and degradation structure the system has a compartmental topology as follows:

- Each compartment K_{α} has an incoming edge from $K_{\alpha+e_i}$ due to protein degradation if $\alpha_i < N_i$ for $i = 1, 2, \dots, n$.
- Each compartment K_{α} has an incoming edge from $K_{\alpha_i,j}$ for $i = 1, 2, \dots, n$ and $j = 1, 2, \dots, \alpha_i - 1$ due to protein production in bursts.

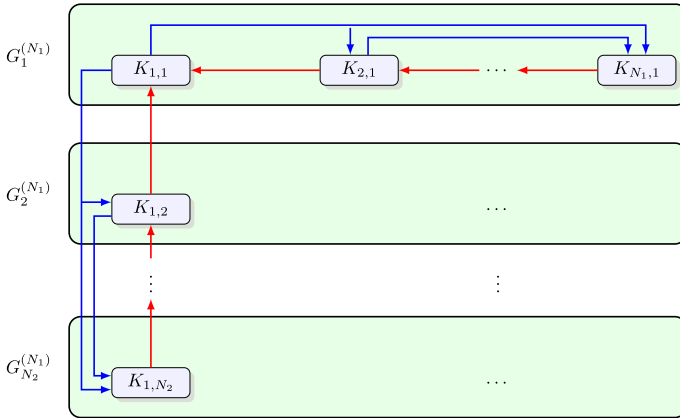


Fig. 1 Compartmental topology of a two-dimensional model. Each subsystem is isomorphic to that of a one-dimensional model

Clearly, the compartmental topology is strongly connected, which property is essential for our further analysis. Based on this structure (and the flattening method) one can easily determine the elements of the matrix $\tilde{\Gamma}^{(N)} \in \mathbb{R}^{N \times N}$ of (8).

To gain further insight into the compartmental topology, let us focus on some low-dimensional (in terms of the PIDE) examples. Figure 1 shows the structure of compartments for a two-dimensional PIDE. Degradations and bursts are denoted with red and blue arrows, respectively. Let $G^{(N_1, N_2)}$ denote the graph in Fig. 1; that is, a compartmental graph of appropriate size corresponding to (8). Notice, that the graph $G^{(N_1, N_2)}$ can be decomposed to the interconnected $G_1^{(N_1)}, G_2^{(N_1)}, \dots, G_{N_2}^{(N_1)}$ graphs that are isomorphic to the compartmental graph of a one-dimensional model of size N_1 . This shows that $G^{(N_1, N_2)}$ is isomorphic to the Cartesian product $G^{(N_1)} \times G^{(N_2)}$. In fact, the $G^{(N_1, N_2, \dots, N_n)}$ compartmental graph of an n -dimensional model (8) is isomorphic to $\times_{i=1}^n G^{(N_i)}$.

CRN representation

For each continuous variable p_α we introduce the species P_α and assign the complex P_α to the compartment K_α . Then the complex composition matrix containing the stoichiometric coefficients of the complexes as its columns is the identity matrix $I \in \mathbb{R}^{N \times N}$, and the reaction structure is identical to the above compartmental topology; that is, the reaction graph is identical (isomorphic) to the compartmental graph and, in particular, is strongly connected. This readily shows that the deficiency of the reaction graph, as defined in CRN theory (Feinberg 2019), is zero as there are N distinct complexes, one linkage class and a spanning tree in the reaction graph of size $N - 1$. Since the system is linear, the reaction vectors corresponding to the edges of the spanning tree span the stoichiometric subspace. Hence the deficiency is indeed $\delta = N - 1 - (N - 1) = 0$.

4.2 Long Time Behaviour

Asymptotic stability

By standard results on compartmental systems, since the truncated system (8) is strongly connected, there is a unique positive equilibrium (that is, a stationary PDF) $\bar{p} \in \mathbb{R}_+^N$ that attracts every admissible initial value (Maeda et al. 1978, Theorem 6).

Remark 2 As a conclusion of the above assertions a mass-action CRN can be assigned to the truncated conservative system (8) whose reaction graph is strongly connected and has deficiency zero. thus, the same assertion follows from CRN theory and, in particular, from the deficiency zero theorem (Feinberg 2019; Anderson 2011).

Furthermore, we also know that the system is Lyapunov stable with the standard entropy-like logarithmic Lyapunov function

$$V(p, \bar{p}) = \sum_{\alpha} \left(p_{\alpha} \log \frac{p_{\alpha}}{\bar{p}_{\alpha}} + \bar{p}_{\alpha} - p_{\alpha} \right).$$

Finally, a well-known result (Maeda and Kodama 1979) shows that for two solutions $p(t)$ and $q(t)$ of (8), the following inequality holds:

$$\|p(t_1) - q(t_1)\|_{L^1} \leq \|p(t_2) - q(t_2)\|_{L^1} \quad t_1 \geq t_2 \geq 0.$$

In particular, if we set $q = \bar{p}$ this shows that the convergence to the unique equilibrium is monotone in the L^1 norm.

4.3 Computing the Equilibrium

We can easily approximate \bar{p} by simulating the system on an appropriately large time interval. However, such a simulation can be computationally expensive and it is not trivial to determine the necessary time interval. Furthermore, in many applications we may not be interested in the time evolution of the system, only in the equilibrium \bar{p} . Instead, relying on the linear nature of the system (8) we may explicitly compute the equilibrium with the following approach.

We can incorporate the conservation into the equilibrium condition as

$$\hat{\Gamma}^{(N)} \bar{p} = \begin{bmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} =: e_1 \tag{9}$$

where $\hat{\Gamma}^{(N)}$ is obtained from $\tilde{\Gamma}^{(N)}$ by replacing the first row with $h\mathbf{1}_N^T \in \mathbb{R}^N$. Since $\tilde{\Gamma}^{(N)}$ has a one-dimensional left kernel (by virtue of the rank-nullity theorem and the fact that zero is a simple eigenvalue, see Foster and Jacquez 1975) spanned by $\mathbf{1}_N$, any

$N - 1$ rows are independent. To see this, assume by contradiction that not any $N - 1$ rows are independent. Then there exists a nonzero vector in the left kernel of $\tilde{\Gamma}^{(N)}$ that has a zero coordinate, but then the left kernel cannot be spanned by $\mathbf{1}_N$. Clearly $\mathbf{1}_N$ is not in the left kernel of $\hat{\Gamma}^{(N)}$ and $\text{Im}\tilde{\Gamma}^{(N)} \subsetneq \text{Im}\hat{\Gamma}^{(n)}$, and thus $\text{rank}\hat{\Gamma}^{(N)} = N$, hence we can find the equilibrium \bar{p} by simply solving the linear system of equations (9).

5 Numerical Experiments

In this section we present some examples from the literature and discuss the memory requirement of the kinetic FVM.

5.1 Examples

In this section we compare the performance of our method to that of SELANSI (Pájaro et al. 2017). For more information about the examples the reader is referred to Pájaro et al. (2017). The numerical simulations have been performed on a computer with Intel(R) Core(TM) i7-8565U CPU @ 1.80GHz and 16 GB of RAM in MATLAB R2022b. The solution (9) is solved with built-in iterative solvers. The final time and time step of the SELANSI simulations are noted for each example.

Example 1: Single Gene Self-Regulation with Positive Feedback

The first example is a GRN consisting of a single gene. The regulation is described by the Hill function

$$c_1(x_1) = \frac{K_1^{H_{11}} + \varepsilon_1 x_1^{H_{11}}}{K_1^{H_{11}} + x_1^{H_{11}}}.$$

We consider a negative Hill coefficient, corresponding to a positive self-regulation. In this case, as described in Pájaro et al. (2015) (see also Friedman et al. 2006) the stationary solution of (4) can be explicitly calculated as follows:

$$\bar{p}(x) = C \rho^{\frac{k_m^1(1-\varepsilon_1)}{H_{11}}}(x) x^{-(1-k_m^1\varepsilon_1)} e^{-\frac{x}{b_1}},$$

where $\rho(x) = \frac{x^{H_{11}}}{K_1^{H_{11}} + x_1^{H_{11}}}$ and $C > 0$ is a constant ensuring that $\bar{p}(x)$ integrates to one. As discussed before, the solution has a singularity at $x = 0$, since it has a factor of $x^{-(1-k_m^1\varepsilon_1)}$. Clearly the exponent is larger than -1 , hence the average of the solution over the first mesh cell will be finite.

Figure 2 shows the simulation results for various L_1 values. Observe that if the domain size is set ideally (Fig. 2a), then both approaches provide a good approximation of the symbolic solution. However, SELANSI cannot handle cases where the domain is too small or too large and skews the solution, see the highlighted section of Fig. 2c. This is assumed to be because (i) SELANSI imposes zero Dirichlet boundary conditions,

(ii) the semilagrangian method is not conservative, thus SELANSI has to manually renormalize in each iteration. Compared to this, our kinetic FVM is nonnegative and conservative and, as noted in Sect. 4, the equilibrium is strictly positive. Thus it can capture the qualitative behaviour of the PIDE even if the domain is not known precisely, which might be the case for previously untested gene regulatory network structures or parameter sets, see Fig. 2b. Table 1 shows the relative error (in the L^2 norm) of the different methods compared to the symbolic solution, computed as follows:

$$E(p, p_{ref}) = \frac{\|p - p_{ref}\|_{L^2}}{\|p_{ref}\|_{L^2}} = \frac{\sqrt{\sum_{i=1}^N (p(x_i) - p_{ref}(x_i))^2}}{\sqrt{\sum_{i=1}^N p_{ref}^2(x_i)}}.$$

The computational times of both methods are depicted in Table 2, from where we can observe that the FVM has also better computational efficiency compared to that of SELANSI.

Example 2: Mutual Activation of Two Genes

In this example we consider Hill functions in the form of

$$c_1(\mathbf{x}) = \frac{K_{12}^{H_{12}} + \varepsilon_1 x_2^{H_{12}}}{K_{12}^{H_{12}} + x_2^{H_{12}}},$$

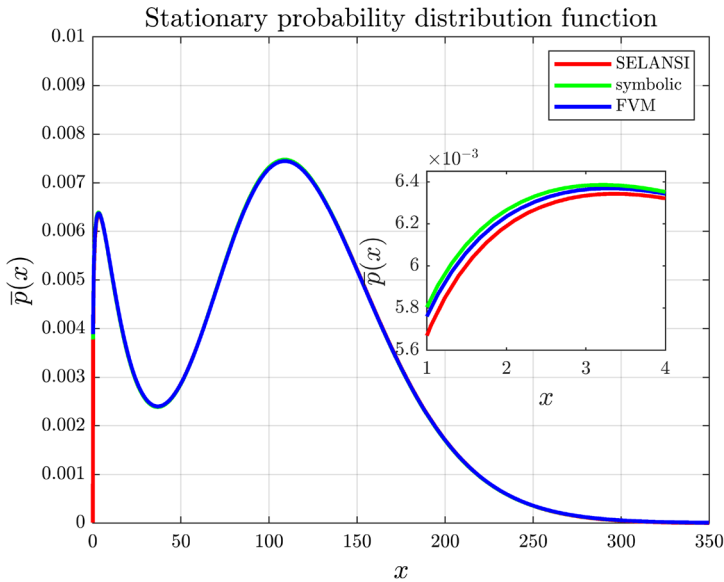
$$c_2(\mathbf{x}) = \frac{K_{21}^{H_{21}} + \varepsilon_2 x_1^{H_{21}}}{K_{21}^{H_{21}} + x_1^{H_{21}}},$$

where $H_{12} < 0$ and $H_{21} < 0$, corresponding to positive cross-regulation or activation. Figure 3a, b respectively show the stationary joint PDF, while Fig. 3c, d respectively show the stationary marginal PDF on various domains, for FVM and SELANSI. Note, that the GRN is symmetric w.r.t. the proteins, thus we only plot one set of marginal PDFs. We can observe the sensitivity of SELANSI to the domain, while the finite volume discretization is quite robust to it. In this example we can see that the solution computed by SELANSI deteriorates not just for too small, but even for too large domains. Since for multidimensional GRNs the symbolic solution of (1) cannot be computed in a straightforward manner, we cannot compute the empirical error as in the case of the one-dimensional example. Instead, we only compare the running times of the two methods, the results of which are collected in Table 3.

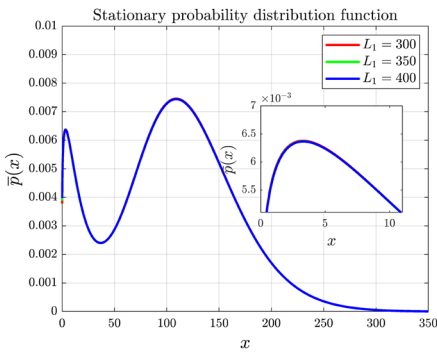
Example 3: Mutual Repression of Two Genes

In this example we consider Hill functions in the form of

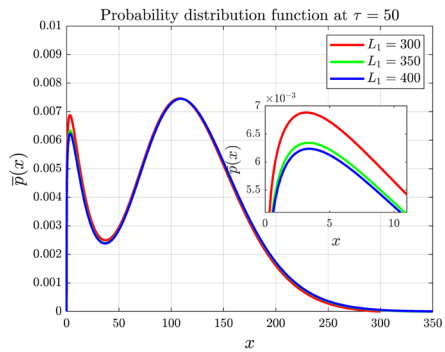
$$c_1(\mathbf{x}) = \frac{K_{12}^{H_{12}} + \varepsilon_1 x_2^{H_{12}}}{K_{12}^{H_{12}} + x_2^{H_{12}}}, \quad c_2(\mathbf{x}) = \frac{K_{21}^{H_{21}} + \varepsilon_2 x_1^{H_{21}}}{K_{21}^{H_{21}} + x_1^{H_{21}}},$$



(a) $L_1 = 350$



(b) PDF with kinetic FVM



(c) PDF with SELANSI

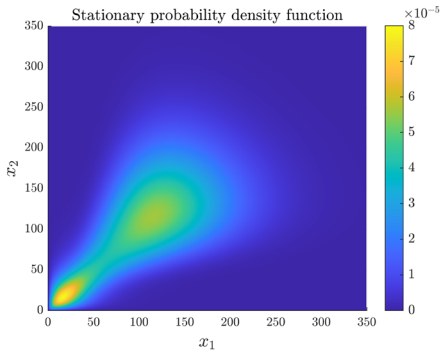
Fig. 2 Self-regulated single gene network with parameters $H_{11} = -4$, $K_1 = 45$, $\varepsilon_1 = 0.15$, $k_m^1 = 3.2 \times 10^{-3}$, $b_1 = 16$ and $\gamma_x^1(x) = 4 \times 10^{-4}$. The simulations are performed with $N_1 = 2000$, $\tau = t\gamma_x^1 = 50$ and $\Delta t = 0.002$ (color figure online)

Table 1 Relative error of the simulation of a one-dimensional GRN on various domains

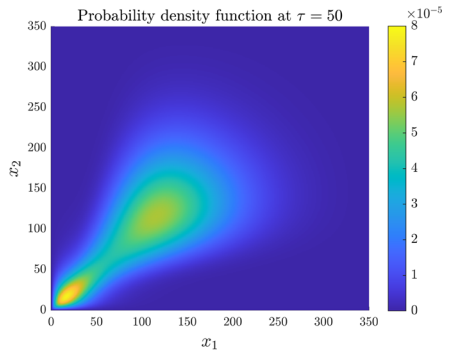
Mesh	FVM ($\times 10^{-3}$)			SELANSI ($\times 10^{-3}$)		
	$L_1 = 300$	$L_1 = 350$	$L_1 = 400$	$L_1 = 300$	$L_1 = 350$	$L_1 = 400$
$2.5 \times 10^4 \times 800$	7.4358	8.6799	9.9106	29.6010	5.4881	9.2314
$2.5 \times 10^4 \times 1200$	4.9886	5.8169	6.6366	28.6252	4.3113	8.3388
$2.5 \times 10^4 \times 1600$	3.7665	4.3870	5.0013	29.5308	3.4255	8.0746
$2.5 \times 10^4 \times 2000$	3.0335	3.5298	4.0209	29.9408	2.8899	7.2511
$2.5 \times 10^4 \times 5000$	1.4271	1.4761	1.6675	31.0859	2.5960	6.0042

Table 2 Average runtime of 100 simulations of a one-dimensional GRN with various mesh sizes

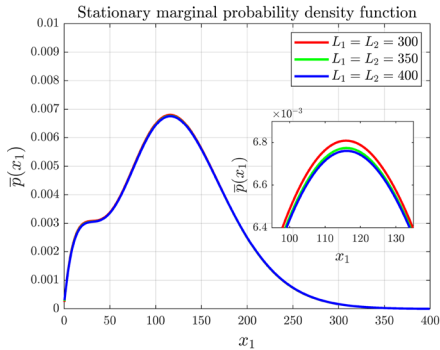
Mesh	FVM	SELANSI
$2.5 \times 10^4 \times 800$	0.0335 s	1.8208 s
$2.5 \times 10^4 \times 1200$	0.0899 s	2.1993 s
$2.5 \times 10^4 \times 1600$	0.1460 s	2.5510 s
$2.5 \times 10^4 \times 2000$	0.3561 s	2.8866 s
$2.5 \times 10^4 \times 5000$	3.8066 s	7.9131 s



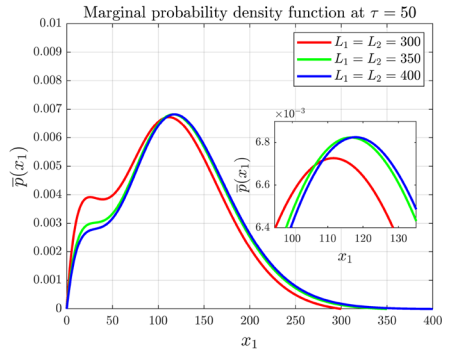
(a) Kinetic FVM with $L_1 = L_2 = 350$.



(b) SELANSI with $L_1 = L_2 = 350$.



(c) Marginal PDF with kinetic FVM.



(d) Marginal PDF with SELANSI.

Fig. 3 Mutual activation with parameters $H_{12} = H_{21} = -4$, $K_{12} = K_{21} = 70$, $\varepsilon_1 = \varepsilon_2 = 0.2$, $k_m^1 = k_m^2 = 3.4 \times 10^{-3}$, $b_1 = b_2 = 18$, $\gamma_x^1(x) = \gamma_x^2(x) = 4 \times 10^{-4}$, $N_1 = N_2 = 400$, $\tau = t\gamma_x^1 = 50$ and $\Delta t = 0.005$ (color figure online)

where $H_{12} > 0$ and $H_{21} > 0$, corresponding to negative cross-regulation or repression. Figure 4 shows the stationary joint PDF and the marginal stationary PDF on multiple domains. Again, the GRN is symmetric w.r.t. the proteins and the same dependence on the domain can be observed in the case of SELANSI. The running time of both methods with various mesh sizes are presented in Table 3.

Table 3 Average runtime of 100 simulations of several two-dimensional GRNs with various mesh sizes

GRN	FVM				SELANSI			
	100 ²	200 ²	300 ²	400 ²	100 ²	200 ²	300 ²	400 ²
Ex. 2	0.2223 s	1.6485 s	5.2159 s	11.5707 s	9.0423 s	21.5449 s	40.6726 s	71.1792 s
Ex. 3	0.2128 s	1.3679 s	4.5709 s	10.3817 s	9.0443 s	21.3982 s	40.7456 s	72.3397 s
Ex. 4	0.3241 s	1.8639 s	5.7925 s	12.3953 s	9.1355 s	22.0259 s	41.9351 s	73.4001 s
Ex. 5	0.7064 s	4.9921 s	17.0973 s	45.0809 s	–	–	–	–

Example 4: Self and Mutual Regulation

In this example we consider two genes, one of which is activated by both, the other is repressed by both. The corresponding Hill functions can be given as follows:

$$c_1(\mathbf{x}) = \frac{\varepsilon_{11}x_1^{H_{11}}x_2^{H_{12}} + \varepsilon_{12}K_{11}^{H_{11}}x_2^{H_{12}} + \varepsilon_{13}x_1^{H_{11}}K_{12}^{H_{12}} + K_{11}^{H_{11}}K_{12}^{H_{12}}}{x_1^{H_{11}}x_2^{H_{12}} + K_{11}^{H_{11}}x_2^{H_{12}} + x_1^{H_{11}}K_{12}^{H_{12}} + K_{11}^{H_{11}}K_{12}^{H_{12}}},$$

$$c_2(\mathbf{x}) = \frac{\varepsilon_{21}x_1^{H_{21}}x_2^{H_{22}} + \varepsilon_{23}K_{21}^{H_{21}}x_2^{H_{22}} + \varepsilon_{12}x_2^{H_{21}}K_{22}^{H_{22}} + K_{21}^{H_{21}}K_{22}^{H_{22}}}{x_1^{H_{21}}x_2^{H_{22}} + K_{21}^{H_{21}}x_2^{H_{22}} + x_2^{H_{21}}K_{22}^{H_{22}} + K_{21}^{H_{21}}K_{22}^{H_{22}}},$$

where $H_{11} < 0$, $H_{21} < 0$, $H_{12} > 0$ and $H_{22} > 0$. We note that the above functions are generalized Hill functions, and thus have to be defined in a separate file for the SELANSI simulation. Figure 5 shows the stationary joint PDF and the marginal stationary PDF on multiple domains. This example is not symmetric w.r.t. the different kind of proteins, thus we plot both marginal density functions. The running time of both methods with various mesh sizes are presented in Table 3.

Example 5: Bacterial Competence

In *Bacillus subtilis*, competence is a probabilistic and transiently differentiated state. In this physiological state bacteria has the capability of DNA uptake from their environment. The phenomena is modelled with a two-dimensional gene regulatory network, consisting of the master regulator self-activated ComK which represses the transcription factor ComS (Süel et al. 2007). Protein degradation is mediated by the MecA complex. After ComK (ComS) binds to the complex an intermediate complex MecA-ComK (MecA-ComS) complex is formed, in which the ComK (ComS) protein is degraded by the ClpP-ClpC proteases (Süel et al. 2006). Instead of explicitly modelling the effects of the MecA complex, the authors consider a variable degradation rate. Using this reduced order stochastic differential equation developed in Süel et al. (2006) a discrete stochastic CME model is presented in Dandach and Khammash (2010), simulated using a Monte-Carlo based Stochastic Simulation Algorithm. A corresponding PIDE is presented in Pájaro et al. (2017) with parameters adapted from the CME model of Dandach and Khammash (2010) as follows: $\alpha_k = 0.0028$, $\beta_k = 0.049$, $\beta_s = 0.057$, $K_k = 100$, $K_s = 110$, $\delta_k = \delta_s = 0.0014$, $\Gamma_k = 500$,

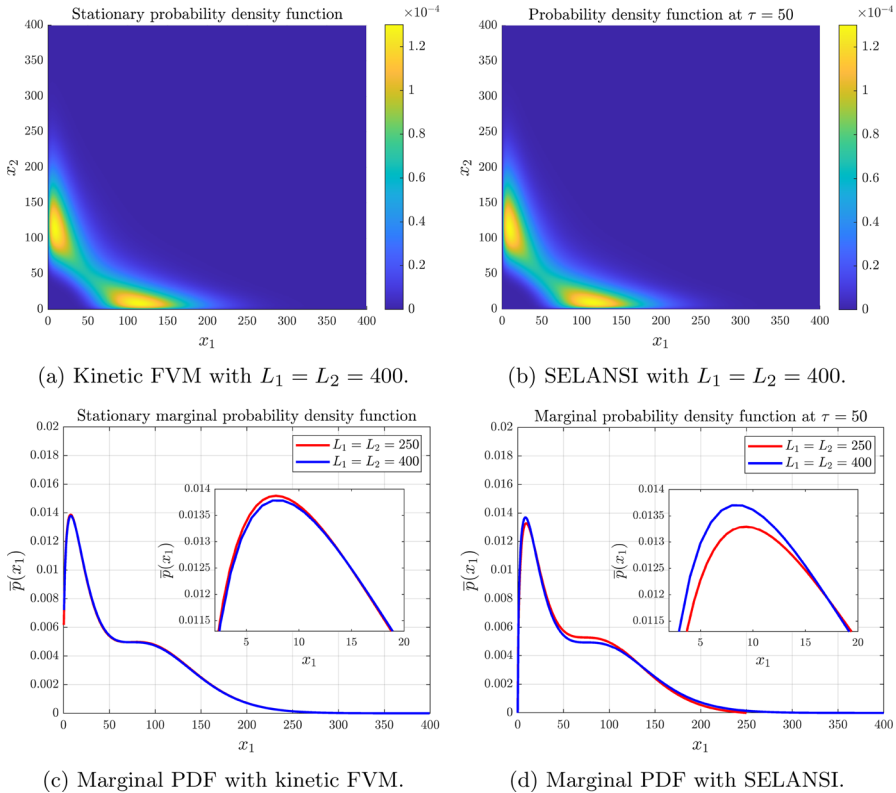


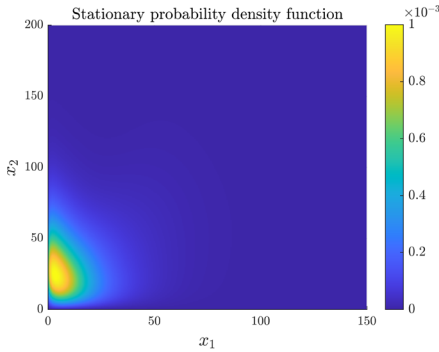
Fig. 4 Mutual repression with parameters $H_{12} = H_{21} = 4$, $K_{12} = K_{21} = 45$, $\varepsilon_1 = \varepsilon_2 = 0.15$, $k_m^1 = k_m^2 = 3.2 \times 10^{-3}$, $b_1 = b_2 = 16$, $\gamma_x^1(x) = \gamma_x^2(x) = 4 \times 10^{-4}$, $N_1 = N_2 = 400$, $\tau = \tau_{\gamma_x^1} = 50$ and $\Delta t = 0.005$ (color figure online)

$\Gamma_s = 50$, $b_1 = 2$, $b_2 = 5$, $k_m^1 = \frac{\alpha_k + \beta_k}{b_1}$, $k_m^2 = \frac{\beta_s}{b_2}$, $\varepsilon_1 = \frac{\alpha_k}{\alpha_k + \beta_k}$, $\varepsilon_2 = 0$, $H_{11} = -2$, $H_{21} = 5$. The coefficient functions are set as:

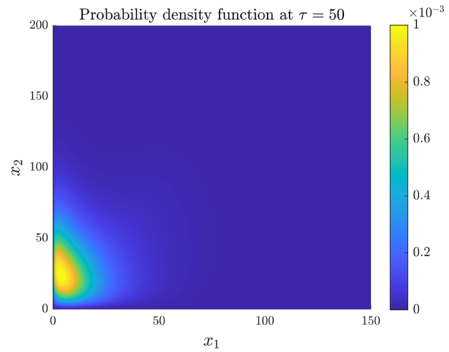
$$c_1(\mathbf{x}) = \frac{K_k^{H_{11}} + \varepsilon_1 x_1^{H_{11}}}{K_k^{H_{11}} + x_1^{H_{11}}}, \quad \gamma_x^1(\mathbf{x}) = \frac{\delta_k \Gamma_k \Gamma_s}{\Gamma_k \Gamma_s + \Gamma_s x_1 + \Gamma_k x_2},$$

$$c_2(\mathbf{x}) = \frac{K_s^{H_{21}} + \varepsilon_2 x_1^{H_{21}}}{K_s^{H_{21}} + x_1^{H_{21}}}, \quad \gamma_x^2(\mathbf{x}) = \frac{\delta_s \Gamma_k \Gamma_s}{\Gamma_k \Gamma_s + \Gamma_s x_1 + \Gamma_k x_2}.$$

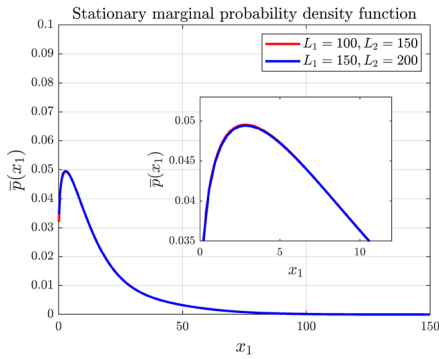
We note that the currently publicly available SELANSI version cannot handle variable degradation rates, thus we could not reproduce the plots of Pájaro et al. (2017). Figure 6 shows the stationary PDFs and its contour plot, both of which are in accordance with the plots of Pájaro et al. (2017). The running times of the kinetic FVM for various mesh sizes are shown in Table 3.



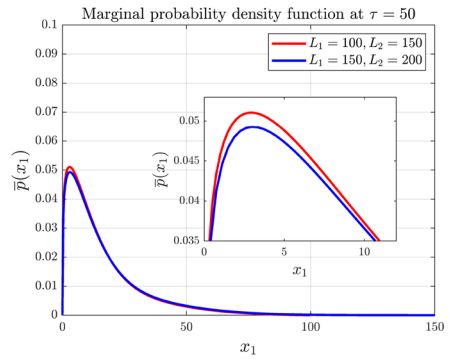
(a) Kinetic FVM with $L_1 = 150, L_2 = 200$.



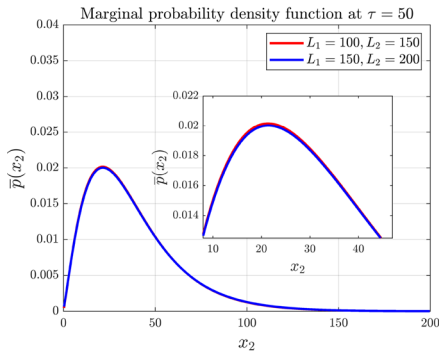
(b) SELANSI with $L_1 = 150, L_2 = 200$.



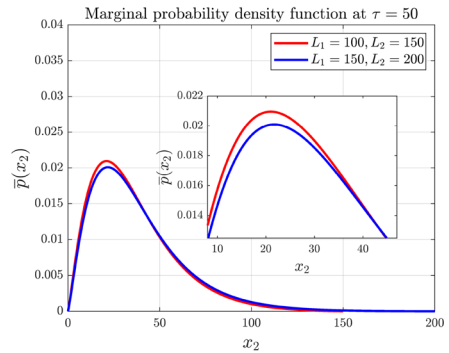
(c) Marginal PDF with kinetic FVM.



(d) Marginal PDF with SELANSI.



(e) Marginal PDF with kinetic FVM.



(f) Marginal PDF with SELANSI.

Fig. 5 Mutual repression with parameters $H_{11} = -4, H_{21} = -6, H_{12} = H_{22} = 2, K_{11} = K_{12} = 45, K_{21} = K_{22} = 70, \varepsilon_{11} = \varepsilon_{21} = 0.002, \varepsilon_{12} = 0.02, \varepsilon_{22} = 0.1, \varepsilon_{13} = \varepsilon_{23} = 0.2, k_m^1 = 4 \times 10^{-3}, k_m^2 = 8 \times 10^{-3}, b_1 = 10, b_2 = 20, \gamma_x^1(x) = \gamma_x^2(x) = 4 \times 10^{-4}, N_1 = N_2 = 400, \tau = t\gamma_x^1 = 50$ (color figure online)

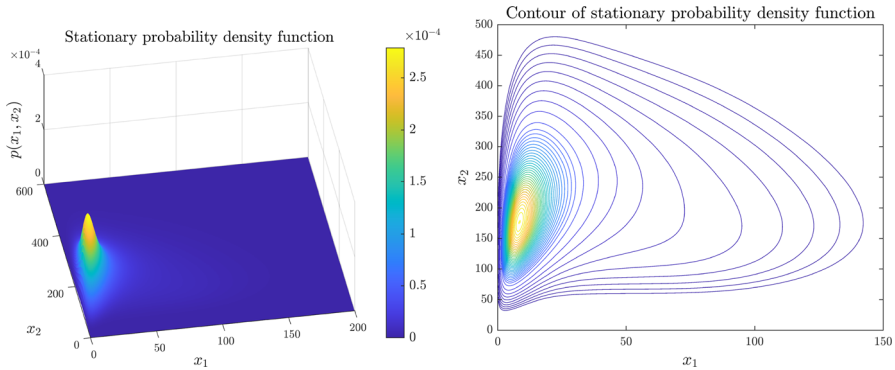


Fig. 6 Kinetic FVM of Example 5 with $N_1 = N_2 = 400$ (color figure online)

5.2 Memory Requirement of the Kinetic FVM

A notable technical challenge in our method is the efficient assembly and storage of the coefficient matrix $\Gamma^{(N)}$. For $n \geq 2$ one should store $\Gamma^{(N)}$ in a sparse representation, but even then the memory requirement can grow quickly. However, we can explicitly calculate the number of nonzero elements of the matrix to aid the design of the simulation. To be precise, the number of nonzero elements of the coefficient matrix corresponding to an n -dimensional PIDE discretized on a grid of size $\prod_{i=1}^n N_i$ is as follows:

$$\begin{aligned}
 & \sum_{i=1}^n \underbrace{\left(\sum_{k=1}^{N_i} (N_i - k) \right)}_{\text{bursting of protein } X_i} \prod_{\substack{j=1 \\ j \neq i}}^n N_j + \sum_{i=1}^n \underbrace{(N_i - 1)}_{\text{degradation of } X_i} \prod_{\substack{j=1 \\ j \neq i}}^n N_j + \underbrace{\prod_{i=1}^n N_i}_{\text{diagonal}} \\
 &= \sum_{i=1}^n \frac{1}{2} (N_i^2 - N_i) \prod_{\substack{j=1 \\ j \neq i}}^n N_j + n \prod_{i=1}^n N_i - \sum_{i=1}^n \prod_{\substack{j=1 \\ j \neq i}}^n N_j + \prod_{i=1}^n N_i \\
 &= -\frac{1}{2} n \prod_{i=1}^n N_i + \frac{1}{2} \left(\sum_{i=1}^n N_i \right) \left(\prod_{i=1}^n N_i \right) + (n + 1) \prod_{i=1}^n N_i - \left(\sum_{i=1}^n \frac{1}{N_i} \right) \left(\prod_{i=1}^n N_i \right) \\
 &= \frac{1}{2} \left(n + 2 + \sum_{i=1}^n N_i - \sum_{i=1}^n \frac{2}{N_i} \right) \left(\prod_{i=1}^n N_i \right).
 \end{aligned}$$

Figure 7 shows the number of nonzero elements on an equidistant grid for a matrix corresponding to a mesh of size $N = 10^{10}$ (that is, the matrix has 10^{20} total elements) as a function of n . The logarithmic scaling suggests that as the dimension of the PIDE is increased, we can increase the number of finite volume cells on the grid even without exceeding the memory limits.

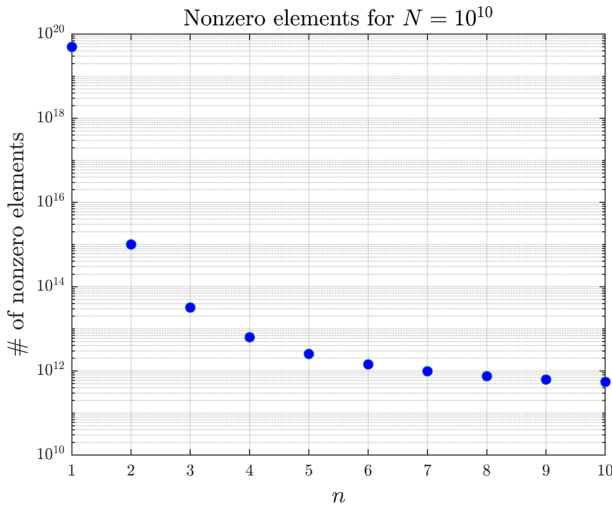


Fig. 7 Memory requirement of an n -dimensional GRN with $N = 10^{10}$ total number of cells

6 Conclusions

A novel discretization scheme was proposed in this paper for the simulation and analysis of multidimensional PIDE models used in the stochastic dynamical description of gene regulatory networks. It was shown that using an appropriate finite volume scheme, a fully conservative linear compartmental dynamics is obtained in ODE form. This representation can be particularly useful for studying the dynamics of the process from a systems biology perspective, since the interpretation of the system of ODEs of a process is usually more intuitive than the original PIDE. In particular, the proposed compartmental description can serve as a basis for designing experiments and solving various analysis and control problems of stochastic gene regulatory networks, since the theoretical properties of the model class are well-known regardless of the dimension of the state variable (van Kampen 2007). The interconnection structure of the discretized system was studied in detail, and it was shown that the associated directed graph is always strongly connected. Therefore, the theory of kinetic and compartmental systems can be used to conclude that the equilibrium of the discretized dynamics representing the stationary distribution of molecules is unique and globally stable for any biologically meaningful parameter values in the PIDE model. Moreover, the stationary distribution can be obtained by solving a set of linear equations without performing the time-domain simulation. The memory requirement of the method can be precisely pre-computed based on which the applicable resolution can be determined. Five illustrative examples were presented to show the operation and performance of the method. Whenever possible, the obtained solutions and running times were compared with those given by the SELANSI toolbox, and these comparisons clearly justified the advantageous properties of the proposed approach both in terms of precision and performance. Further work will be focused on feedback control design based on the semidiscretized models.

Acknowledgements The authors acknowledge the helpful discussions with M. Kovács. M.P. acknowledges funds from the Galician Government (grants ED431C 2022/47 and ED431G 2019/01, both including FEDER financial support). I.O.M. acknowledges MCIN/AEI/10.13039/501100011033 funding through grant PID2021-127888NA-I00 (COMPSYNBIO). G.Sz. acknowledges the support of the Hungarian National Research, Development and Innovation Office (NKFIH) through the grant 145934. M.A.V. acknowledges the support of the ÚNKP-22-3-I-PPKE-72 National Excellence Program of the Hungarian Ministry for Innovation and Technology from the source the NFKIH.

Funding Open access funding provided by Pázmány Péter Catholic University.

Data Availability All data generated or analysed during this study are included in this published article. Our implementation along with the examples found in this manuscript can be found at the following GitHub repository: [mihalyvagy/PIDE](https://github.com/mihalyvagy/PIDE).

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Anderson DF (2011) A proof of the global attractor conjecture in the single linkage class case. *SIAM J Appl Math* 71:1487–1508
- Anderson DH (2013) *Compartmental modeling and tracer kinetics*, vol 50. Springer, Berlin
- Angeli D (2009) A tutorial on chemical reaction network dynamics. *Eur J Control* 15:398–406
- Ay A, Arnosti DN (2011) Mathematical modeling of gene expression: a guide for the perplexed biologist. *Crit Rev Biochem Mol Biol* 46:137–151
- Barbuti R, Gori R, Milazzo P, Nasti L (2020) A survey of gene regulatory networks modelling methods: from differential equations, to Boolean and qualitative bioinspired models. *J Membr Comput* 2:207–226
- Berg OG (1978) A model for the statistical fluctuations of protein numbers in a microbial population. *J Theor Biol* 71:587–603
- Berg OG (1978) A model for the statistical fluctuations of protein numbers in a microbial population. *J Theor Biol* 71:587–603
- Benman A, Plemmons RJ (1994) *Nonnegative matrices in the mathematical sciences*. SIAM, Philadelphia
- Bokes P, Singh A (2019) Controlling noisy expression through auto regulation of burst frequency and protein stability. In: Češka M, Paoletti N (eds) *Hybrid systems biology*. Springer, Cham, pp 80–97
- Bokes P, King JR, Wood ATA, Loose M (2011) Multiscale stochastic modelling of gene expression. *J Math Biol* 65:493–520
- Bokes P, King JR, Wood ATA, Loose M (2013) Transcriptional bursting diversifies the behaviour of a toggle switch: hybrid simulation of stochastic gene expression. *Bull Math Biol* 75:351–371
- Bortner C, Meshkat N (2022) Identifiable paths and cycles in linear compartmental models. *Bull Math Biol* 84:53
- Brown RF (1980) *Compartmental system analysis: state of the art*. IEEE Trans Biomed Eng BME-27:1–11
- Cañizo JA, Carrillo JA, Pájaro M (2019) Exponential equilibration of genetic circuits using entropy methods. *J Math Biol* 78:373–411
- Chellaboina V, Bhat SP, Haddad WM, Bernstein DS (2009) Modeling and analysis of mass-action kinetics. *IEEE Control Syst* 29:60–78
- Cobelli C, Romanin-Jacur G (1976) Controllability, observability and structural identifiability of multi input and multi output biological compartmental systems. *IEEE Trans Biomed Eng BME-23:93–100*
- Dandach SH, Khammash M (2010) Analysis of stochastic strategies in bacterial competence: a master equation approach. *PLoS Comput Biol* 6:e1000985

- Dar RD et al (2012) Transcriptional burst frequency and burst size are equally modulated across the human genome. *Proc Natl Acad Sci* 109:17454–17459
- Dar RD, Razooky BS, Weinberger LS, Cox CD, Simpson ML (2015) The low noise limit in gene expression. *PLoS ONE* 10:e0140969
- De Jong H (2002) Modeling and simulation of genetic regulatory systems: a literature review. *J Comput Biol* 9:67–103
- Elgart V, Jia T, Fenley AT, Kulkarni R (2011) Connecting protein and mRNA Burst distributions for stochastic models of gene expression. *Phys Biol* 8:046001
- Érdi P, Tóth J (1989) *Mathematical models of chemical reactions: theory and applications of deterministic and stochastic models*. Manchester University Press, Manchester
- Eymard R, Gallouët T, Herbin R (2000) *Finite methods*. Elsevier
- Feinberg M (2019) *Foundations of Chemical Reaction Network Theory*. Springer, Cham
- Fernández C, Faquir H, Pájaro M, Otero-Muras I (2022) Feedback control of stochastic gene switches using PIDE models. *IFAC-PapersOnLine* 55:62–67
- Foster DM, Jacquez JA (1975) Multiple zeros for eigenvalues and the multiplicity of traps of a linear compartmental system. *Math Biosci* 26:89–97
- Friedman N, Cai L, Xie XS (2006) Linking stochastic dynamics to population distribution: an analytical framework of gene expression. *Phys Rev Lett* 97:168302
- Godfrey K (1983) *Compartmental models and their application*. Academic Press, London
- Haddad WM, Chellaboina V, Hui Q (2010) *Nonnegative and compartmental dynamical systems*. Princeton University Press, Princeton
- Jacquez JA, Simon CP (1993) Qualitative theory of compartmental systems. *SIAM Rev* 35:43–79
- Kaern M, Elston TC, Blake WJ, Collins JJ (2005) Stochasticity in gene expression: from theories to phenotypes. *Nat Rev Genet* 6:451–464
- Karlebach G, Shamir R (2008) Modelling and analysis of gene regulatory networks. *Nat Rev Mol Cell Biol* 9:770–780
- LeVeque RJ (2002) *Finite volume methods for hyperbolic problems*. Cambridge University Press, Cambridge
- Mackey MC, Tyrant-Kamińska M (2015) The limiting dynamics of a bistable molecular switch with and without noise. *J Math Biol* 73:367–395
- MacNeil LT, Walhout AJ (2011) Gene regulatory networks and the role of robustness and stochasticity in the control of gene expression. *Genome Res* 21:645–657
- Maeda H, Kodama S (1979) Some results on nonlinear compartmental systems. *IEEE Trans Circuits Syst* 26:203–204
- Maeda H, Kodama S, Ohta Y (1978) Asymptotic behavior of nonlinear compartmental systems: nonoscillation and stability. *IEEE Trans Circuits Syst* 25:372–378
- McAdams H, Arkin A (1997) Stochastic mechanisms in gene expression. *Proc Natl Acad Sci* 94:814–819
- Meshkat N, Sullivant S, Eisenberg M (2015) Identifiability results for several classes of linear compartment models. *Bull Math Biol* 77:1620–1651
- Pájaro M, Alonso AA, Vázquez C (2015) Shaping protein distributions in stochastic self-regulated gene expression networks. *Phys Rev E* 92:032712
- Pájaro M, Alonso AA, Otero-Muras I, Vázquez C (2017) Stochastic modeling and numerical simulation of gene regulatory networks with protein bursting. *J Theor Biol* 421:51–70
- Pájaro M, Otero-Muras I, Vázquez C, Alonso AA (2017) SELANSI: a toolbox for simulation of stochastic gene regulatory networks. *Bioinformatics* 34:893–895
- Pájaro M, Otero-Muras I, Vázquez C, Alonso AA (2019) Transient hysteresis and inherent stochasticity in gene regulatory networks. *Nat Commun* 10:4581
- Peter IS, Davidson EH (2015) *Genomic control process: development and evolution*. Academic Press, London
- Rap PE (1986) In oscillations and chaos in cellular metabolism and physiological systems. In: Holden AV (ed) *Chaos*, vol 9. Princeton University Press, London, pp 179–208
- Reuter GEH (1957) Denumerable Markov processes and the associated contraction semigroups on l . *Acta Math* 97:1–46
- Sequeiros C, Vázquez C, Banga JR, Otero-Muras I (2022) Automated design of synthetic biocircuits in the stochastic regime. *IFAC-PapersOnLine* 55:630–634

- Sequeiros C, Otero-Muras I, Vazquez C, Banga JR (2022) Global optimization approach for parameter estimation in stochastic dynamic models of biosystems. *IEEE/ACM Trans Comput Biol Bioinf* 20:1971–1982
- Smolen P, Baxter DA, Byrne JH (2000) Mathematical modeling of gene networks. *Neuron* 26:567–580
- Süel GM, Garcia-Ojalvo J, Liberman LM, Elowitz MB (2006) An excitable gene regulatory circuit induces transient cellular differentiation. *Nature* 440:545–550
- Süel GM, Kulkarni RP, Dworkin J, Garcia-Ojalvo J, Elowitz MB (2007) Tunability and noise dependence in differentiation dynamics. *Science* 315:1716–1719
- van Kampen NG (2007) *Stochastic processes in physics and chemistry*, vol 1, 3rd edn. Elsevier

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.