Analysing Biochemical Oscillation through Probabilistic Model Checking

Paolo Ballarini [1] and   Radu Mardare [1] and   Ivan Mura [1]

Abstract

Automated verification of stochastic models has been proved to be an effective technique for the analysis of a large class of stochastically behaving systems. In this paper we show how stochastic model-checking can be effectively applied to the analysis of biological systems. We consider a few models of biological systems taken from the literature, and we consider both their encodings as ordinary differential equations and Markovian models. We show that stochastic model-checking verification of biological systems can complement both deterministic and stochastic simulation techniques when dealing with dynamical properties of oscillators. We demonstrate how stochastic model-checking can provide exact quantitative characterization of properties of systems exhibiting oscillatory behavior, providing insights that cannot be obtained with differential equations models and that would require a large number of runs with stochastic simulation approaches.

Key words: Biological oscillators, Ordinary differential equations,
Markov models, Stochastic model checking, Simulation

1   Introduction

The quantitative analysis of the dynamics of biological systems is a fundamental task in *systems biology*, a new research field that focuses on the systematic study of complex interactions using an integrative approach rather than a reductive one [14]. Even simple biochemical networks of interacting proteins can show surprisingly complex behavior, a behavior that cannot be understood looking at the evolution of the single components but that instead requires a systemic analysis approach.

Traditionally, the study of time-dependent dynamics of biological systems has been addressed with deterministic approaches, based on ordinary differ-

---
[1] The Microsoft Research - University of Trento Centre for Computational and Systems Biology, Piazza Manci 17, Povo-Trento 38100, Italy
e-mail: {ballarini,mardare,mura}@cosbi.eu

ential equation (ODE, hereafter) models. ODEs provide quite an adequate abstraction for capturing the interactions and transformation of biochemical species, and come with a consolidated set of computational tools for model definition and solution [12]. More recently, stemming from the foundational work of Gillespie [10], a discrete approach to the modeling of biochemical systems has been gaining consensus, especially to model systems in which the continuous approximation does not seem to be justified, for instance when only a few molecules exist for some species.

The relationships between the continuous deterministic and the discrete stochastic representation of a given system are quite interesting and still subject of research. When the number of molecules that constitute a molecular network is low, the stochastic modeling may represent a more suitable tool to represent and analyze the dynamics of the system. On the other hand, as the number of molecules grows, abstracting discrete number of molecules into continuous concentration levels and representing evolution of dynamics through a system of coupled ODEs provides very accurate representations and also has the advantage of not suffering from the state-space explosion problem that plagues stochastic modeling tools. Moreover, stochastic models are mostly solved via simulation, which may require performing a substantial number of simulation runs to compute statistically relevant results.

When dealing with biological systems that exhibit complex patterns of behavior, it may be difficult to check whether the dynamics of the system satisfies some interesting properties. One of such properties, quite often vaguely defined, is the oscillatory behavior that characterizes a variety of biological systems [4].Whereas it is comparatively easy to hypothesize the presence of oscillations through visual inspection of simulated time courses, another matter is to precisely verify whether the oscillations will continue forever or rather will definitely stop, either by being progressively damped down or abruptly interrupted. In this respect, the abstraction adopted by ODE models has some advantages, because the deterministic approach ensures that any periodic evolution will indefinitely repeat. However, as we will see in this paper, such properties may appear to hold only because of the continuous approximation. On the other hand, results obtained from discrete stochastic models can still provide clues about the existence of oscillatory behaviors, but the noise introduced by stochastic fluctuations may make them appear less pronounced. Furthermore, when oscillations do exist but their phase, period and amplitude is also varying stochastically, a multi-run analysis of simulation traces over the time domain does not help in characterizing the long-range behavior of the system.

In this paper we consider two examples of models of biochemical systems taken from the literature, which we use as case studies to demonstrate the ad-

vantages of a formal approach to the definition and verification of properties of biological systems, including oscillatory behaviors. The major contribution of this paper is the definition of an approach to the formal specification of properties, based on Probabilistic Computation Tree Logic (PCTL, hereafter) [11] and Continuous Stochastic Logic (CSL) [1]. PCT/CSL formulas can be used to precisely define properties of system dynamics, such as convergence of a variable to a fixed value, bound oscillation around a fixed value, perpetual alternation of growth and decrease phases, relationships among multiple variables. Such formulas define a query-language in which a user can encode a property of the systems he intends to verify. Probabilistic model checkers that accept PCTL formulas defined over the variables of a discrete stochastic model can then be used to verify whether the property holds of the system. In this paper, we show how the PRISM [15] model checker can be used for this purpose. We apply our formal approach to the example models to show how system properties can be characterized and to demonstrate the insights that can be obtained into system dynamics with the verification with PRISM. We show the obvious advantages provided by the exact quantitative solutions returned by a model-checking tool. In contrast, simulation of stochastic models can only provide statistically approximate solutions. Our proposed approach applies to general models of biological systems, its current limitation being the fact that only models having a bounded state space are amenable to analysis.

The rest of this paper is organized as follows: in Section 2 we introduce the basics about biochemical oscillators illustrating a specific system which will be used as the running example throughout the remainder of the paper. In Section 3 we investigate the use of temporal logic as a means to encode relevant features of oscillating systems. In Section 4 we briefly describe probabilistic model checking. In Section 5 we analyse the structural properties of the Markovian model of the our running example, the verification of which by means of the PRISM model checker is illustrated in Section 6. We summarise our contribution in the final section of the paper.

2 Biochemical oscillations

A biochemical system consists of a number of correlated (type of) entities, referred to as *reactants*, and a number of, time consuming, interactions, referred to as *reactions*. Each reactant belongs to a given *species*. Reactions are descriptions of the system's dynamics and they indicate how the production/consumption of reactants is regulated by other reactants. A biochemical system is fully characterised by: a description of the system's reactions (including the timing information of each reaction, referred to as *reaction rate*), and the initial amount of reactants of each species. In this work we are interested in analysing a specific behaviour of certain biochemical systems, namely

oscillation. Informally we say that a biochemical system exhibits an *oscillating behaviour* when the system's reactions result in a (periodic) fluctuation of the level of the reactants, around a given value. To this aim, we consider a specific example taken from the literature, known as the *3-way oscillator*, which we use as the running example of the paper.

2.1 The 3-way oscillator: transient oscillation

We consider a simple system of three biochemical reactions involving three species of molecules, namely: A, B and C. The reactions are positive feedback arranged in a loop fashion and their chemical equation form is:

$$A+B \xrightarrow{r_A} B+B \qquad\qquad B+C \xrightarrow{r_B} C+C \qquad\qquad C+A \xrightarrow{r_C} A+A \quad (2.1)$$
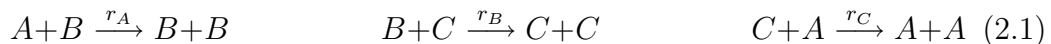
where r_A, r_B and r_C are the kinetic constants of the reactions. The meaning of reactions (2.1) is intuitive. For example, the first equation indicates that a molecule of A and a molecule of B are consumed in order to produce two molecules of B and that happens at a constant rate r_A [^2]. Such system has been introduced in [3] and it has been referred to as the *3-way oscillator* for it can be shown that the level of molecules of each species oscillates in time. Let a_0, b_0 and c_0 be the initial number of molecules of the three species. The system enjoys the following properties: first the total number of molecules in the system $INV = a_0 + b_0 + c_0$ is invariant, as the are not synthesis nor degradation reactions, and because of the stoichiometry of the reactions. Second the system remains *alive* meaning that the system reactions are all possible, as long as all reactants are present: as soon as one species get extinct the system irreversibly enter a terminating path at the end of which INV molecules have been cumulated into either one of the remaining species. For this reason we refer to the oscillation of the 3-way oscillator as *transient* (i.e. it damps off in finite time). We will see how the oscillation of the 3-way oscillator can be turned into a *permanent* one (i.e. one that never stops). Based on our running example we now briefly describe the two standard modeling approaches used in system biology.

2.2 Continuous Deterministic Modeling of Biological Systems

A deterministic model of a system is one in which the parameters and variables are not subject to random fluctuations, so that the system is, at any time, entirely defined by the initial conditions chosen. Biological systems have been extensively modeled with systems of coupled ODEs at various abstractions levels, ranging from molecular interactions [18] up to predator-preys

[^2]: Similar considerations hold for the other reactions.

population evolution [20]. Variables representing biochemical entities of a deterministic model are continuous and their value $\boldsymbol{X}(t)$ at time t represents the amount of the biochemical entities they represent, i.e. the state of the system at that time. The state $\boldsymbol{X}(t_0)$ of the system at time t_0 defines the initial condition. Changes of the system state along time are described by a set of differential equations, in the form $\boldsymbol{X}'(t) = F(\boldsymbol{X}(t))$ with $F : \mathcal{R}^n \to \mathcal{R}^n$, n being the number of variables in the model. The system of ODEs can be solved either analytically or via simulation (numerical integration) to determine the state of the system at any time $t \geq t_0$, and to determine possible stationary solutions. A continuous deterministic model of the 3-way oscillator can be straightforwardly derived from the chemical equations (2.1). It has 3 real-valued variables a, b and c, where $x \in \{a, b, c\}$ represents concentration of species $X \in \{A, B, C\}$. The changes in the amount of species is described by the following system of ODEs:

$$\dot{a} = -r_A ab + r_C ac$$
$$\dot{b} = r_A ba - r_B bc \qquad (2.2)$$
$$\dot{c} = -r_C ac + r_B bc$$

where $\dot{x} = \frac{dx}{dt}$ denotes the first derivative (with respect to time) of $x \in \{a, b, c\}$.

Notice that the rate constants in (2.2) are multiplied by the concentrations of reactants to account for the concurrent execution of multiple instances of the same reaction according to the mass-action law. Given the concentrations of the 3 biochemical species at time $t = t_0$, it is possible to solve the system of ODEs above to obtain the state of the system at time $t \geq t_0$. Many tools exist that can perform this task, some of which are also able to automatically obtain the system of ODEs from the system's chemical equations form. Figure 1(a), Figure 1(b) show solutions [3] of the system (2.2) plotted onto the phase-space and phase-plane respectively. Those plots display that the concentrations of the three species follow indeed a very regular oscillating behaviour evidenced by the solution trajectories forming a neat triangular orbit around the initial concentration point.

We observe that, although a useful method to gain insights about a system behaviour, continuous deterministic modeling has some downsides. For example, as we will see, oscillation for the system (2.1) is transient but such a fundamental characteristic cannot be evinced through ODEs analysis. Furthermore, the system of ODEs (2.2) allows for some trivial fixed points, such as $a_0 = b_0 = c_0$ (given that $k_a = k_b = k_c$). The solution of (2.2) for an initial condition corresponding to any such fixed point will indicate that the system does not evolve at all, whereas, in reality, the system will certainly evolve for any initial condition such that $a_0 > 0$, and $b_0 > 0$ and $c_0 > 0$.

---

[3] Such solutions are calculated with the following rates values: $k_a = k_b = k_c = 1$

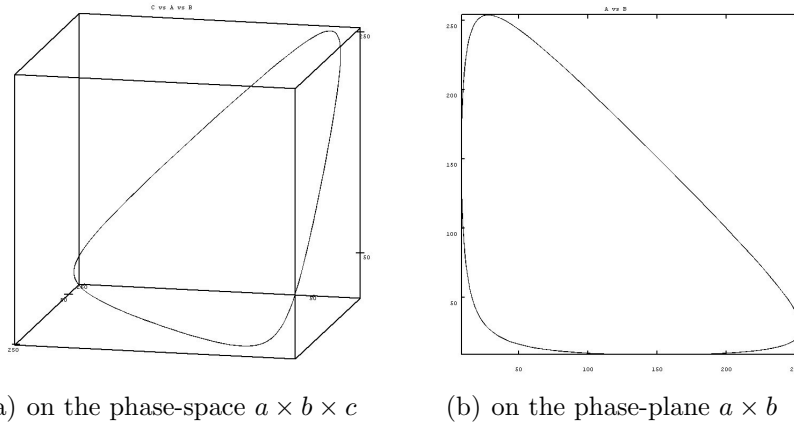(a) on the phase-space $a \times b \times c$ (b) on the phase-plane $a \times b$

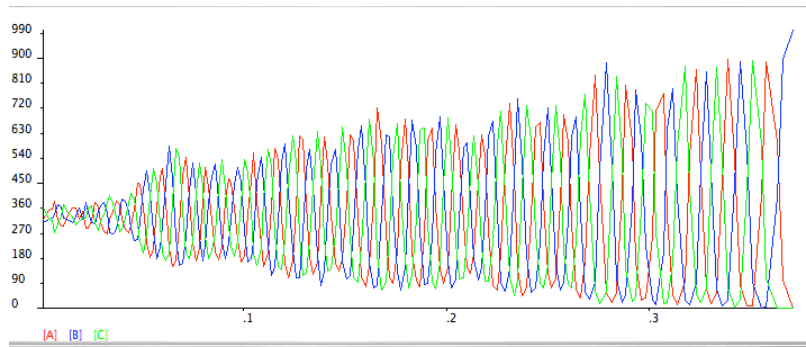Fig. 1. ODEs solutions of the 3-way oscillator with $(a_0, b_0, c_0) = (100, 200, 10)$

Fig. 2. A simulation trace of the 3-way oscillator with initial configuration $a_0 = b_0 = c_0 = 333$

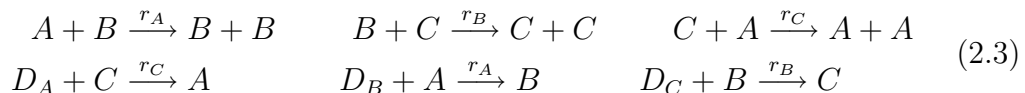### 2.3 Discrete Stochastic Modeling of Biological Systems

As the number of entities in biological systems is finite and discrete, it is possible describing biological systems as a collection of discrete state variables, whose value over time change as a consequence of interactions that stochastically happen in the system. It has been proved by Gillespie in [10] that, under homogeneity conditions, the time to the occurrence of the next reaction in a biochemical system can be very well approximated through a negative exponential probability distribution. This result provides the theoretical ground for representing and analyzing the dynamics of biochemical systems through continuous-time Markov chain (CTMC)[19] models. Gillespie's result has lead to the implementation of many software tools for the stochastic simulation of biological systems (see for example [17,5,7]). Such tools take as input a description of modeled system and through application of the Gillespie algorithm they calculate an approximated time course of the system's reactants. Figure 2 shows a plot of the output of a single simulation trace for the 3-way oscillator model (with initial condition $a_0 = b_0 = c_0 = 333$) obtained through

the Cyto-Sim [17] simulation tool. By observing the simulation trace in Figure 2 we can evince the following: first, the molecules of the 3 species A, B and C do follow an oscillating course; secondly, the amplitude of oscillation is not constant; thirdly, species B is the *oscillation damper* (i.e. oscillation ends with 999 molecules of B and none of both A and C). Such observations, however, are true only for that (very) specific simulation run. Other simulation runs may (and will) provide us with different insights. Therefore, if we want to use simulation for devising meaningful information about the general behaviour of a system we need to calculate statistics over a large number of simulation runs.

Alternative approaches to the simulation of discrete stochastic models, are given by Markovian analysis techniques [19] and/or probabilistic/stochastic model checking [11],[1]. In the following section we introduce the basics about temporal logic and model-checking and we will show how relevant properties of systems of biochemical reactions can be stated as (probabilistic) temporal logic queries which can be automatically verified against a discrete stochastic model. First, though, we introduce a *permanent oscillation* variant of our running example.

2.4 The 3-way oscillator with doping: permanent oscillation

In [3] Cardelli shows that in order to make the oscillation of the simple 3-way oscillator, perpetual, it suffices to add a *doping* reaction for each of the 3 species A, B and C, where a doping reaction is one that is able to produce a molecule of a species $X \in \{A, B, C\}$ from a molecule of a doping substance D_X and one of another species. Practically speaking the chemical equations form of the 3-way oscillator with doping is:

$$
\begin{array}{lll}
A + B \xrightarrow{r_A} B + B & B + C \xrightarrow{r_B} C + C & C + A \xrightarrow{r_C} A + A \\
D_A + C \xrightarrow{r_C} A & D_B + A \xrightarrow{r_A} B & D_C + B \xrightarrow{r_B} C
\end{array}
\tag{2.3}
$$

where D_A, D_B and D_C are the doping substances for A, B and C respectively. The role of the doping reactions (the bottom three reactions in (2.3)) is to avoid that the system blocks once a species get extinct: the presence of at least one molecule of each doping species is enough to guarantee that a species will not get extinct (a species may have zero molecules at some point in time, but it is guaranteed that from there new molecules of the species will be generated through doping). As a result it can be shown that oscillation, for the "doped" version of the 3-way oscillator, is never ending. We will use this model to illustrate, in a comparative fashion, how model checking techniques can be used for the analysis of both non-perpetual and perpetual oscillating behaviours.

3 On the logical characterisation of oscillation

We consider temporal logic to express *oscillation* related queries. Given a model's variable v we describe as *oscillation* (with respect to v) those evolutions for which v fluctuates around a value $v = k$. In logical terms (permanent) oscillation of a system may be expressed through the following temporal property: *"always in the future, the variable v departs from and reaches the value $v = k$ infinitely often"*. We consider branching-time temporal logic CTL [8] and its probabilistic extensions PCTL [11] and CSL [1] as formal means to expressing oscillation related properties. In the following we briefly introduce CTL and PCTL/CSL: for a detailed treatment the reader is referred to the literature [8,11,1].

In CTL, formulae are built upon a set of atomic propositions (AP) and are evaluated against *Kripke structure models*. Two basic temporal operators are used in combination with propositional logic's *conjunction* ($\wedge$), *disjuntion* ($\vee$) and *negation* ($\neg$): a *next-state* operator (X), for referring to properties of successor states (i.e. X (a) is true in s, denoted $s \models X$ (a), if a is true in some successor of s), and an *until* operator (U) for referring to properties which hold in future states (i.e. ($\psi \, U \phi$) is true in s if and only if from s it is possible to reach a future state $s' \models \phi$ through a sequence of states satisfying ψ). *Eventually in the future* (F (a) $\equiv$ ($true \, U \, a$)) and *always in the future* ($G \equiv \neg F(\neg a)$) are special cases of until formulae. CTL temporal operators are path quantified, either *existentially* (EX, EU) or *universally* (AX, AU). Existentially quantified formulae must be satisfied by at least a path starting from the current state, as opposed to universally quantified formulae, which must be satisfied by every path starting from the current state. Permanent oscillation (with respect to a variable v) can be characterised in CTL by the following formula:

$$\phi_{CTL} \equiv AG(((v\!=\!k) \rightarrow EF(v\!\neq\!k)) \wedge ((v\!\neq\!k) \rightarrow EF(v\!=\!k))) \qquad (3.1)$$

which states that, at any point in the future, if the system's evolution reaches a state such that $v\!=\!k$ then it has to be possible to reach a future state such that $v\!\neq\!k$ and vice versa.

The probabilistic CTL [11] (PCTL) and the Continuous Stochastic Logic [1] (CSL) are logic to state properties of, respectively, *discrete time* and *continuous time Markov chains* models. PCTL and CSL formulae are of the same form as in CTL except that path operators are continuously quantified, rather than existentially or universally. Hence, for example, the formula $P_{\trianglelefteq p}(\psi \, U \, \phi)$ is satisfied in a state s of a Markov chain model if the measure of probability of those paths satisfying ($\psi \, U \, \phi$) is bounded by p (with $\trianglelefteq \in \{\leq, <, >, \geq\}$ and $p \in [0, 1]$). Relying on basics propositional equivalence the CTL expression of oscillation (3.1) can be reformulated in PCTL/CSL terms resulting in the
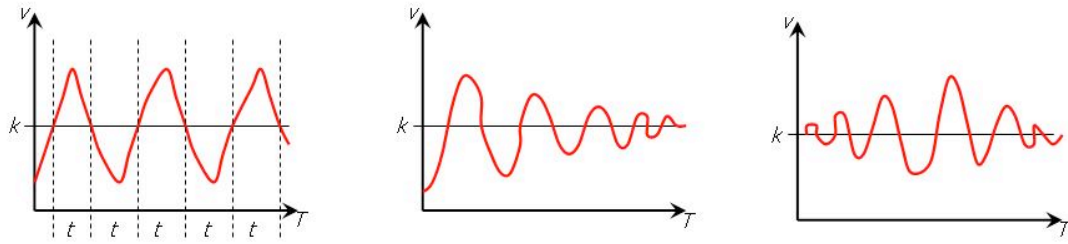
(a) permanent, regular oscillation (b) irregular transient oscillation (c) irregular transient oscillation

Fig. 3. Examples of *regular* vs *irregular* oscillation

following probabilistic formula:

$$\phi_{PCTL} \equiv P_{\leq 0}[\top\ U\ ((((v{=}k)\wedge P_{\leq 0}(\top\ U\ (v{\neq}k))))\vee(((v{\neq}k)\wedge P_{\leq 0}(\top\ U\ (v{=}k)))))] \tag{3.2}$$

which states that *"evolutions for which once a state $v = k$ is reached then it cannot be left"* have a null probability. Other important aspects of oscillation can be coded in temporal logic, *noise filtering* and *periodicity* or, more generally, *regularity* being two such examples. Fluctuations whose amplitude is below a certain (well defined) threshold classify as *noise*, and should not be detected as oscillations. If n is the noise threshold for a given system then existence of *noise filtered oscillations permanence* can be encoded in CTL as [4]:

$$\phi_{CTL} \equiv AG(((v{=}k) \rightarrow EF((v{>}k{+}n)\vee(v{<}k{-}n))\wedge((v{>}k{+}n)\vee(v{<}k{-}n) \rightarrow EF(v{=}k))) \tag{3.3}$$

Oscillation regularity aspects, on the other hand, regard the periodicity as well as the amplitude of fluctuations around k. Plots in Figure 3 are examples of different cases of oscillation. Figure 3(a) shows the most regular type of oscillation, one whose amplitude and periodicity are both always constant. Two examples of irregular oscillations are depicted in Figure 3(b) and Figure 3(c). Encoding of *oscillation regularity* aspects in temporal logic terms, is not a trivial task.

4 Probabilistic Model Checking

Probabilistic model checking is a formal technique for the verification of systems that can be modelled in terms of stochastic processes. It extends classical model checking [6], to the probabilistic framework. The basic idea behind (probabilistic) model checking is as follows: an algorithm is developed

---

[4] PCTL version of (3.3) is straightforward

which takes as input a (stochastic) model M of the considered system (often expressed through some formal modelling language, such as, for example, stochastic process algebras or stochastic Petri Nets [16]) and a property ϕ expressed as a (probabilistic) temporal logic formula. The algorithm then automatically check ϕ against M, and provides a *positive* output, if M satisfies ϕ (denoted $M \models \phi$) or a negative one if that is not the case (denoted $M \not\models \phi$). The verification of ϕ against M is achieved through an exhaustive exploration of M, hence the output of model checking is exact, as opposed to the output of (stochastic) simulation which is inherently approximated. The main difference between classical and probabilistic model checking is in the nature of the model they are referred to, and, as a consequence, in the type of verification they allow for. In classical model checking, a model is given by a *labelled transition system*, a graph which captures the system's evolutions without expressing any timing information. With probabilistic model checking models are Markov processes (i.e. labelled transition systems enriched with stochastic information), thus inherently contain timing and probabilistic information. As a consequence classical model checking is suitable for *qualitative verification* of a system's model, such as, for example *"does the oscillation of a system eventually stop?"*. On the other hand probabilistic model checking allows for *quantitative verification* of a model: properties such as *"what is the probability that the oscillation will stop within time T?"* can be stated and verified through probabilistic model checking. Finally the expressiveness of probabilistic model checking analysis is further enhanced by the possibility of *reward based verification*. A Markov chain model may be enriched with meaningful *state rewards* and/or *transition rewards*, and reward-based properties, such as, for example, *"what is the expected amplitude of oscillation within time T?"*, may then be verified against it. Before proceeding with the probabilistic model checking verification of the 3-way oscillator model (Sec. 6), we first set the basics about reasoning on oscillations with temporal logic.

5 Markov Chain model of the 3-way oscillator

Based on the chemical equations (2.1) a CTMC model of the 3-way oscillator can be obtained straightforwardly. Figure 4 illustrates the state-space of the CTMC model of the 3-way oscillator corresponding to the initial conditions: $a_0 = b_0 = c_0 = 3$. Each state in Figure 4 is given as a triple of integers (a, b, c) representing the corresponding number of molecules of each species in the state, whereas the labels on each arc represent the transitions rate [5]. We observe that the CTMC in Figure 4 enjoys a specific structure: each path along the perimeter of a given triangle corresponds to oscillations of constant

---

[5] Note that each transition rate is *marking dependent*, i.e. its value depends on the distribution of molecules on that state.
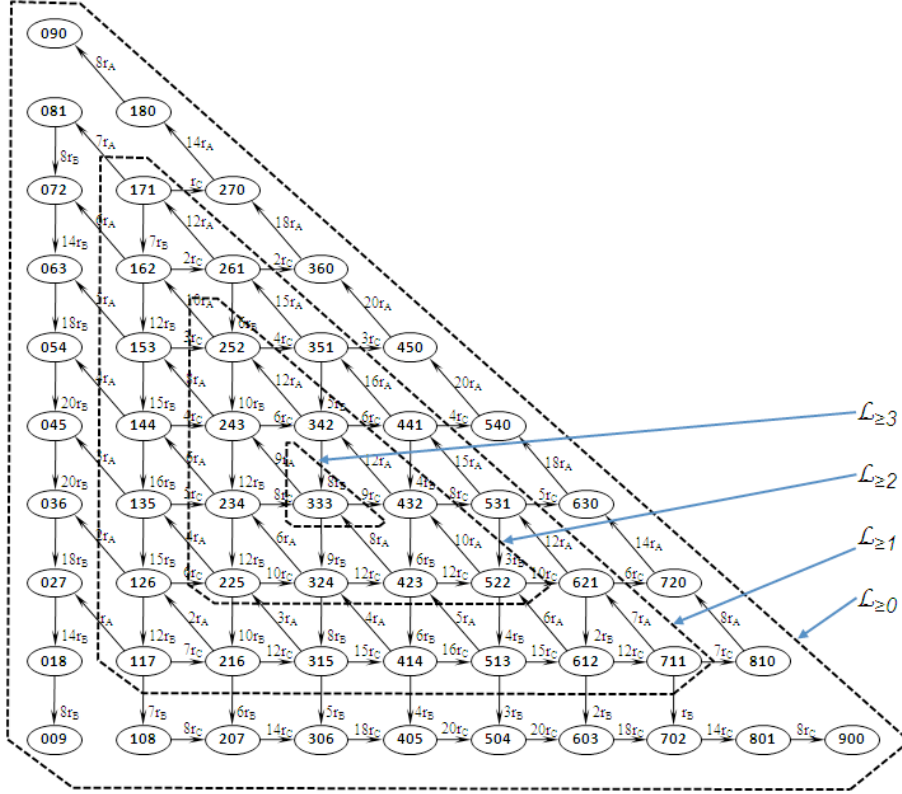
Fig. 4. State space of the 3 species oscillator with initial state $a_0 = b_0 = c_0 = 3$

amplitude [6] . We refer to the set of states on the perimeter of such a "triangle" as *level k states* , formally: $L_k = \{(a, b, c) \in S : min(a, b, c) = k\} \subset S$ and to the set of states contained in the level k triangle as $L_{\geq k} = \{(a, b, c) \in S : min(a, b, c) \geq k\} \subseteq S$. Counting the number of states in $L_{\geq k}$ is necessary in order to perform significant reward based analysis of the 3-way oscillator CTMC model (see Section 6). The total number of states in the CTMC is proportional to the system invariant ($INV = a_0 + b_0 + c_0$) and is given in the the following proposition. Furthermore an iterative rule for computing the number of states in $L_{\geq k}$ is also defined.

Proposition 5.1 *Let $M = (S, Q, s_0)$ be the CTMC model of the three species oscillator with state-space S and initial state $s_0 = (a_0, b_0, c_0) = (n, n, n)$ and $INV = a_0 + b_0 + c_0$.*

- *the number of states in S is given by:*

$$|S| = \frac{(INV + 2) \cdot (INV + 1)}{2} \tag{5.1}$$

---

[6] Note that states in each such "*k*-triangular" path are such that the minimum between A, B and C is equal to k.

INV	$|S|$	$k \leq n$	$|L_{\geq k}|$
10	66	10	1
20	231	9	10
50	1326	8	28
100	5151	7	55
200	20301	6	91
500	125751	5	136
1000	501501	4	190
2000	2003001	3	253
5000	12507501	2	325
10000	50015001	1	406

(a) total number of states   (b) number of k-level states for $1 \leq k \leq 10$

Fig. 5. Total number of states and k-level states in the 3-way oscillator CTMC

.

- *the number of states contained in the level-k triangle can be iteratively calculated as follows:*

$$|L_{\geq(k-1)}| = |L_{\geq(k)}| + 9 \cdot k \qquad \forall k : 1 \leq k \leq n \qquad (5.2)$$

  and the number of state on level-k triangle's perimeter is given by $|L_k| = 9(n - k)$. Note that, trivially, $|L_{\geq n}| = 1$.

Through (5.1) we can straightforwardly assess the state-space explosion as a function of the model initial state (i.e. as a function of INV). On the other hand (5.2) will be used to normalise the reward models described in the next section (Section 6) and which allow us to quantify the likelihood of the system to oscillate within *level k* as a function of the number of states in the model. Tables in Figure 5(a) and Figure 5(b) shows sample values obtained by application of (5.2) and (5.1).

6 Property specification and verification with PRISM

In this section we present results of probabilistic model checking verification of the PRISM model of the 3-way oscillator both in its standard and "doped" version. For the sake of space we omit the PRISM code, the interested reader can find it in [2].

6.1 Verification of the 3-way oscillator, with and without doping

The following queries, first described informally in natural language, then encoded as PRISM formulae, have been considered for verification.
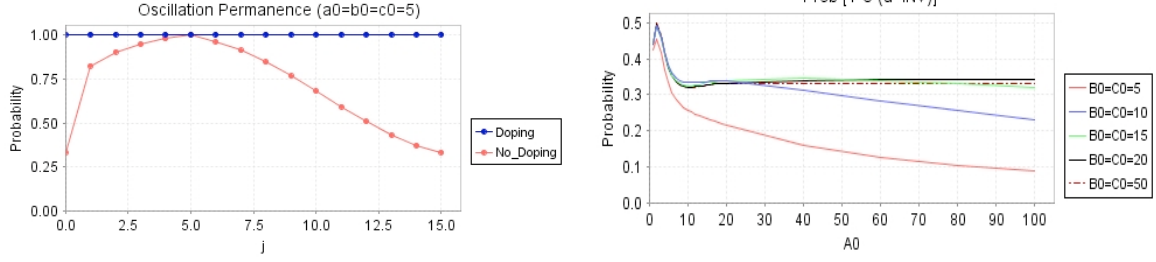
(a) The 3-way oscillator without doping has a non-null probability of never returning into state $a = j$

(b) Probability of oscillation to terminate in A

Fig. 6. Probability of permanent oscillation and probability of oscillations to terminate in a given state
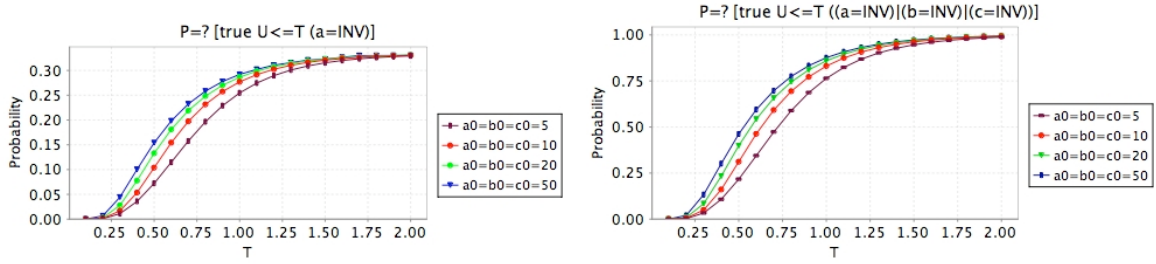
(a) Probability of oscillation to terminate in species A within time T

(b) Probability of oscillation to terminate within time T

Fig. 7. Probability of oscillation to terminate within a time bound T

**A.** *What is the probability that the oscillation will never terminate ?*

$$P_{=?}[\top \ U((((a \neq j)|P_{\geq 1}(\top \ U(a \neq j)))\&((a=j)|P_{\geq 1}(\top \ U(a=j)))))] \qquad (6.1)$$

Results of verification of (6.1) (referred to the initial state $a_0 = b_0 = c_0 = 5$), are shown in Figure 6(a). They confirm that oscillation is permanent for the 3-way oscillator with doping whereas it is not for the standard version of it: the probability of ever returning into state $a = j$ is less than 1 $\forall j \in [0, INV]$ (note that for $a = 5$ such probability is equal to 1 simply because of the initial state $a_0 = 5$). It should be noted that (6.1) provides a formal means to automated verification of oscillation termination, something which cannot be achieved through simulation.

B. *What is the probability that oscillation terminates in species $X \in \{A, B, C\}$?*

$$P_{=?}[\top \ U(a = INV)] \qquad (6.2)$$

Results for (6.2) are depicted in Figure 6(b). Specifically the probability that
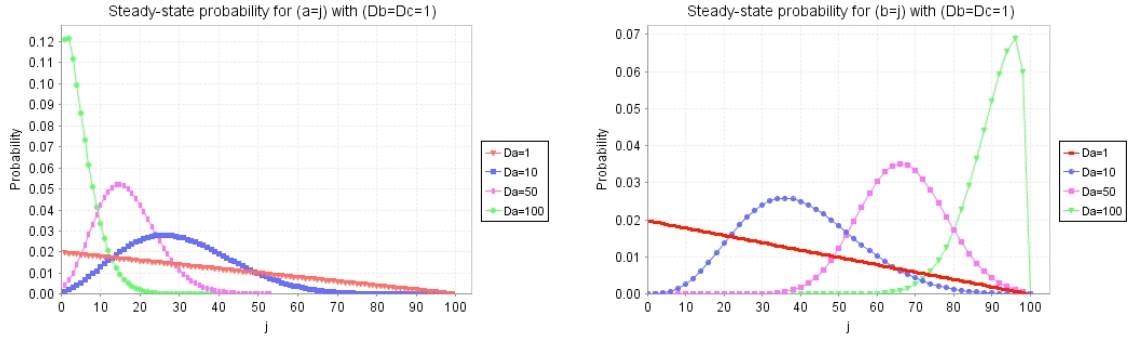
(a) Steady-state probability of A as a function of doping D_A

(b) Steady-state probability of B as a function of doping D_A

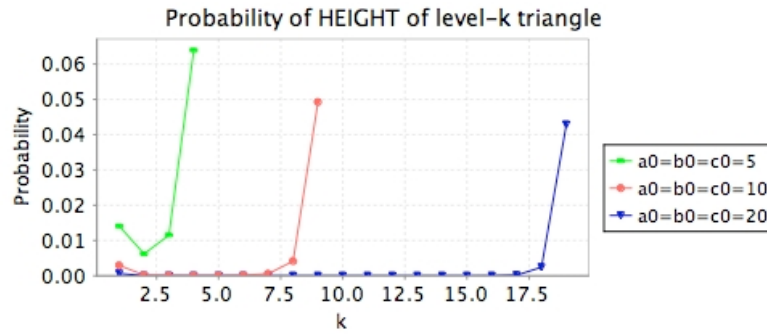Fig. 8. Oscillator with doping: steady-state probability of A and B as a function of doping D_A

Fig. 9. Probability of oscillation of amplitude k

oscillation ends in A is plotted against the initial number of A molecules (a_0), and for different (equal) initial number of B and C molecules $(b_0 = c_0)$. Plots in Figure 6(b) reveal a maximum corresponding to $a_0 = 2$, which indicates a particular probabilistic tendency of the 3-way oscillator, that is: the smaller the initial value of species X, the higher the probability that oscillation terminates in X.

C.1 *What is the probability that oscillation terminates in species X within time T?*

$$P_{=?}[\top \ U^{\leq T}(a = INV)] \tag{6.3}$$

C.2 What is the probability that the oscillation eventually terminates within time T?

$$P =?[\top \ U^{\leq T}((a = INV) \vee (b = INV) \vee (c = INV))]) \tag{6.4}$$

Results of verification of (6.3) and (6.4) are shown in Figure 7(a) and Figure 7(b) respectively. They show a rather counter intuitive peculiarity of the oscillator which is: the higher is the initial population, the faster the oscilla-
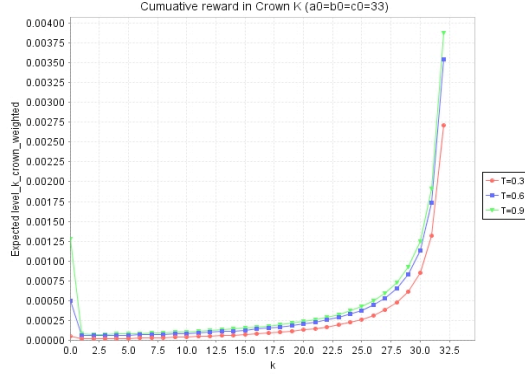
Fig. 10. Expected *level-k* within T with reward $\frac{1}{|L_k|}$ given to each *level-k* state and initial state $a_0 = 33, b_0 = 33, c_0 = 33$

tor will reach its absorbing state. Note that because these experiments are referred to uniform initial population (i.e. $a_0 = b_0 = c_0$), then the limit probability in Figure 7(a) is 1/3. Finally, plots in Figure 7(a) and Figure 7(b) provide a quantification of the time to absorbtion: oscillation stops certainly within time $T = 2$.

D What is the probability of an oscillation of amplitude k?

$$P = ?[((a = k) \wedge (b > k))U((a = k) \wedge (b = k) \wedge (c = (3 \cdot a_0 - 2 \cdot k)))] \quad (6.5)$$

An oscillation of amplitude k corresponds to the path on the perimeter of the level-k triangle (see Figure 4). Each side of the level-k triangle can be characterised by a CSL formula. For example (6.5) identifies the "height" of the triangle. Through verification of (6.5) (and similar formulae for the remaining two sides of the triangle), we can get a quantification of the likelihood of an oscillation of amplitude k. Figure 9 shows results of verification of (6.5) for different initial amount of molecules. The U-shaped plots in Figure 9 indicate that (very) small and (very) large amplitude oscillations are more likely than those whose amplitude is in between.

E (reward based). *What is the expected level-k reward cumulated within time T?*

$$R\{level\_k\_reward\} = ?[C^{\leq T}] \quad (6.6)$$

where *"level_k_reward"* is a reward model, that assigns a reward of $\frac{1}{|L_k|}$ to each level-k state, where the normalising constant $|L_k|$ is defined as in Proposition 5.1. Verification of (6.6) leads to plots in Figure 10. Those curves indicate that, on average, the system spend more time in the orbits closer to the initial state $a_0 = b_0 = c_0$.

F (stready-state). *What is the probability that the number of molecules of species X is j on the long-run?*

$$S_{=}=?[(a = j)] \tag{6.7}$$

such formula is referred to "doped" version of the 3-way oscillator only (i.e. the standard 3-way oscillator does not allow for a steady-state distribution). Results for (6.7) are depicted in Figure 8(a). Specifically the steady-state probability for $(a = j)$ is plotted against the amount of A's doping (D_A) for different (equal) initial number of B's and C's doping (i.e. $D_B = D_C = 1$). When the amount of doping is equal for all three species the probability at steady state decreases linearly when we move from outer orbits towards inner ones. On the other hand, as we increase the amount of doping of a species (A in this example) while keeping the doping for the other species constant, then the more "doped" species becomes less likely to "dominate" the others. Such rather counter intuitive behaviour may be explained by looking at the nature of reactions (2.3): in fact the more doping we add, the more likely we are to synthesis new molecules of A, hence the more likely we are to consume them in the production of new B. As a result, opposite to what one would expect, augmenting the doping of species X is actually going to augment (the average number of molecules of) species X', where X' is the successor of X, assuming $A < B < C$ as order.

7   Conclusion

In this paper we have proposed a comparative study of modelling techniques for the analysis of biochemical systems. We have focused on a specific class of systems, ones whose behaviour is characterised by existence of oscillations in the quantities of reactants they consist of. First we have demonstrated how useful insight about an oscillating system can be gained by application of standard system biology's modelling techniques, such as solution of continuos deterministic (ODEs) model and simulation of a discrete stochastic model. However we have seen that both such methods lack to fulfil the modeller needs for analysis of the considered system. We have then argued that formal languages, such as temporal logic, may be applied to encoding relevant characteristic of an oscillator, providing a formalisation of the *oscillation permanence* in terms of two popular logic: CTL and PCTL. We have then developed a Markovian model of both a *transient* oscillator, known as the 3-way oscillator as well as of its *permanent oscillation* variant. We have demonstrated the effectiveness of automated temporal logic verification, by coding such models in the probabilistic model checker PRISM. we have been able to perform a thorough probabilistic analysis of several reachability as well as reward based logical formulae, which provided us with interesting outcomes

which could not be obtained otherwise. Future developments of this work include the extension of the proposed methodology to the verification of more complex biological systems, such as the cell-cycle, a model of which is currently being developed and analysed.

7.1 Related work

Application of model checking techniques to system biology has been proposed in some other works. In [13], Kwiatkowska *et al.* developed a non trivial stochastic model of the complex Fibroblast Growth Factor (FGF) signalling pathway. They used PRISM to code the FGF pathway and developed a number of probabilistic properties, some of which reward-based, to verify relevant properties of the signalling mechanism. In [9] Gilber *et al.* propose a general overview about formal modelling techniques suited to the verification of biological systems. Based on Petri Net model of the ERK signal transduction pathway, they identify what type of formalisms/verification-technique, is suitable to what type of analysis the modeller is interested in. Hence in the discrete-state modelling framework non-probabilistic model checking (i.e. standard temporal logic) has to be used to perform *qualitative analysis*, whereas probabilistic/stochastic is needed for *quantitative analysis*. Finally in the continuous-state modelling framework, ODEs solution as well as Linear Time Logic with constraints (LTLc) are to be used for model analysis.

References

[1] C. Baier, B. Haverkort, H. Hermann, and J.-P. Katoen. Model-checking algorithms for continuous-time Markov chains. *IEEE Trans. on Software Eng.*, Vol. 29(6):pp. 524–541, June 2003.

[2] P. Ballarini, R. Mardare, and I. Mura. Query-based verification of biochemical oscillations through probabilistic model checking, TR 05/2008, available at http://www.cosbi.eu. Technical report, CoSBi, 2008.

[3] L. Cardelli. Artificial biochemistry. Technical report, The Microsoft Research - CoSBi, 2006.

[4] M. Cavaliere and S. Sedwards. Modelling cellular processes using membrane systems with peripheral and integral proteins. In Springer, editor, *Proceedings of the fourth International Conference on Computational Methods in Systems Biology (CMSB2006), LNBI-LNCS 4210*, 2006.

[5] F. Ciocchetta and J. Hillston. Bio-PEPA: an extension of the process algebra PEPA for biochemical networks. In *Proceedings of FBTC 2007*, volume 194 of *Electronic Notes in Theoretical Computer Science*, pages 103–117, 2008.

[6] E. Clarke, O. Grumberg, and D. Peled. *Model Checking.* MIT Press, 1999.

[7] L. Dematté, C. Priami, and A. Romanel. Modelling and simulation of biological processes in BlenX. *SIGMETRICS Performance Evaluation Review*, 35(4):32–39, 2008.

[8] E. Clarke, E. Emerson, and A. Sistla. Automatic verification of finite-state concurrent systems using temporal logic specifications. *ACM Transactions on Programming Languages and Systems*, Vol. 8(2):pp. 244–263, 1986.

[9] D. Gilbert, M. Heiner, and S. Lehrack. A unifying framework for modelling and analysing biochemical pathways using petri nets. In *In proceedings CMSB 2007, (Computational Methods in Systems Biology), Editors: M.Calder and S.Gilmore, Published by Springer-Verlag in LNCS/LNBI*, 2007.

[10] G. T. Gillespie. Exact stochastic simulation of coupled chemical reactions. *J Phys Chem.*, 81(25):2340–2361, 1977.

[11] H. A. Hansson and B. Jonsson. A framework for reasoning about time and reliability. In *Proc. 10th IEEE Real-Time Systems Symposium*, pages 102–111, Santa Monica, Ca., 1989. IEEE Computer Society Press.

[12] S. Hoops, S. Sahle, R. Gauges, C. Lee, J. Pahle, N. Simus, M. Singhal, L. Xu, P. Mendes, and U. Kummer. COPASI — a COmplex PAthway SImulator. *Bioinformatics*, 22:3067–3074, 2006.

[13] John Heath, Marta Kwiatkowska, Gethin Norman, Dave Parker, and Oksana Tymchyshyn. Probabilistic model checking of complex biological pathways. *Theoretical Computer Science*, 391:239–257, 2008.

[14] H. Kitano, editor. *Foundations of Systems Biology.* MIT Press, 2001.

[15] M. Kwiatkowska, G. Norman, and D. Parker. Probabilistic symbolic model checking with probabilistic symbolic model checking with PRISM. In J. Katoen and P. Stevens, editors, *Proceedings of the 8th International Conference on Tools and Algorithms for Construction and Analysis of Systems (TACAS 2002)*, volume 2280 of *Lecture Notes in Computer Science*, 2002.

[16] M. A. Marsan, G. Balbo, G. Conte, S. Donatelli, and G. Franceschinis. *Modelling with Generalized Stochastic Petri Nets.* Wiley, 1995.

[17] S. Sedwards. and T. Mazza. Cyto-sim: A formal language model and stochastic simulator of membrane-enclosed biochemical processes. *Bioinformatics*, 2007.

[18] J. C. Sible and J. J. Tyson. Mathematical modeling as a tool for for investigating cell cycle control networks. *Methods*, 41:238–247, 2007.

[19] W. J. Stewart. *Introduction to numerical solution of Markov Chains.* Princeton, 1994.

[20] V. Volterra. Variations and fluctuations of the number of individuals in animal species living together. *Animal Ecology*, 1931.