From Simple Regulatory Motifs to Parallel Model Checking of Complex Transcriptional Networks

J. Barnat, L. Brim, I. Černá, S. Dražan, and D. Šafránek [1,2]

Faculty of Informatics, Masaryk University, Czech Republic

Abstract

The central mechanism which drives every living cell is protein synthesis, the so-called transcription, which is realized according to the genetic code. There are complex regulatory interactions that control transcription of genes to proteins. Owing to their inherent complexity, analysis of dynamical models of such interactions requires a scalable computational approach. In this paper we employ parallel LTL model checking for a case study of selected dynamic properties of an in silico model of transcription in Bacillus subtilis, a bacterium living in soil. Moreover, we show the general fact that crucial LTL properties characterising transcriptional dynamics can be inferred from network motifs commonly studied in systems biology.

Key words: transcriptional networks, model checking, B. subtilis

1 Introduction

The traditional reductionistic style of biological research is nowadays turning towards systematic integrative paradigm, the so-called *systems biology*. There appear plenty of databases of biological knowledge keeping both structural and functional aspects of living organisms. Existence of such databases allows building of *in silico models* that predict functionality of living cells.

Metabolism of any living cell is driven by proteins. Proteins are synthesised according to the genetic code in the process of *transcription*. Transcription of genes to proteins is controlled by complex regulatory interactions. These interactions are driven by specific proteins, so-called *transcription factors*, which collaborate on activation (increase) or repression (decrease) of particular gene transcription. For each cell type the transcriptional regulation is described by a *transcriptional regulatory network* (TRN) in which nodes represent proteins

[1] Email: `safranek@fi.muni.cz`

[2] This work has been supported by the FP6 project No. NEST-043235 (EC-MOAN) and by the Academy of Sciences of CR grant No. 1ET408050503.

This paper is electronically published in
Electronic Notes in Theoretical Computer Science
URL: `www.elsevier.nl/locate/entcs`

and genes. Protein nodes are interconnected with gene nodes by edges which denote transcriptional regulation. TRNs make the bottom level of complex biological networks and pathways. Even when taken separately, TRNs relevant to the most of pathways can be themselves very complex (containing 10s of genes). A general scheme of TRN is depicted in Figure 1. There is a particular network of 8 genes from which m proteins (each denoted X_i) are produced. Additionally, the process of regulation is controlled by n external signals which determine interface for interaction with other functional layers in a cell.

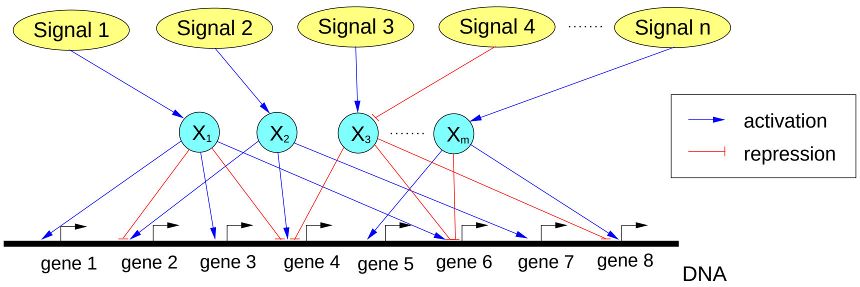

Fig. 1. Scheme of a transcriptional regulatory network

The most widely-used modelling frameworks for the analysis of the dynamics of TRNs, i.e., the prediction of how concentration of species evolve in time, are based on deterministic approach of ordinary differential equations [17] (ODE). The reduction of continuous models to discrete automata by a sequence of approximations and abstractions allows formal methods for the automated verification of properties of discrete transition systems to be applied [6]. One method which can be employed here is model checking. The most interesting properties of TRN dynamics can be characterised by certain parts of the interaction networks – so-called *network motifs*. Role of network motifs in dynamic behaviour can be mathematically analysed [15]. However, such mathematical analysis is very complicated and cannot be performed automatically. Therefore, model checking appears to be a suitable tool for algorithmic analysis of reduced ODE models. Properties which can be analysed by model checking include the properties introduced by particular network motifs. Recent studies on biologically-relevant properties identified the need for both branching-time temporal operators, able to express multi-stability properties (reachability of several different equilibrium states) [12,7], and linear-time temporal operators, able to capture oscillations of protein concentrations [2,7] as well as temporal ordering of events.

When dealing with complex TRNs, standard model-checking techniques do not provide acceptable response times for answering user queries and scalable parallel model-checking algorithms are thus needed. A typical well-understood TRN analysed by biologists consists of around 10 state variables. Owing to deep dynamical dependences among state variables, the state-space explosion arises. In consequence, discrete reductions of such relatively small models expand to hundreds of millions of states. It is also worth noting that simulation algorithms which compute the dynamics of TRNs by discrete overapproximation of ODE solutions, which is a very accurate method, have also high time complexity (i.e., in the method that we consider here the time needed to compute successor states for a given state grows exponentially with the number of variables).

In order to deal with such complexity of transcriptional regulatory networks, powerful computational methods are inevitable. One of the most critical limitations in applying current approaches to modelling and functional analysis is their pure scalability. Large models require powerful computational methods, the hardware infrastructure is available (clusters, GRID, multi-core computers), but the parallel (distributed) algorithms for model analysis are still under development.

In general, while substantial work on model-checking qualitative as well as quantitative properties of deterministic models of TRNs has been already achieved, to our best knowledge, no attempts to use parallel model checkers are known.

1.1 Our Contribution

In this paper we present a case study of applying parallel model checking to analysis of transcriptional regulatory networks using the GeNeSim, an extension of the parallel model-checker DiVinE [3]. We employ the piece-wise affine discrete abstraction method in which differential equations of the original ODE model are reduced to a system of piece-wise linear differential equations (PLDE) as introduced in [11]. Our distributed state space generator is based on the respective qualitative simulation method implemented (as a sequential algorithm) in *Genetic Network Analyser (GNA)* [10].

The DiVinE distributed state space generator allows on-the-fly generation of the transition graph giving thus in many circumstances the possibility to analyse properties of even larger networks as opposed to the explicit representation as used in the GNA simulator. This allows to check biologically interesting liveness properties on larger models than is possible with traditional sequential approach. In this case study we demonstrate that parallel model checking can extend the possibilities of qualitative analysis of Bacillus subtilis bacteria presented in [9].

From our experiences on collaboration with biologists we have found it difficult to fully take advantage of the strong power of temporal logic. It is mainly because of the fact that the experimental biological research principally differs from the formal approach of computer science. At present, it is far from reality that a temporal logic formalism can be directly used by biologists. To this end, we try to reveal biologically interesting qualitative LTL properties from the notion of network motifs which is well known and understood in the biological community.

This work extends our previous work [2] in two directions. At first, we introduce a pool of biologically relevant LTL properties which are systematically derived from transcriptional motifs which frequently appear in biological databases [14,16]. At second, the experiments of the Bacillus subtilis case study presented in this paper are aimed at demonstrating how the motif-specific properties can be checked on complex networks that combine several motifs together.

1.2 Related Work

The use of model checking for the analysis of biological networks has attracted much attention [6,19,12]. The individual approaches differ in models and model-checking tools used. Our approach is based on qualitative hybrid models as proposed in [11] and implemented in GNA [10]. Another tool is the Robust Verification of Gene Networks (RoVerGeNe) [4]. This tool complements the GNA by focusing on synthesis of quantitative parameters from PLDE models w.r.t. temporal properties of the transcriptional dynamics.

Besides GNA there are some other sequential approaches for model-checking of biological pathways, the most of which can be adapted also for analysis of TRNs. The BIOCHAM workbench [7] provides an interface to symbolic model checker NuSMV and the enumerative CADP verification toolbox; the interface is based on a simple language for representing biochemical networks. The workbench provides mechanisms to reason about reachability, existence of partially described stable states, and some types of temporal behaviour. There is also a biological extension of the probabilistic model checker PRISM [13], which allows quantitative analysis of biological pathways. However, none of the tools mentioned above employ the parallel model checking approach.

2 Preliminaries

Before we start with stating the properties of individual transcriptional motifs, we briefly explain basic principles of PLDE systems and their qualitative simulation. The precise definition of these principles is given in [11]. Finally, we also give a syntax of the LTL logic as used in this paper.

2.1 PLDE Systems

Assume a network with n proteins. Let the concentration of ith protein in a time instant t, where $i \in \{1, ..., n\}$, be denoted by the variable $x_i(t)$. The respective PLDE system consists of a set of n equations. For each protein there is an equation describing how its concentration changes in time:

$$\frac{dx_i}{dt} = \sum_{l \in L} \kappa_{il} \varrho_{il}(\langle x_1, ..., x_n \rangle) - \gamma_i x_i$$

where

- L is a finite index set.
- For each $i \in \{1, .., n\}, l \in L$, κ_{il} is a constant expressing the rate of protein production.
- $\varrho_{il} : \mathcal{R}^n \to \{0, 1\}$ is a discrete input function that denotes the transcriptional regulation.
- γ_i is a constant expressing the rate of exponential decay of protein i.

In general, each equation of the system consists of two terms – the positive production term and the negative degradation term. The latter term is naturally nonzero and describes instability of proteins among other species in the cell. The former term describes the transcriptional regulation, in particular, the intensity of protein production w.r.t. the current conditions in the cell. More precisely, the maximal production rate is given by the production constant κ which is additionally regulated by the input function ϱ.

The input function ϱ in general depends on the current concentration of all proteins in the system. In piece-wise linear approximation the (multidimensional) input function has a discrete range and is given by a product of elementary (one dimensional) step functions. These step functions qualitatively characterise edges in TRNs, in particular, the transcriptional activation or repression of the target gene by a certain transcriptional factor w.r.t. its given threshold concentration. In a simplified case, the input function has the form:

$$\varrho_{il}(\langle x_1, ..., x_n \rangle) = \prod_{j=1}^{n} s^*(x_j, \theta_j^i)$$

where $s^* : \mathcal{R} \to \{0, 1\}$, $* \in \{+, -\}$, denotes a so-called step function defined for the protein concentration x_j and its threshold θ_j^i by the following expressions:

$$\text{activation}: \quad s^+(x_j, \theta_j^i) = \begin{cases} 1, \text{ if } x_j > \theta_j^i, \\ 0, \text{ if } x_j < \theta_j^i, \end{cases}$$

$$\text{repression}: \quad s^-(x_j, \theta_j^i) = 1 - s^+(x_j, \theta_j^i)$$

2.2 Qualitative Analysis of PLDE Systems

For given initial conditions (initial concentrations of protein species), the possible evolution of protein concentrations in time can be predicted by simulation. The algorithm of qualitative simulation for PLDE models relies on the fact that input functions given as products of step functions have discrete ranges. Up-to this approximation, the concentration values of proteins are discretely abstracted into several open intervals between respective thresholds, and additionally, the discrete points equal just to the threshold values.

For each variable there is a finite number of regulation configurations determined by particular values of the step functions. In each such configuration the transcriptional regulation tends towards an equilibrium state (a so-called *local equilibrium*) determined for some $L' \subseteq L$ by the equation:

$$\frac{dx_i}{dt} = 0 \Leftrightarrow x_i = \frac{\sum_{l \in L'} \kappa_{il}}{\gamma_i}$$

If we consider a variable x_i then for each particular configuration of the binary values returned by step functions appearing in the respective equation the equilibrium concentration value is given just by the expression above. In order to simulate the dynamics of the PLDE system, we set for each variable the inequality condition which mutually relates all the thresholds defined for that variable, and moreover, we also put each local equilibrium concentration value between some two succeeding thresholds.

Evolution of protein concentrations starting from the given initial values is then determined by the state transition system in which each state is characterised by the vector of current discrete concentration levels of the proteins (open intervals or just the thresholds). Because each protein is naturally degraded (γ_i is always nonzero), values of each variable are bounded by some maximal level and hence the system has finite number of states [11].

The outgoing transitions are computed differently for states where all variables are set to values approximated by an open interval (so-called *regulatory states*) and for states where at least one variable has a value equal to some threshold (so-called *switching states*). Each transition directs the qualitative solution to continue in some of the succeeding states in the global phase space. In the former case, the transitions are determined directly by a direction vector leading from the current state to its respective local equilibrium state. In the latter case, the direction vector is computed by an intricate algorithm which relies on analysis of the system dynamics in neighbouring states. In general, time complexity of computing outgoing transitions for a given state is exponential in number of variables in the system. This is an imposition due to the accuracy of the overapproximative simulation and also a motivation for employing the parallel model checking method for analysis of PLDE systems.

2.3 Qualitative Properties

We use LTL formulae for expressing temporal properties of qualitative state transition systems. The atomic propositions are declared in the following way.

Let $\mathcal{S}$ be a PLDE system consisting of n variables each of which is denoted x_i, $i \in \{1, ..., n\}$. Moreover, for each x_i let T_i denote the set of all thresholds appearing in the equation for $\frac{dx_i}{dt}$, $T_i = \{\theta_i^j | j \in I_i, I_i \subseteq \{1, ..., n\}\}$. The set of atomic propositions $AP(\mathcal{S})$ is then defined in the following way.

For each $\theta_i^j \in T_i$

$$p \in AP(\mathcal{S}) \overset{df}{\Leftrightarrow} p \equiv x_i = \theta_i^j \mid x_i < \theta_i^j \mid x_i > \theta_i^j \mid x_i \leq \theta_i^j \mid x_i \geq \theta_i^j.$$

An atomic proposition p is satisfied by a path starting at a state s just if the particular concentration level x_i in s satisfies the inequality stated by p. Atomic propositions can be composed by usual logic connectives. Temporal formulae are build from the atomic propositions in traditional way.

3 PLDE and LTL Approach to Transcriptional Motifs

In this section, we characterise the basic transcriptional motifs modelled as PLDE systems. For each motif we introduce a set of LTL properties which describe qualitatively its significant dynamical properties. We focus on motifs that embody qualitatively exceptional properties, in particular, we do not include motifs that manipulate timing aspects of transcriptional interactions (e.g., single input modules, feed-forward loops,... [18]).

3.1 Autoregulation Motifs

Autoregulation is the simplest motif which appears very frequently in TRNs [1]. It deals with one variable (one protein) which regulates the production of its own. Such a regulation determines the resulting amount of the protein concentration which is achieved when the system reaches a global equilibrium state. In particular, the regulation is controlled by a step function which depending on the current amount of the protein concentration returns its increment. Depending on the character of the step function there are two types of autoregulation – negative and positive.

3.1.1 Negative Autoregulation

In the case of negative autoregulation, the input function is given by a single step function denoting repression. In particular, this means that the protein represses production of its own. The PLDE system consists of a single equation having the following form:

$$\frac{dx}{dt} = \kappa s^{-}(x, \theta) - \gamma x$$

The only local equilibrium state of this system is characterised by the concentration value $\frac{\kappa}{\gamma}$. By positioning this value above or below the threshold, we get different dynamic behaviour. In Table 1 there is given a characteristic LTL property for each of these positions.

The first line shows the situation when the local equilibrium concentration is positioned below the threshold. It implies that the production rate is relatively low and hence the autoregulation does not affect the transcription at the concentration level below the threshold, whereas for the values above the threshold the autoregulation switches off the production. Both aspects lead to the observation that this autoregulation motif forces the resulting global equilibrium state to appear at levels below the threshold.

$0 < \frac{\kappa}{\gamma} < \theta$	$\mathbf{FG}(x \leq \theta)$
$\theta < \frac{\kappa}{\gamma} < max$	$\mathbf{FG}(x = \theta)$

Table 1
Characteristic Properties of Negative Autoregulation

The second line represents the situation when the local equilibrium concentration is set to a value greater than the threshold. In this case it implies that the production rate is relatively high and thus the autoregulation leads the system from all initial concentration levels to a global equilibrium, which is characterised by the concentration value equal to the threshold.

3.1.2 Positive Autoregulation

Positive autoregulation is determined by a single activation step function. Here the protein at certain concentration level activates its own production. The respective PLDE system is analogous to the previous one:

$$\frac{dx}{dt} = \kappa s^{+}(x, \theta) - \gamma x$$

Similarly to the case of negative autoregulation there are two different local equilibrium positions. They are listed in Table 2. In the first case, the behaviour is just the same as in the previous case. However, the case when the local equilibrium concentration is set above the threshold has different behaviour. In particular, a kind of bistability arises. In other words, there are two different global equilibrium states to which the system can lead. In such a setting, the positive autoregulation behaves like a switch that decides whether the system stabilises at the maximal concentration level or at the minimal level. Both situations are possible.

$0 < \frac{\kappa}{\gamma} < \theta$	$\mathbf{FG}(x \leq \theta)$
$\theta < \frac{\kappa}{\gamma} < max$	(1) $((x < \theta) \Rightarrow \mathbf{G}(x < \theta)) \wedge ((x > \theta) \Rightarrow \mathbf{G}(x > \theta))$ (2) $(x = \theta) \Rightarrow (\mathbf{GF}(x > \theta) \wedge \mathbf{GF}(x < \theta))$

Table 2
Properties of Positive Autoregulation

The LTL property (1) describes the existence of the two global equilibrium states — a necessary and sufficient condition of bistability. In PLDE approximation, both stable states are reachable from a single specific qualitative state – a so-called bistable switch. However, the existence of such a state cannot be directly expressed in LTL. Therefore we state at the least a formula (2) which contradicts the existence of a bistable switch.

3.2 Feedback Loop Motifs

Another significant category of transcriptional motifs is characterised by a group of at least two different transcriptional factors which mutually interact in a circular manner. In particular, such motifs have the form of cyclic graphs. Feedback loops, depending on the type of individual interactions in the circle, embody behaviour similar to autoregulation (imposing stability or multi-stability), or they can lead to concentration oscillations [20] (e.g., the well-known circadian rhythm). We consider properties of minimal loops consisting of two proteins.

3.2.1 Double-negative feedback loop

In this kind of a feedback loop each of the two proteins represses the other. If we denote the proteins X and Y, respectively, the relevant PLDE system is defined by the following equations:

$$\frac{dx}{dt} = \kappa_x s^-(y, \theta_y^x) - \gamma_x x \qquad \frac{dy}{dt} = \kappa_y s^-(x, \theta_x^y) - \gamma_y y$$

$\theta_x^y < \frac{\kappa_x}{\gamma_x} < max_x$ $\theta_y^x < \frac{\kappa_y}{\gamma_y} < max_y$	(1) $((x > \theta_x^y \wedge y < \theta_y^x) \Rightarrow \mathbf{G}(x > \theta_x^y \wedge y < \theta_y^x)) \wedge ((x < \theta_x^y \wedge y > \theta_y^x) \Rightarrow \mathbf{G}(x < \theta_x^y \wedge y > \theta_y^x))$ (2) $(x = \theta_x^y \wedge y = \theta_y^x) \Rightarrow ((\mathbf{GF}(x > \theta_x^y) \wedge \mathbf{GF}(x < \theta_x^y)) \vee (\mathbf{GF}(y > \theta_y^x) \wedge \mathbf{GF}(y < \theta_x^y)))$

Table 3
Properties of double-negative feedback loop

From all the possible different local equilibria configurations we consider the most representative one, given in Table 3. In such a setting, the system is bistable, and moreover, there is a bistable switch in the state characterised by the atomic property stating that concentration value of each protein is equal to the respective threshold.

The formula (1) states that two different global equilibrium states exist in the system. Similarly as in the case of bistability in positive autoregulation, the property of bistable switch is not expressible in LTL. However, we give the formula (2) which involves counterexamples of a bistable switch.

3.2.2 Double-positive feedback loop

This kind of two-protein feedback loop has the similar properties as the double-negative loop. The PLDE system differs only in that both step functions are positive in this case.

$\theta_x^y < \frac{\kappa_x}{\gamma_x} < max_x$ $\theta_y^x < \frac{\kappa_y}{\gamma_y} < max_y$	(1) $((x > \theta_x^y \wedge y > \theta_y^x) \Rightarrow \mathbf{G}(x > \theta_x^y \wedge y > \theta_y^x)) \wedge ((x < \theta_x^y \wedge y < \theta_y^x) \Rightarrow \mathbf{G}(x < \theta_x^y \wedge y < \theta_y^x))$ (2) $(x = \theta_x^y \wedge y = \theta_y^x) \Rightarrow ((\mathbf{GF}(x > \theta_x^y) \wedge \mathbf{GF}(x < \theta_x^y)) \vee (\mathbf{GF}(y > \theta_y^x) \wedge \mathbf{GF}(y < \theta_x^y)))$

Table 4
Properties of double-positive feedback loop

The most representative configuration and the respective bistability properties are stated in Table 4. The property (1) differs from the respective property of the negative loop only in atomic propositions. The bistable switch contradiction (2) remains the same.

3.2.3 Incoherent feedback loop

The most interesting two-protein feedback motif is based on circular interaction in which the first of the two proteins acts as a repressor of the second protein production while the second protein acts as an activator of the first protein (or vice-versa). Formally, the corresponding PLDE system for such a situation can be of the following form:

$$\frac{dx}{dt} = \kappa_x s^-(y, \theta_y^x) - \gamma_x x \qquad \frac{dy}{dt} = \kappa_y s^+(x, \theta_x^y) - \gamma_y y$$

The representative configuration of local equilibria is determined by the same inequalities as in the case of the previous loop motifs. In such a situation, the motif leads to oscillatory behaviour (when initial concentrations are suitably set). The oscillation means that each of both species periodically changes its concentration value from the level below the threshold to the level above the threshold and vice-versa. Moreover, the oscillatory behaviour is in this case kept on for the entire time-live of the cell. LTL properties which guarantee the oscillatory behaviour for each of the proteins are given in Table 5 as properties (1) and (2).

$\theta_x^y < \frac{\kappa_x}{\gamma_x} < max_x$	(1) $(\mathbf{G}((x_a \leq \theta_a^b) \Rightarrow \mathbf{F}(x_a > \theta_a^b))) \wedge (\mathbf{G}((x_a \geq \theta_a^b) \Rightarrow \mathbf{F}(x_a < \theta_a^b)))$
$\theta_y^x < \frac{\kappa_y}{\gamma_y} < max_y$	(2) $(\mathbf{G}((x_b \leq \theta_b^a) \Rightarrow \mathbf{F}(x_b > \theta_b^a))) \wedge (\mathbf{G}((x_b \geq \theta_b^a) \Rightarrow \mathbf{F}(x_b < \theta_b^a)))$

Table 5
Characteristic properties of incoherent feedback loop

4 GeNeSim: Parallel Model Checker for PLDE Models

GeNeSim is build on the top of the DiVinE library that offers common functions needed to develop a parallel or distributed enumerative model checker. The only extension to the library that was necessary, was the extension of the state generator to a state generator tailored for the specific input provided by GeNeSim GUI [2]. For the structure of GeNeSim implementation and connection to DiVinE see Figure 2.

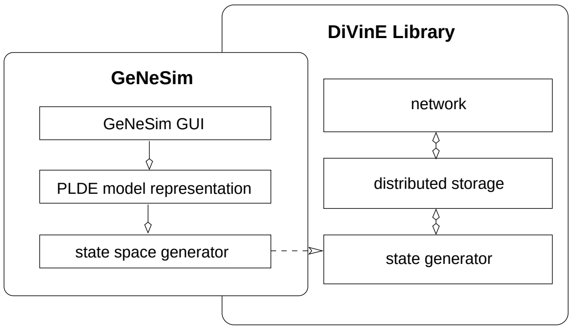

Fig. 2. How is GeNeSim embodied into DiVinE.

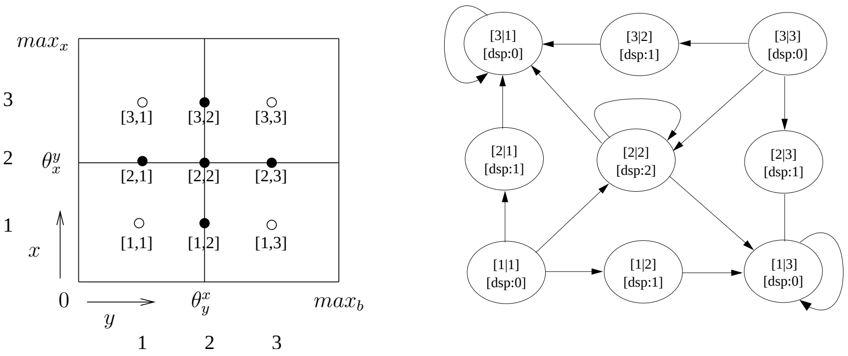

Fig. 3. A qualitative state space and a transition system generated by GeNeSim

4.1 PLDE model representation and state generator

The central component of GeNeSim is the state generator. The hierarchy of DiVinE state space representation classes is extended with new structures that represent symbolically the PLDE model in C++. A PLDE model is represented as a container of variables which occur in respective PLDEs. Each state variable contains a set of production rates (may be empty). Each production rate is defined as a container of *regulation terms*. Each regulation term is responsible for the activation or deactivation of the relevant production rate constant. We support two forms of regulation terms – negative and positive. The positive term is defined as a direct product of step functions whereas the negative term is defined as a negation of a product of step functions. Both kinds of regulation terms allow all possible types of PLDE models to be encoded in GeNeSim. Negative terms are necessary for modelling of transcription factors which have the form of multi-protein complexes, as showed in [11].

The GeNeSim state generator implements the DiVinE methods `get_initial()` for determining the initial state of the system and `get_succs()` that computes a container of successors of a given state. By these two methods the implicit representation of the qualitative state transition system is implemented. Such a representation allows integration of GeNeSim with DiVinE distributed on-the-fly LTL model checking algorithm.

An example of a simulation transition system generated by GeNeSim is depicted in Fig. 3 (on the right). It represents the state transition system of the PLDE model of the double-negative loop motif presented in the previous section. The information included in the individual states denotes the address of the state and the so-called *direction set property*. The address of each domain is given as a vector of discrete concentration levels of respective proteins in the symbolic concentration space. Open intervals are represented by odd numbers and threshold levels by even numbers (see Fig. 3 (on the left)).

Direction set property (dsp) expresses the information characterising the possible directions that the system can follow from a switching state. Because of the discrete range of step functions, PLDEs are not defined at threshold points. In consequence, the behaviour at switching states is predicted by overapproximation to all meaningful situations. In this point, three scenarios can arise. The first possibility is the existence of a global equilibrium at respective threshold concentration levels – this situation is denoted by the self edge. The second possibility is keeping concentration at respective threshold levels for a finite time period ($dsp = 2$). The last situation is immediate change in concentration levels to values above or below thresholds ($dsp = 1$). The dsp has no sense for regulatory states, because in their case the concentration is represented by open intervals between thresholds and thus the last situation cannot occur.

5 Experiments

In Section 4 we have stated several qualitative properties of dynamics of simple transcriptional motifs. It is experimentally proved [18] that complex TRNs of real cells are composed from such elementary motifs. In this section we show how model checking can be employed to investigate the presence of the key motif properties in complex networks. In particular, we present a case study on analysis of a concrete transcriptional network responsible for sporulation of a soil bacteria Bacillus subtilis. The considered network, presented in [9], consists of 11 variables and its transition system is measured in hundreds of millions of states.

By employing parallel model checking, we demonstrate how the previous detailed dynamical analysis of this network [9] can be extended to automatized decision of presence of selected motif properties. In particular, we focused on the following properties:

(a) stability of low concentration level of AbrB protein caused by negative autoregulation through σ_a promoters
(b) stability of threshold concentration of AbrB protein in the same conditions as in (a)
(c) bistability of SinI protein under and above the lowest threshold
(d) nonexistence of a bistable switch in SinI protein behaviour
(e) oscillation of SinI protein concentration around the lowest threshold
(f) oscillation of SinI protein concentration around the highest threshold

We have considered a subset of the entire transition system reachable from the inital state characterised by low concentration levels of all the proteins. We have found out that the properties (a), (b), and (f) were satisfied, whereas all the others were not. As the PLDE model overapproximates the original continuous model, we can say directly that also solutions of the precise mathematical model satisfy the properties (a), (b), and (f). Moreover, the oscillation of the SinI protein goes with the biological study presented in [20]. However, because of the overapproximation, we cannot directly interpret the negative

Model	States	Trans.	1 node	5 nodes	10 nodes	15 nodes	20 nodes
(a)	1176457	4140397	36882.2 s	14340.8 s	9894.57 s	8537.45 s	7839.13 s
(b)	1369357	5214972	42293.6 s	18779.1 s	13484.6 s	11386.9 s	10477.7 s
(c)	927602	2521760	10966.7 s	7202.44 s	4806.69 s	4282.16 s	3870.28 s
(d)	1612281	3651085	29368.2 s	16131.4 s	11123.8 s	9116.36 s	8380.39 s
(e)	1384467	5412303	28146.4 s	15489.6 s	10756.6 s	8993.03 s	8308.86 s
(f)	1369955	5656968	28445.8 s	14795.7 s	10383.7 s	8834.25 s	8194.06 s

Table 6
Time and space needed for model checking of considered properties

results, and further mathematical and biological analysis of counterexamples has to be done.

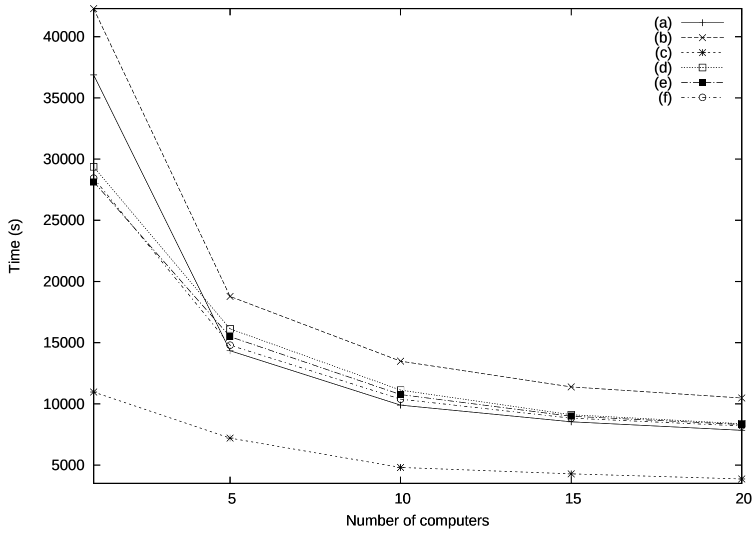

Fig. 4. Scaling of the experiments

We have employed the OWCTY model checking algorithm [8] on a homogeneous cluster of up-to 20 computers each based on dual-core Pentium 4 2× 2.6GHz with 4 GB RAM. Times needed for single computation of the properties on considered numbers of nodes are presented in Table 6. The values show that the most time and space consuming computation concerned the property (b). Scaling is showed graphically in Figure 4.

6 Conclusions

Results presented in the previous section show that the parallel approach can accelerate simulation and model checking of PLDE approximation of TRN dynamics. It allows models of the size relevant to current biological knowledge (around 10 variables), to be satisfactorily analysed. This has been demonstrated on a case study of a TRN responsible for sporulation of a well-known bacteria. A significant advantage of the parallel approach is the possibility to scale the computation to arbitrary number of computers while reducing the time needed for the analysis.

From the space complexity point of view, the on-the-fly approach is suitable for this kind of models because in contrast to the explicit approach it does not deal with the entire state space representation. Therefore it allows analysis of larger networks than is possible by the sequential approach. However, the exponential time complexity of the successor state generator still decreases the practicability of direct model checking of dynamics for all possible initial conditions.

To summarise, our work revisits in terms of parallel model checking the algorithm [11] for overapproximative quantitative analysis of a specific kind of biological networks. Although the scaling achieved is not ideal, our approach can be satisfactorily used for model checking of TRNs with up-to 15 genes. The most crucial is the expense of successor generation. Additionally, we have also showed how elementary network motifs can be used as the source of interesting dynamical properties. For the future work we aim to speed-up the generation of successors, possibly by considering an underapproximation of PLDE system solutions, and that way to reach better time results and scalability.

References

[1] U. Alon. *An Introduction to Systems Biology: Design Principles of Biological Circuits*. Chapman & Hall/CRC, 2006.

[2] J. Barnat, L. Brim, I. Černá, S. Dražan, and D. Šafránek. Parallel Model Checking Large-Scale Genetic Regulatory Networks with DiVinE. In *From Biology to Concurrency and Back, to appear in ENTCS*, 2007.

[3] J. Barnat, L. Brim, I. Černá, P. Moravec, P. Ročkai, and P. Šimeček. DiVinE – A Tool for Distributed Verification (Tool Paper). In *Computer Aided Verification*, volume 4144/2006 of *LNCS*, pages 278–281. Springer, 2006.

[4] G. Batt, C. Belta, and R. Weiss. Model checking genetic regulatory networks with parameter uncertainty. In A. Bemporad, A. Bicchi, and G. Butazzo, editors, *HSCC'07*, volume 4416 of *LNCS*, pages 61–75. Springer, 2007.

[5] G. Batt, D. Bergamini, H. de Jong, H. Garavel, and R. Mateescu. Model checking genetic regulatory networks using GNA and CADP. In *SPIN'04*, volume 2989 of *LNCS*, pages 158–163. Springer, 2004.

[6] G. Bernot, J-P. Comet, A. Richard, and J. Guespin. Application of Formal Methods to Biological Regulatory Networks: Extending Thomas' Asynchronous Logical Approach with Temporal Logic. *Journal of Theoretical Biology*, 229(3):339–347, 2004.

[7] L. Calzone, F. Fages, and S. Soliman. BIOCHAM: an environment for modeling biological systems and formalizing experimental knowledge. *Bioinformatics*, 22(14):1805–1807, 2006.

[8] I. Černá and R. Pelánek. Distributed explicit fair cycle detection (set based approach). In T. Ball and S.K. Rajamani, editors, *SPIN'03*, volume 2648 of *Lecture Notes in Computer Science*, pages 49 – 73. Springer Verlag, 2003.

[9] H. de Jong, J. Geiselmann, G. Batt, C. Hernandez, and M. Page. Qualitative simulation of the initiation of sporulation in Bacillus subtilis. *Bulletin of Mathematical Biology*, 66(2):261–300, 2004.

[10] H. de Jong, J. Geiselmann, C. Hernandez, and M. Page. Genetic network analyzer: qualitative simulation of genetic regulatory networks. *Bioinformatics*, 19(3):336–344, 2003.

[11] H. de Jong, J-L. Gouz, C. Hernandez, T. Sari, M. Page, and J. Geiselmann. Qualitative simulation of genetic regulatory networks using piecewise-linear models. Technical Report RR-4407, INRIA-Helix, 2002.

[12] et.al. G. Batt. Analysis and verification of qualitative models of genetic regulatory networks: A model-checking approach. In *IJCAI*, pages 370–375, 2005.

[13] J. Heath, M. Kwiatkowska, G. Norman, D. Parker, and O. Tymchyshyn. Probabilistic model checking of complex biological pathways. In C. Priami, editor, *CMSB'06*, volume 4210 of *Lecture Notes in Bioinformatics*, pages 32–47. Springer Verlag, 2006.

[14] A. Huerta, H. Salgado, D. Thieffry, and J. Collado-Vides. Regulondb: A database on transcriptional regulation in escherichia coli. *Nucleic Acids Res.*, 26(1):55–60, 1998.

[15] M. Kaufman and R. Thomas. Emergence of complex behaviour from simple circuit structures. *Comptes Rendus Biologies*, 326:205–214, 2003.

[16] Y. Makita, M. Nakao, N. Ogasawara, and K. Nakai. Dbtbs: database of transcriptional regulation in bacillus subtilis and its contribution to comparative genomics. *Nucleic Acids Res.*, 32:75–77, 2004.

[17] T. Mestl, E. Plahte, and S.W.Omholt. A mathematical framework for describing and analysing gene regulatory networks. *Journal of Theoretical Biology*, 176(2):291–300, 1995.

[18] R. Milo, S. Shen-Orr, S. Itzkovitz, N. Kashtan, D. Chklovskii, and U. Alon. Network motifs: simple building blocks of complex networks. *Science*, 298(5594):824–827, October 2002.

[19] M. Schaub, Th. Henzinger, and J. Fisher. Qualitative networks: a symbolic approach to analyze biological signaling networks. *BMC Systems Biology*, 1:4, 2007.

[20] C.A. Voigt, D.M. Wolf, and A.P. Arkin. The bacillus subtilis sin operon: an evolvable network motif. *Genetics*, 169(3):187–202, 2005.