Stable States of Boolean Regulatory Networks Composed Over Hexagonal Grids

Pedro L. Varela[a,b,1] Inês Lynce[a,2] Vasco Manquinho,[a,3]
Claudine Chaouiya[b,4] Pedro T. Monteiro[a,5]

[a] *INESC-ID/Instituto Superior Técnico, University of Lisbon, Portugal*

[b] *Instituto Gulbenkian de Ciência, Oeiras, Portugal*

Abstract

Cellular processes are governed by complex molecular regulatory networks. To understand the dynamics emerging from these networks, a popular approach relies on a Boolean abstraction. These Boolean regulatory networks define qualitative models with discrete dynamics in which properties of interest relate to the so-called attractors and their reachability. When considering multi-cellular systems, cell-cell communication must be accounted by properly inter-connecting cellular network models. This is done through logical composition rules that define cell-cell communication, leading to a (composed) model of the regulatory control of the whole.

This work focuses on Boolean models composed over hexagonal grids, which suitably represent simple epithelia. Stable states embody stable patterns of the grid, *i.e.*, ensembles of differentiated cell types, each characterized by a specific pattern of gene expression. Identification of these stable states is a challenging problem due to the combinatorial explosion of the dimensions of the model state space and to the potentially huge number of solutions.

Following the formalisation of model composition, we present a SAT-based method to identify the stable states of Boolean models composed over hexagonal grids. This approach is applied to a few prototypical case studies, illustrating counter-intuitive dependencies of the number of stable states on: the cellular model, the composition rule and the grid characteristics (*i.e.*, dimensions and border conditions).

Keywords: Stable patterns, multi-cellular systems, Boolean regulatory networks, model composition, SAT encoding

1 Introduction

Molecular components (genes, transcription factors, proteins, etc.) interact and define complex networks, which control cellular processes (differentiation, proliferation, death, etc.). Appropriate computational tools are needed to analyse the be-

[1] Email:Pedro.Varela@tecnico.ulisboa.pt
[2] Email:Ines.Lynce@tecnico.ulisboa.pt
[3] Email:Vasco.Manquinho@tecnico.ulisboa.pt
[4] Email:chaouiya@igc.gulbenkian.pt
[5] Email:Pedro.Tiago.Monteiro@tecnico.ulisboa.pt

haviours of such regulatory networks. In this context, the logical formalism defines discrete models in which each regulatory component is associated with a Boolean (or multi-valued) variable and a logical function defining the evolution of this variable depending on the component's regulators [2,5,7]. Such models are amenable to formal methods to characterise the network dynamics and, in particular, to identify attractors and their reachability properties. Attractors are of the utmost importance as they relate to long term behaviours of the modelled processes. An attractor reduced to a single state is referred to as a *stable state*, which generally embodies a differentiated cellular state. Complex attractors encompass several states and denote oscillatory behaviours.

Various methods have been devised to determine stable states in logical models, a problem known to be NP-hard [3]. Using a Binary Decision Diagram representation of the dynamics, Garg *et al.* proposed an efficient algorithm to recover the stable states of Boolean models [17]. Also relying on the exploration of the dynamics, a SAT-based bounded model checking approach allows to find all the attractors of Boolean models under a synchronous update [14]. In contrast, some methods search for the stable states as fixed points of the transition function. For example, constraint programming can be used to solve stability equations [12], or yet the table states of multi-valued logical models are recovered by properly defining a decision diagram representation of the stability condition which allows to recover the stable states [23].

Dealing with models of multi-cellular systems worsens the problem of the combinatorial explosion of, not only the size of the state spaces, but also, potentially, of the number of stable states. This precludes the use of methods currently available, calling for novel approaches.

In this work, we consider Boolean models of simple epithelia (*i.e.* single layers of cells) defined as the composition of cellular models over 2D grids of hexagonal cells. We propose to use a SAT-based approach to determine all the stable states of such models. Cellular models were defined using GINsim, a tool for the logical modelling of regulatory networks equipped with SAT encoding features (available at http://ginsim.org) [22]. We assume that all the cells are controlled by the same internal regulatory network, which is modelled using the logical formalism. A composed model is defined by specifying logical rules governing the interactions between instances of this cellular model. Using a similar rational, several logical models of pattern formation within multi-cellular systems have been published, particularly in the fruit fly development [4,15,26]. In the same vein, for multi-cellular logical models under an asynchronous update, Mendes *et al.* relied on a process algebra framework to define a compositional procedure that uncovers the stable states reachable from an initial condition, a very challenging problem, since the dynamics has to be considered [21].

This paper is organised as follows. Section 2 introduces the basics of the logical framework, covering the modelling of single cell regulatory networks and their composition. This composition is specified through integration functions that define the rules of cell-cell communication, and are appropriately expressed in terms

of cardinality constraints. Section 3 then presents two SAT encodings to retrieve the stable states of a composed model over a 2D hexagonal grid. Using the integration functions, the first encoding assembles the stability conditions of all the single cell model instances. The second encoding specifies compatibility rules between abstracted representations of the cellular stable states. Besides discussing the performances of both encodings, Section 4 is mainly devoted to uncovering trends in the evolution of the numbers of stable states, using case studies to identify the parameters that matter. We conclude with a discussion of our current results and future work.

2   Logical Regulatory Graphs and Composition

This section introduces the logical modelling formalism and model composition. We refer to [2,5,7] for further details on the modelling framework.

2.1   Logical Regulatory Graph (LRG)

As formalised in Definition 2.1, an LRG is a graph in which the nodes embody genes or, more generally, regulatory components and the directed, signed edges represent regulatory interactions between these components. Moreover, variables are associated with components to denote their states (ON/OFF or yet *true/false*), and logical functions specify the evolution of these variables.

Definition 2.1 A *Logical Regulatory Graph* is defined by a triplet $N = (G, U, F)$ with:

- $G = \{g_i\}_{i \in L_G}$ is the indexed set of the *internal components* (L_G being the corresponding set of indices);

- $U = \{u_i\}_{i \in L_U}$ is the indexed set of the *input components* (L_U being the corresponding set of indices, with $L_G \cap L_U = \emptyset$);

- each component of $G \cup U$ is associated with a Boolean variable: $\forall i \in L_G \cup L_U, x_i \in \{0, 1\}$ denotes the state of the component with index i;

- the state space S is given by $\{0, 1\}^{|G \cup U|}$ and $x \in S$ denotes a state of the LRG;

- $F = \{f_i\}_{i \in L_G \cup L_U}$ is the *transition function* specified by the *logical functions* defining the evolution of the component states; $\forall i \in L_G \cup L_U$, $f_i : S \to \{0, 1\}$, and at any state x the future value of x_i is given by $f_i(x)$.

Input components can be viewed as receptors at the cellular membrane, receiving external signals. They represent the environment and, in the context of a cellular model, their levels are assumed constant. Therefore, the logical function associated to any input component is the identity: $\forall u_i \in U, \forall x \in S, \ f_i(x) = x_i$. Figure 1 provides an illustration of a simple LRG. From now on, to denote a component, we will indistinctly use its index i rather than g_i (resp. u_i) for an internal component (resp. input component).

In this work, we focus on Boolean LRGs as specified in Definition 2.1 (in contrast

A) Regulatory graph   **B)** Logical functions   **C)** Stable states

u_0	g_1	g_2	g_3	
0	0	0	0	SS1
0	1	1	1	SS2
1	0	0	0	SS1

$$f_0(x) = x_0$$

$$f_1(x) = \neg x_0 \wedge x_3$$
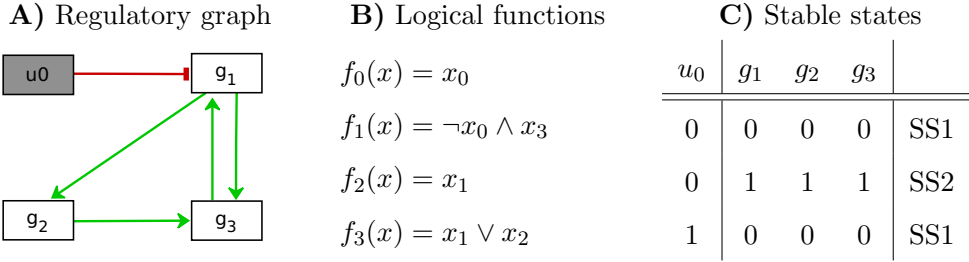
$$f_2(x) = x_1$$

$$f_3(x) = x_1 \vee x_2$$

Figure 1. The logical formalism illustrated: **A)** LRG with 3 internal components ($G = \{g_1, g_2, g_3\}$) and a single input component ($U = \{u_0\}$); activations are represented by green arrows (*e.g.*, g_1 activates g_2 and g_3), the red blunt arrow represents an inhibition (u_0 inhibits g_1). **B)** Logical functions defining the evolution of the variables ($\neg$ denotes the logical negation, $\wedge$ the conjunction and $\vee$ the disjunction). **C)** The two model stable states, where a stable state is characterised by the values of the internal components and may be possible for different values of the input component.

to multi-valued LRGs whose variables can take a broader range of discrete values). This is not a limitation because any multi-valued LRG can be formally mapped to a Boolean LRG [13]. Indeed, this mapping was been used in our first case study in Section 4.

An LRG defines a discrete dynamics over its (finite) state space, for each state $x \in S$, if $x_i < f_i(x)$ then x_i will increase (*i.e.* x_i will be set to 1 or *true*, reflecting the fact that g_i is activated); if $x_i > f_i(x)$ then x_i will decrease (*i.e.* x_i will be set to 0 or *false*, reflecting the fact that g_i is inhibited); if $x_i = f_i(x)$ then x_i remains unchanged *i.e.* g_i is stable in state x.

Definition 2.2 Given an LRG $N = (G, U, F)$ and its state space S, a state $x \in S$ is *stable* if and only if it satisfies the *stability condition*:

$$F(x) = x. \tag{1}$$

Hence, stable states are the fixed points of the function F. Note that to satisfy the stability condition above, it is enough to verify that $\forall g_i \in G$, $f_i(x) = x_i$. This is because input components are maintained constant.

2.2 Composition of Logical Regulatory Graphs

Cells communicate through signals emitted by specific internal components that, in biological terms, correspond to *e.g.* secreted proteins; these signals are received by input components of neighbouring cells. For a multi-cellular system encompassing p communicating cells, we build a composed logical model consisting of p LRGs instances properly connected by defining integration functions, which specify how signals are combined to define the levels of some input components (see Figure 2). Definition 2.3 formalises this composition procedure.

Definition 2.3 Given p identical LRGs $N^k = (G^k, U^k, F^k)$, $k = 1, \ldots, p$, the composition rule $\mathcal{M}(u_i^k) = (Z_i^k, h_i^k)$, for all $u_i^k \in U^k$, is defined over the p sets of input components with:

- $Z_i^k \subseteq \bigcup_{k' \neq k} G^{k'}$, the set of internal components influencing u_i^k and,

- $h_i^k : \{0,1\}^{|Z_i^k|} \to \{0,1\}$, the *logical integration function* defining the value of u_i^k.

If $Z_i^k \neq \emptyset$, u_i^k is a *mapped input component*, otherwise it remains an input component and h_i^k is the identity function.

Then the composition of the p LRGs, denoted $\bigotimes_{\mathcal{M}} \{N^k\}_{k=1,\ldots,p}$, is an LRG $N = (G, U, F)$ where:

- the set of *internal components* is $G = \left(\bigcup_{k=1,\ldots,p} G^k \right) \cup \left(\bigcup_{k=1,\ldots,p} \widetilde{U}^k \right)$, where $\widetilde{U}^k = \{u_i^k \in U^k \text{ s.t. } Z_i^k \neq \emptyset\}$ is the set of mapped input components of N^k;

- the set of *input components* is $U = \bigcup_{k=1,\ldots,p} \left(U^k \setminus \widetilde{U}^k \right)$;

- the state space is $S = \{0,1\}^{|G \cup U|}$;

- a mapping *ind* reindexes all the components with, $\forall k = 1, \ldots, p$:
  $\forall g_i^k \in G^k \cup \widetilde{U}^k$, $ind(g_i^k) \in L_G$ and $\forall u_i^k \in U^k \setminus \widetilde{U}^k$, $ind(u_i^k) \in L_U$;

- the logical regulatory functions are defined adequately: $\forall j \in L_U$, f_j is the identity function and $\forall j \in L_G$,
  - if the component is such that $ind(g_i^k) = j$ is in G^k, *i.e.*, is an internal component of N^k, then

$$\forall x \in S, \ f_j(x) = f_i^k(x_{|S^k}),$$

  where $x_{|S^k}$ is the projection of x onto S^k, the state space of N^k;
  - if the component such that $ind(u_i^k) = j$ is in $\widetilde{U}^k$ with $\mathcal{M}(u_i^k) = (Z_i^k, h_i^k)$, *i.e.*, is a mapped input of N^k, then

$$\forall x \in S, \ f_j(x) = h_i^k(x_{|S^k}).$$

Figure 2 illustrates this model composition. For simplicity, we assume that the LRGs to be composed are identical instances. Moreover, the composition rules of the several instances of any input component are also assumed to be identical. The generalisation to the composition of distinct LRGs would be straightforward (see [21]) as well as the consideration of distinct composition rules.

Given a composed model $\bigotimes_{\mathcal{M}} \{N^k\}_{k=1,\ldots,p}$, its stable states are the solutions of Equation 1. However, the following property asserts that we can also recover these stable states from those of the single cell LRGs N^k. For notation convenience, we assume that the mapping *ind* verifies:

- $\forall k \in \{1, \ldots, p\}$, $\forall c_i^k, c_{i'}^k \in G^k \cup U^k$ such that $i < i'$, $ind(c_i^k) < ind(c_{i'}^k)$ (the order of the components in N^k is preserved) and,

- $\forall k, k' \in \{1, \ldots, p\}$ such that $k < k'$ and $\forall c_i^k \in G^k \cup U^k$, $\forall c_i^{k'} \in G^{k'} \cup U^{k'}$, $ind(c_i^k) < ind(c_i^{k'})$ (the order of the LRG instances is preserved).

Property 1 *Given a composed model $(G, U, F) = \bigotimes_{\mathcal{M}} \{N^k\}_{k=1,\ldots,p}$ and its state space $S = \bigcup_{k=1,\ldots,p} S^k$, a state $x = (x^1, x^2, \ldots, x^p) \in S$ (with $x^k \in S^k, k = 1, \ldots, p$),*

is stable if and only if:

$$\forall k = 1, \ldots, p, F^k(x^k) = x^k \ and \tag{2}$$

$$\forall j \in L_G \ s.t. \ j = ind(g_i^k) \ with \ g_i^k \in \widetilde{U}^k, \ f_j(x) = x_j. \tag{3}$$

Property 1 asserts that a state of the composed model is stable if and only if it is defined as the concatenation of *compatible* stable states of its constituent LRGs. More precisely, Equation 2 indicates that the states of each LRG must be stable when not affected by the composition. Furthermore, Equation 3 specifies the *compatibility rule* for the composition of stable states instances: the values of the mapped input components must remain stable considering the values of their regulators (internal components of neighbouring LRGs, as specified by $\mathcal{M}$). Although rather straightforward, the proof is cumbersome due to convoluted notation.

In practice, to compose LRGs, one needs to specify how they influence each other. This is done through the composition rule $\mathcal{M}$, where the definition of the sets Z_i^k depends on the topology of the system. Here, as previously mentioned, we consider that all the cells encompass the same LRG, are hexagonal and arranged over a 2D grid. This leads to the definition of a simple *neighbouring relation*,

A) Model composition, 2x2 cells

B) Integration function: OR

$$h_0^1(x) = x_1^2 \vee x_1^3$$

at-least-1 neighbour with $x_1 = 1$

Stable states

x_1^1	x_2^1	x_3^1	x_1^2	x_2^2	x_3^2	x_1^3	x_2^3	x_3^3	x_1^4	x_2^4	x_3^4
0	0	0	0	0	0	0	0	0	0	0	0
1	1	1	0	0	0	0	0	0	0	0	0
0	0	0	1	1	1	0	0	0	0	0	0
0	0	0	0	0	0	1	1	1	0	0	0
0	0	0	0	0	0	0	0	0	1	1	1
1	1	1	0	0	0	0	0	0	1	1	1

C) Integration function: AND

$$h_0^1(x) = x_1^2 \wedge x_1^3$$

at-least-v neighbours with $x_1 = 1$

Stable states

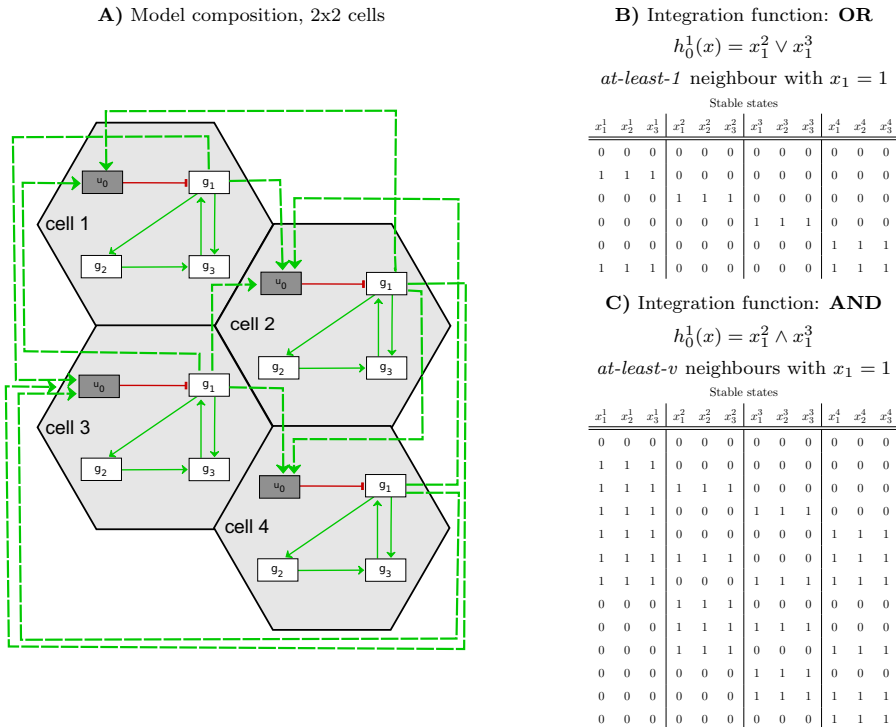| x_1^1	x_2^1	x_3^1	x_1^2	x_2^2	x_3^2	x_1^3	x_2^3	x_3^3	x_1^4	x_2^4	x_3^4
0	0	0	0	0	0	0	0	0	0	0	0
1	1	1	0	0	0	0	0	0	0	0	0
1	1	1	1	1	1	0	0	0	0	0	0
1	1	1	0	0	0	1	1	1	0	0	0
1	1	1	0	0	0	0	0	0	1	1	1
1	1	1	1	1	1	0	0	0	1	1	1
1	1	1	0	0	0	1	1	1	1	1	1
0	0	0	1	1	1	0	0	0	0	0	0
0	0	0	1	1	1	1	1	1	0	0	0
0	0	0	1	1	1	0	0	0	1	1	1
0	0	0	0	0	0	1	1	1	0	0	0
0	0	0	0	0	0	1	1	1	1	1	1
0	0	0	0	0	0	0	0	0	1	1	1

Figure 2. Composition of the toy model of Figure 1. **A)** A 2×2 2D hexagonal grid, displaying the cellular networks and cell-cell communication; the integration input u_0 is regulated by (*i.e.* integrates signals from) g_1 of neighbouring cells (*e.g.* $Z_0^1 = \{g_1^2, g_1^3\}$ and $Z_0^2 = \{g_1^1, g_1^3, g_1^4\}$). **B)** Stable states of the composed model when the integration function of u_0 is a disjunction. **C)** Stable states of the composed model when the integration function of u_0 is a conjunction. Specification of the integration function as a cardinality constraint is indicated, v denoting the number of neighbours (v=2 for cell 1, v=3 for cell 2). Note that, different integration functions lead to different set of stable states.

depending on a distance: cells have up to 6 direct neighbours (contacting cells on the grid, *i.e.* at distance 1) and possibly more neighbours at greater distances. If the neighbouring relation specifies a distance 2, a cell has up to $6 + 12 = 18$ neighbours. The number of neighbours thus depends on the distance, but also on the position of the cell in the grid (less neighbours for cells at the grid borders than at its center).

3 SAT encoding

In GINsim, a software tool for the logical modelling of cellular regulatory networks [22], the stability condition of a model component (Equation 1) is encoded in the form of an ordered decision diagram, and the conjunction of these conditions for all the components is efficiently computed [23]. More precisely, the resulting decision diagram encodes the stability conditions of the whole model as a Disjunctive Normal Form (DNF) formula, where each conjunctive clause in the formula is given by a path from the root to a leaf labelled *true*. Nevertheless, as SAT solvers support Conjunctive Normal Form (CNF) formulas, we retrieve the equivalent CNF formula. This CNF formula includes one clause for each path from the root to a leaf labelled *false*, each clause being the disjunction of the negated literals defining the path. Our SAT encoding relies on this export feature of GINsim for stability conditions of cellular models (as defined in Equation 1).

Logical integration functions are more easily specified as *cardinality constraints* over the value(s) of internal components across neighbours. We denote a cardinality constraint by $\#_q(y_1, \ldots, y_v)$ where q is the cardinal restriction, v is the number of variables over which the restriction applies, and $\#$ is an operator in $\{\leq, \geq\}$, representing *at-most-q* or *at-least-q*, respectively. The compatibility rule as stated in Equation 3 thus requires, for each mapped input component u_i^k, the encoding of a constraint $u_i^k \iff \#_q(y_1, \ldots, y_v)$, that is:

$$\left(\neg u_i^k \vee \#_q(y_1, \ldots, y_v)\right) \wedge \left(u_i^k \vee \neg\#_q(y_1, \ldots, y_v)\right). \tag{4}$$

The SAT community has addressed the encoding of cardinality constraints, focusing mostly on the *at-most-1* constraint [9,19,24]. A survey of these encodings goes beyond the scope of this paper, and we refer to [8,16] for additional details. In this work, we use the classical Binomial encoding. Different SAT encodings generate different formulas (varying in the numbers and sizes of the clauses, and possibly requiring additional variables). The performance of a SAT solver depends on the problem characteristics and on a trade-off between the numbers and sizes of the clauses and the number of variables [1,24].

Our first approach relies on the composition rule to combine the stability conditions of all the cellular model instances, thus defining the stability condition of the multi-cellular, composed model (*i.e.,* "compose first, compute afterwards"). Our second approach computes the stable states of the cellular model, which are then combined using the compatibility rules as specified in Equation 3 (*i.e.,* "compute

first, compose afterwards").

3.1 Stability condition of a composed model

The specification of a composed model includes p LRG instances $N^k, k = 1, \ldots, p$ and the composition rule $\mathcal{M}(u_i^k) = (Z_i^k, h_i^k)$ for each mapped input component. For each mapped input component, the integration function is defined as a *cardinality constraint* over a set of internal components of neighbouring cells (elements of Z_i^k). If an input receives signals from distinct internal components (*i.e.*, $\exists g_{i_1}^{k'}$ and $g_{i_2}^{k'}$ in Z_i^k with $i_1 \neq i_2$), its integration function may combine cardinality constraints over these distinct regulators. For simplicity, here we consider that this is not the case, *i.e.*, an input component is mapped to the same, unique internal component of neighbouring cells. The integration function h_i^k is thus defined over a set of Boolean variables, $(y_1, \ldots, y_v)$, where each y_i refers to a $x_i^{k'}$, the level of the internal component $g_i^{k'}$ in Z_i^k, and v is the number of neighbours of N^k. To encode the stable states of the composed model, we first encode the stability condition of all internal components (Equation 2), and then the integration function of the mapped inputs (Equation 3), expressed as a cardinality constraint (Equation 4).

In the sequel, we will refer to this SAT encoding as *Stability condition*.

3.2 Composition of cellular stable states

Here, we first determine the set of all the stable states of the LRG to be composed and, to each of those states a Boolean variable is associated. The compatibility rules (see Property 1) are then encoded as follows, for each mapped input component u_i:

$$\#_q(\sigma_1^{u_i}, \ldots, \sigma_v^{u_i}) \Longrightarrow \Sigma^{u_i}, \tag{5}$$
$$\neg\#_q(\sigma_1^{u_i}, \ldots, \sigma_v^{u_i}) \Longrightarrow \Sigma^{\neg u_i}, \tag{6}$$

where, for all $j = 1, \ldots, v$, the Boolean variable $\sigma_j^{u_i}$ is *true* if the jth neighbour is in a stable state for which the internal component mapped to u_i is at the required level to activate u_i. In other words, each $\sigma_j^{u_i}$ accounts for the disjunction of the stable states compatible with a *true* value of the input component u_i. Moreover, Σ^{u_i} (resp. $\Sigma^{\neg u_i}$) is the disjunction of the Boolean variables associated to the LRG stable states for which $u_i = 1$ (resp. $u_i = 0$).

Unlike the equivalence in Equation 4, we have an implication because a stable state can be possible for both values of the input component, and thus the associated Boolean variable will appear in both Σ^{u_i} and $\Sigma^{\neg u_i}$ (see example in 4.1).

Note that if there was no cell-cell communication (*i.e.*, no mapped inputs), the number of stable states would be the cardinal of Σ to the power of the number of LRG instances in the composed model. This gives an upper bound of the number of stable states of a composed model, and when some inputs are mapped, the compatibility rules lead to a lower number of stable states.

In the sequel, we will refer to this SAT encoding as *Compatibility condition*.

4 Case studies

We first present a case study, which corresponds to a concrete regulatory module and mainly serves to compare our two SAT encodings for the identification of the stable states of a composed model. Then, three synthetic models are presented to illustrate the diverse parameters that influence the numbers of stable states.

Note that, because the number of stable states may grow exponentially with the number of cells in the grid, to perform the tests presented below, we settle for counting the number of stable states instead of characterising their actual identity. To this intend, we use the model counter *Cachet* [27].

4.1 Segment-Polarity

Along its development, the fruit fly embryo is segmented in repeated units, which give rise to specific anatomical structures of the adult fly (*e.g.* thorax, wings, etc.). This process is controlled by a set of genetic modules among which the segment-polarity module establishes the borders between the embryo segments. Here, we consider the logical model proposed by Sánchez *et al.* for this regulatory module as depicted in Figure 3-A) [26]. To account for the formation of the segmental borders across the embryo, Sánchez *et al.* composed the segment-polarity model over a row of 6 cells. Figure 3 further provides the cellular model stable states, the integration functions used for model composition in addition to compatibility rules and characteristics of the two SAT encodings (*i.e.* the Stability condition *versus* the Compatibility condition, see Section 3). The performances of these encodings are assessed by increasing the number of rows in the hexagonal grid and thus the number of LRG instances in the composed model (see Figure 4).

As shown in Figure 4-A, the Compatibility condition encoding yielded better performances, allowing us to explore grids up to 6 rows, while the Stability condition encoding only permitted the counting of stable states for grids up to 4 rows (given a timeout of 14000 seconds). This gain in performance is mainly due to the fact that, for this specific model, the Compatibility condition encoding requires less variables and less clauses (see Figure 3-E). In general, we expect that the Compatibility condition encoding should perform better than the Stability condition encoding when the number of LRG components is higher than the number of cellular stable states. This is the case for the segment-polarity model as well as for most realistic models. To analyse the remaining case studies, the Compatibility condition encoding was used.

Moreover, Figure 4-B shows that, although drastically limited by the composition (cell-cell communication), the number of stable states increases exponentially with the number of rows.

Although the number of stable states increases exponentially, some of pairs of states are symmetric, as observed by Sánchez *et al.* and as illustrated in Figure 3-F [26]; indeed, there are 65 stable states for a 6×1 grid but, if we discard symmetrical states (*i.e.,* considering only one of them), this number drops to 29. This symmetry property could be used to reduce the search space of grid stable configurations.
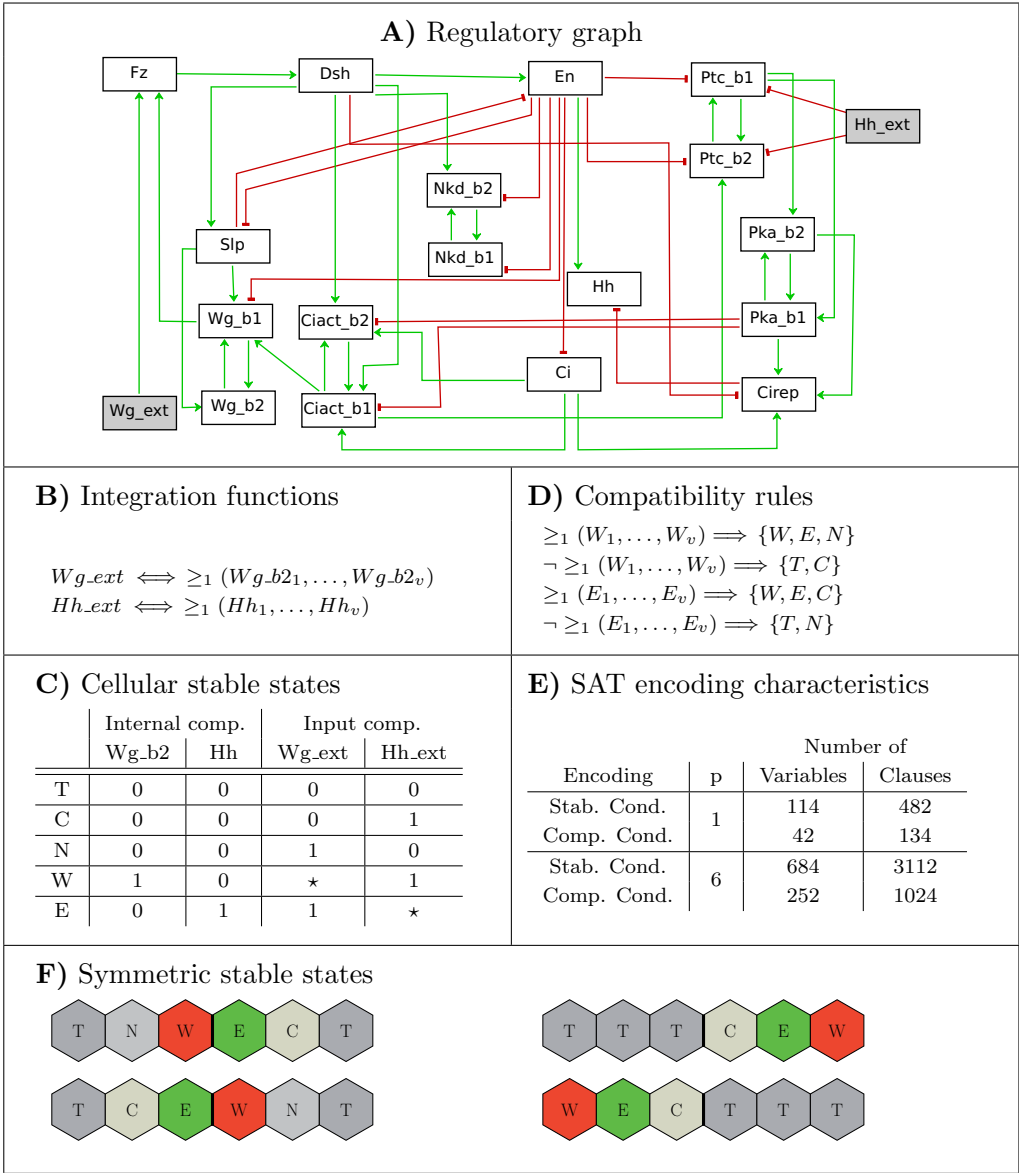
A) Regulatory graph

B) Integration functions

$$Wg\_ext \iff \geq_1 (Wg\_b2_1, \ldots, Wg\_b2_v)$$
$$Hh\_ext \iff \geq_1 (Hh_1, \ldots, Hh_v)$$

D) Compatibility rules

$$\geq_1 (W_1, \ldots, W_v) \implies \{W, E, N\}$$
$$\neg \geq_1 (W_1, \ldots, W_v) \implies \{T, C\}$$
$$\geq_1 (E_1, \ldots, E_v) \implies \{W, E, C\}$$
$$\neg \geq_1 (E_1, \ldots, E_v) \implies \{T, N\}$$

C) Cellular stable states

	Internal comp.		Input comp.	
	Wg_b2	Hh	Wg_ext	Hh_ext
T	0	0	0	0
C	0	0	0	1
N	0	0	1	0
W	1	0	$\star$	1
E	0	1	1	$\star$

E) SAT encoding characteristics

		Number of	
Encoding	p	Variables	Clauses
Stab. Cond.	1	114	482
Comp. Cond.		42	134
Stab. Cond.	6	684	3112
Comp. Cond.		252	1024

F) Symmetric stable states

Figure 3. Segment-Polarity model: **A)** LRG as defined in [26], where multi-valued components have been duplicated following the Boolean mapping defined in [13]; *e.g.* Wg_b1 and Wg_b2 replace the multi-valued Wg. **B)** Integration functions of the two input components: *e.g.* Wg_ext is 1 iff *at-least-1* neighbour has Wg_b2 equals to 1. **C)** The segment-polarity LRG has 5 stable states; the table names those states and provides the corresponding values of the inputs and relevant internal components (that influence the integration functions). The symbol $\star$ indicates both values 0 and 1. **D)** Compatibility rules: *e.g.* if *at-least-1* neighbour is in state W, then the considered LRG instance can be either in state W, E or N. **E)** CNF characteristics for the two SAT encodings: number of variables and clauses for grids of sizes 1×6 and 6×6. **F)** Illustration of two pairs of symmetric stable states over a 1×6 grid as described in [26].

Finally, it may be worth noting that here all possible stable states are counted and that reachability from specific initial conditions is not considered, which would not only greatly reduce the number of (reachable) stable states, but also provide grid stable states with a biological relevance.
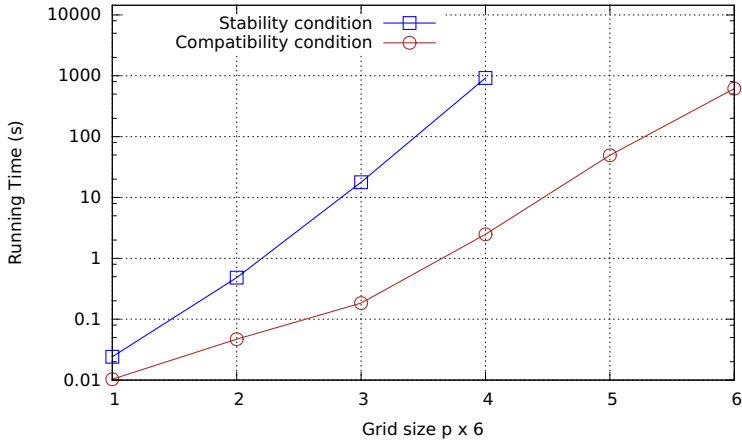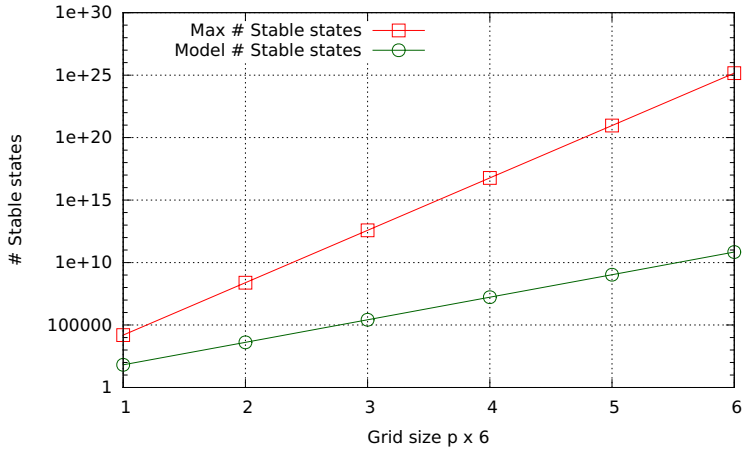
Figure 4. Segment-polarity module composed over a $p \times 6$ hexagonal grid (composition rules given in Figure 3-D). **A)** Running times for the two SAT encodings when varying p the number of rows. **B)** Number of stable states when varying p the number of rows (this number obviously does not depend on the chosen SAT encoding); curves depict the number of stable states of the composed model (in green) and 5^{6p}, the upper bound of the number of stable states (in red). All plots are on a logarithmic scale.

4.2 Toy model

We define the composition of the toy model defined in Figure 1 over a 4×4 grid, with a composition rule set to: *at-least-q* neighbours with g_1 at level 1 ($x_1 = 1$). Assuming that the neighbouring distance is 2, we observe that the number of stable states depends on the value of q in the cardinality constraint. However, due to the borders of the grid, not all the cells have the same number of neighbours (see Figure 5). To assess how borders affect the number of stable states, we consider an horizontal roll-over (*i.e.,* defining an hexagonal cylinder), which makes the left border of the grid adjacent to the right border, and a double roll-over (*i.e.,* defining an hexagonal torus).

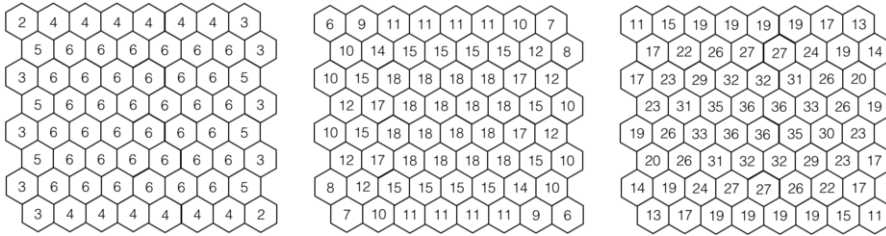Results show that the number of stable states of the composed model increases

Figure 5. Numbers of neighbours in a 8 × 8 hexagonal grid, for distances 1, 2 and 3. Numbers is the cells indicate the number of their neighbours.

with q, independently of the border existence (Figure 6-A). This is not surprising, because as q (the number of required neighbours) tends to v (the number of actual neighbours), the restriction on the state of neighbouring cells is harder to fulfill (as more cells are required to be in SS2). If the cardinality constraint is not satisfied, more configurations are possible, since a cell may be either in SS1 or SS2, explaining the increase of the number of stable states.

The compatibility rule influences the number of stable states. Indeed, modifying the toy model in such a way that the compatibility rule asserts that a cell is in state SS1 or SS2 when the cardinality constraint is satisfied and otherwise is necessarily in SS1, a similar reasoning explains the decreasing number of stable states (Figure 6-B); as the cardinality constraint is harder to satisfy, less cells are in SS2, and the grid quickly evolves to a stable configuration where all the cells are in SS1.

The consideration of roll-over limits the effect of borders, meaning that more cells are able to satisfy the cardinality constraints for higher q. For the toy model, roll-over leads to a decrease of the number of stable states, with the cylinder and the torus having a very close number of stable states. On the other hand, for the modified toy model the influence of the roll-over on the number of stable states is more visible, since limiting the border effect increases the number of cells that are likely to comply with the cardinality constraint, resulting in a higher number of stable states.

4.3   Model of lateral inhibition

Lateral inhibition is a mechanism widely used in developmental patterning, and it has been the subject of significant modelling studies (*e.g.,* see [11]). Notch signalling is one of the most classical pathways implementing lateral inhibition, where cell-cell interaction is mediated by the transmembrane proteins Notch and Delta. The simplest model assumes that Notch is activated by Delta from neighbouring cells and , in turn, activated Notch inhibits Delta in the cell, which generates a positive feedback loop between neighbouring cells. The Delta-Notch mediated signalling plays a crucial role is establishing the diversity of cell types within a tissue in the embryonic development of a wide range of organisms [10]. Here, we consider a lateral inhibition synthetic model, presented in Figure 7.

Figure 8 illustrates the influence of the neighbouring distance on the number of stable states. Considering signalling distances ranging from 1 to 3, we observe that
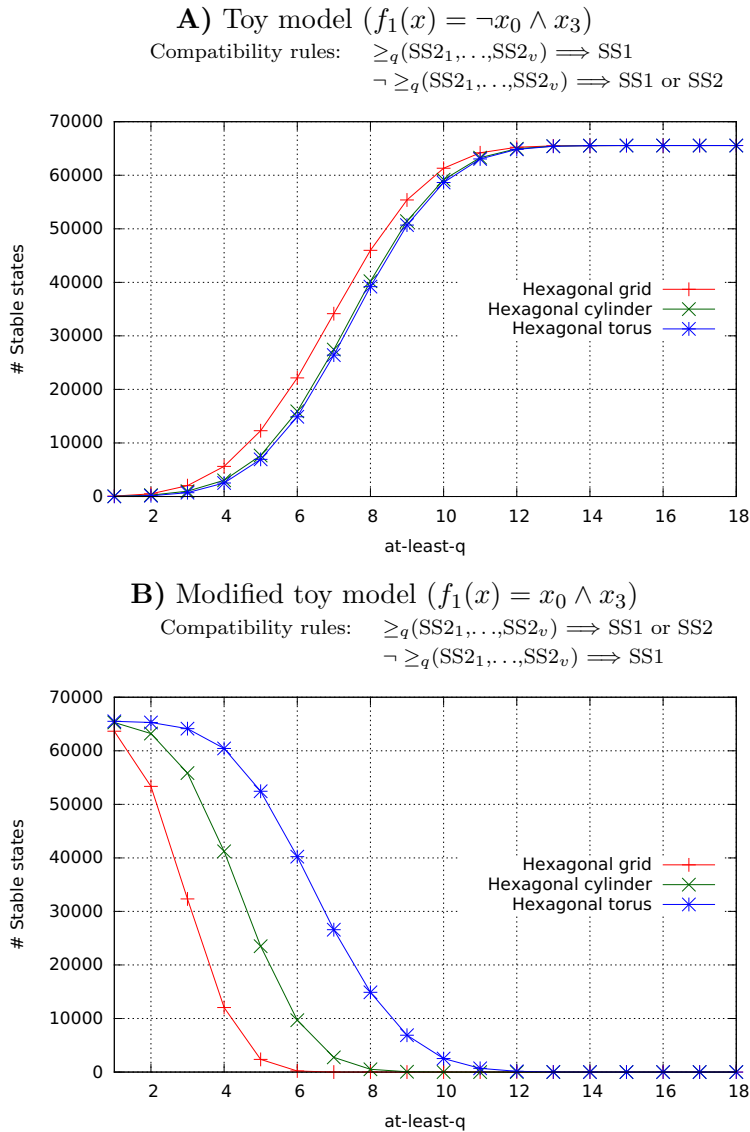
A) Toy model $(f_1(x) = \neg x_0 \land x_3)$

Compatibility rules: $\geq_q(\text{SS2}_1,\ldots,\text{SS2}_v) \implies \text{SS1}$
$\neg \geq_q(\text{SS2}_1,\ldots,\text{SS2}_v) \implies \text{SS1 or SS2}$

B) Modified toy model $(f_1(x) = x_0 \land x_3)$

Compatibility rules: $\geq_q(\text{SS2}_1,\ldots,\text{SS2}_v) \implies \text{SS1 or SS2}$
$\neg \geq_q(\text{SS2}_1,\ldots,\text{SS2}_v) \implies \text{SS1}$

Figure 6. Number of stable states for models composed over a 4×4 cellular configuration, with three different border conditions and varying q in the cardinality constraint *at-least-q*. **A)** Toy model (as defined in Figure 1): the logical function of the internal component g_1 is recalled, the compatibility rules are given (LRG stable states given Figure 1), and the numbers of stable states are plotted. **B)** Modified toy model, where the input now activates g_1: evolution of the number of stables is reverted.

the number of stable states decreases with the distance. Indeed, as the neighbouring distance increases, the cardinality constraint is easier to fulfill. More precisely, as illustrated in Figure 9, if a cell is in state SS2, its neighbours are necessarily in SS1, and the largest the distance the more cells are restricted by this compatibility rule. Moreover, increasing the size of the grid leads to an increase of the number of stable states, independently of the distance.
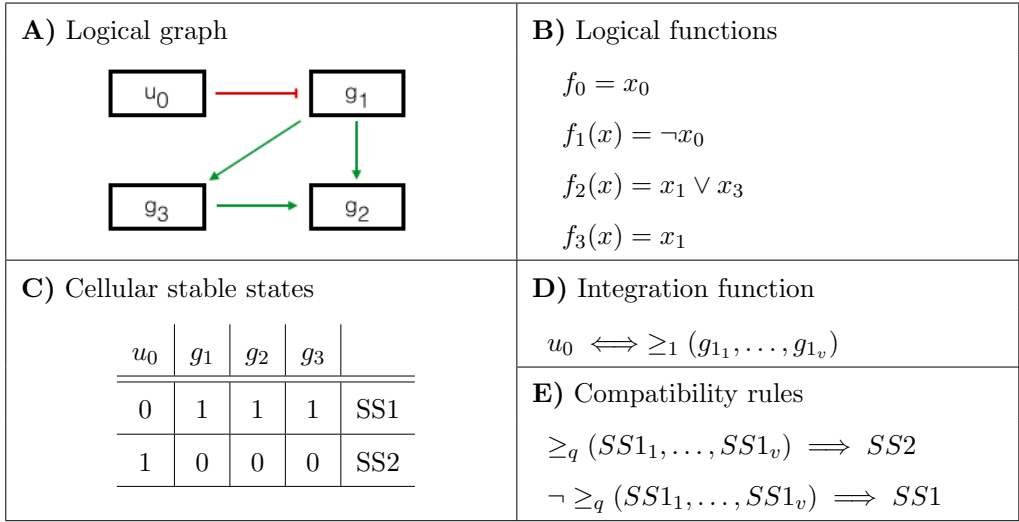
A) Logical graph	**B)** Logical functions

A) Logical graph

$u_0 \longmapsto g_1$

$g_3 \longrightarrow g_2$

B) Logical functions

$f_0 = x_0$

$f_1(x) = \neg x_0$

$f_2(x) = x_1 \lor x_3$

$f_3(x) = x_1$

C) Cellular stable states

u_0	g_1	g_2	g_3	
0	1	1	1	SS1
1	0	0	0	SS2

D) Integration function

$$u_0 \iff \geq_1 (g_{1_1}, \ldots, g_{1_v})$$

E) Compatibility rules

$$\geq_q (SS1_1, \ldots, SS1_v) \implies SS2$$

$$\neg \geq_q (SS1_1, \ldots, SS1_v) \implies SS1$$

Figure 7. Lateral inhibition model: **A)** LRG with 3 internal components and 1 input component. **B)** Logical functions governing the values of the internal components. **C)** Stable states of the single cell model. **D)** Integration function: the mapped input is 1 iff *at-least-1* neighbour has g_1 at level 1. **E)** Compatibility rules: if *at-least-1* of its neighbours is in state SS1, a cell is then in SS2.
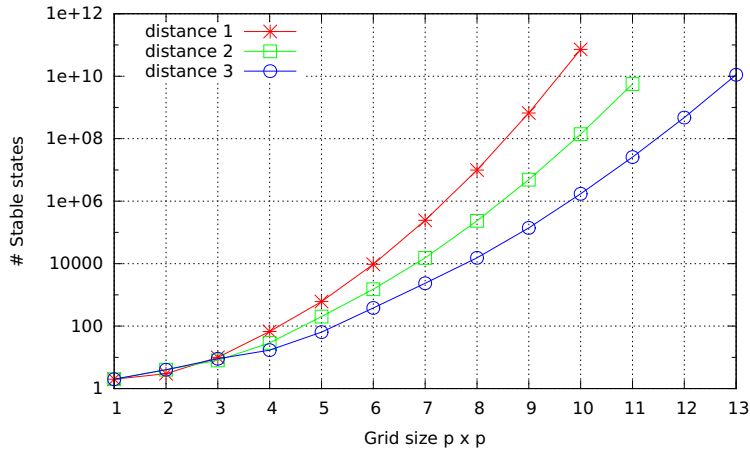
Figure 8. Lateral inhibition model composed over a $p \times p$ grid: evolution of the numbers of stable states, for different neighbouring distances and varying the grid size p. Plots are on a logarithmic scale.

4.4 Model of lateral induction

Similarly to lateral inhibition, lateral induction is a signalling mechanism for pattern formation [25]. Figure 10 illustrates a lateral induction synthetic model. Unlike the lateral inhibition model (Figure 7), the input component u_0 now activates g_1, which in turns activates the input in neighbouring cells.

To assess the influence of the neighbouring distance on the number of stable states, we reproduce the same experiment as in Figure 8, *i.e.*, we vary the size of the grid for different distances (see Figure 11). However, now if a cell is surrounded by q cells in state SS2, it also adopts this state. Hence, for a small q (*e.g.*, $q = 1$),
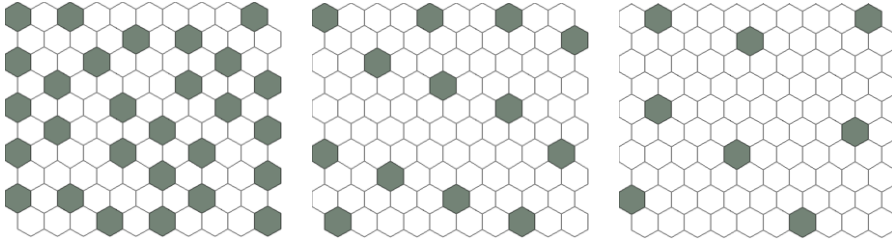
Figure 9. Three stable states for the model composed over a 10×10 grid, where each cell encompasses the lateral inhibition model (Figure 7), considering a cardinality constraint of *at-least-1* neighbour at distances 1, 2 and 3, respectively. Green cells are in state SS2, whereas white cells correspond to cells in SS1.

A) Logical graph

B) Logical functions

$$f_0 = x_0$$
$$f_1(x) = x_0$$
$$f_2(x) = x_1 \vee x_3$$
$$f_3(x) = x_1$$

C) Cellular stable states

u_0	g_1	g_2	g_3	
0	0	0	0	SS1
1	1	1	1	SS2

D) Integration function

$$u_0 \iff \geq_q (g_{1_1}, \ldots, g_{1_v})$$

E) Compatibility rules

$$\geq_q (SS2_1, \ldots, SS2_v) \implies SS2$$
$$\neg \geq_q (SS2_1, \ldots, SS2_v) \implies SS1$$

Figure 10. Lateral induction model: **A)** LRG with 3 internal components and 1 input component. **B)** Logical functions governing the internal components. **C)** Stable states of the single cell model. **D)** Integration function: the mapped input is 1 iff *at-least-q* neighbours have g_1 at level 1. **E)** Compatibility rules: if *at-least-q* of its neighbours is in state SS2, a cell is then in SS2.

all cells are in the same state (either SS1 or SS2), thus leading to an homogeneous stable pattern of the whole. This behaviour does not depend on the size of the grid, but it is disrupted whenever some cells, and not all, comply with the cardinality constraint. In Figure 11, we observe that an increase on the number of stable states with the size of the grid is only possible for intermediate values of q, *i.e.*, only when cells have less neighbours than required by the cardinality constraint, $q \geq v$, or when the cardinality constraint is above half the number of neighbours, $q > v/2$. The first cells to be in this interval are the cells at the borders of the grid (see Figure 5). In this case, the grid stable states are no longer homogeneous, and their numbers depend on the size of the grid. This shows that increasing the size of the grid or the neighbouring does not necessarily lead to an increase of the number of stable states.
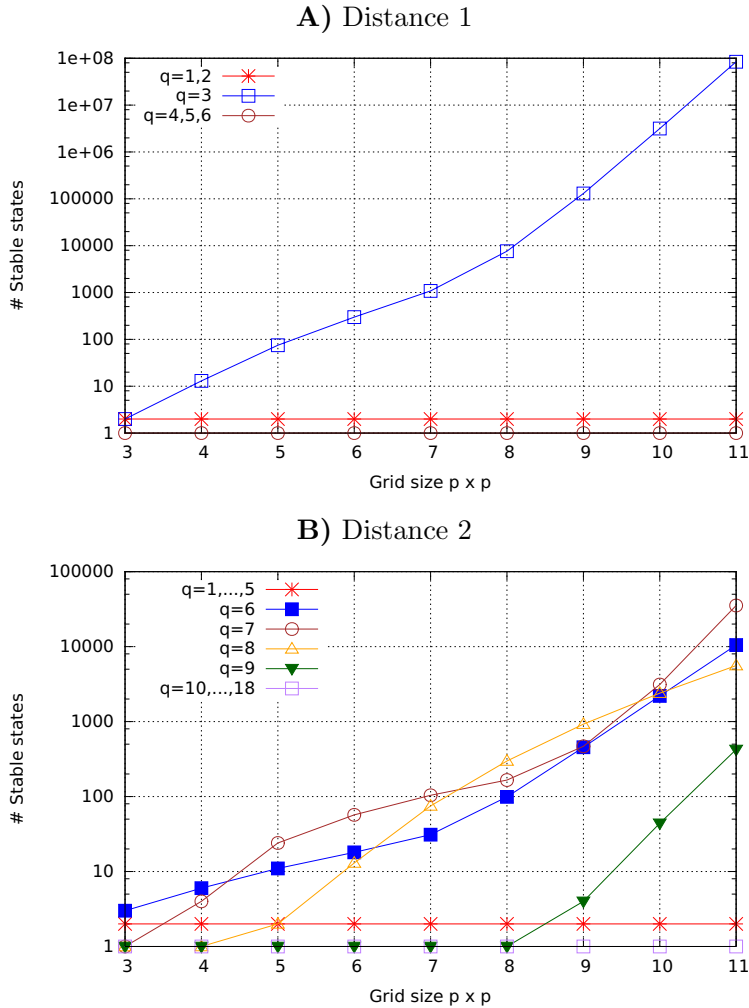
A) Distance 1

B) Distance 2

Figure 11. Lateral induction model composed over a $p \times p$ grid: evolution of the numbers of stable states varying p, the grid size, and for different values of q in the cardinality constraint. A) Neighbouring distance is fixed to 1 whereas in B) the distance is 2. Plots are on a logarithmic scale.

5   Discussion

The identification of attractors in Boolean networks has received great attention, albeit mainly focusing on the single-cell case (*i.e.*, cellular models). In this paper, we have addressed the identification of stable states of multi-cellular systems defined over 2D hexagonal grids. To this end, we proposed two SAT encodings. The first defines the tability condition of a composed model as an appropriate combination of stability conditions of the constituent cellular models. This encoding is similar to previous work as it results in searching stable states of a large model [18,14]. The second encoding uses compatibility rules to combine (previously determined) cellular stable states across neighbouring cells. For most models, this abstraction should yield a significant improvement in performance (Figure 4-A). We show that models composed over grids of cells may yield an exponential number of stable

states. For the case studies presented here, the number of stable states ended up being the limiting factor for the grid sizes considered, as the computational cost of counting all stable states surpassed the timeout (4 hours).

Moreover, we demonstrate that the evolution of the number of stable states depends on parameters such as q, the number of neighbours required to be in a specific state, the neighbouring distance as well as the grid size and the border conditions.How the number of stable states evolves with q greatly depends on the cellular model and on the composition rule (*e.g.* a state being possible if neighbouring cells are (or not) in the same state).

As future work, we aim to exploit the grid topology as well as intrinsic model characteristics, to use symmetry breaking and thus cut down the number of solutions [6,20]. This will allow the consideration of ever larger grids. Furthermore, there is room for generalisation of the framework presented here, as integration functions (and hence the compatibility rules) may depend on distinct internal components. Finally, we intend to integrate the SAT Compatibility condition encoding in EpiLog (`http://epilog-tool.org`), a software tool dedicated to the definition, simulation and visualisation of qualitative models of pattern formation on epithelia [28].

Acknowledgment

References

[1] I. Abío, R. Nieuwenhuis, A. Oliveras, and E. Rodriguéz-Carbonell. A parametric approach for smaller and better encodings of cardinality constraints. In *Principles and Practice of Constraint Programming – CP 2013*, volume 8124 of *Lecture Notes in Computer Science*, pages 80–96. Springer, 2013.

[2] W. Abou-Jaoudé, P. Traynard, P. T. Monteiro, J. Saez-Rodriguez, T. Helikar, D. Thieffry, and C. Chaouiya. Logical modeling and dynamical analysis of cellular networks. *Frontiers in Genetics*, 7:94, 2016.

[3] T. Akutsu, S. Kuhara, O. Maruyama, and S. Miyano. Identification of gene regulatory networks by strategic gene disruptions and gene overexpressions. In *Proc. 9th ACM-SIAM Symposium Discrete Algorithms – SODA 1998*, pages 695–702, 1998.

[4] R. Albert and H. G. Othmer. The topology of the regulatory interactions predicts the expression pattern of the segment polarity genes in *Drosophila melanogaster*. *Journal of Theoretical Biology*, 223(1):1–18, 2003.

[5] R. Albert and J. Thakar. Boolean modeling: a logic-based dynamic approach for understanding signaling and regulatory networks and for making useful predictions. *Wiley Interdisciplinary Reviews: Systems Biology and Medicine*, 6(5):353–69, 2014.

[6] F. A. Aloul, K. A. Sakallah, and I. L. Markov. Efficient symmetry breaking for Boolean satisfiability. *IEEE Transactions on Computers*, 55(5), 2006.

[7] D. Bérenguier, C. Chaouiya, P. Monteiro, A. Naldi, E. Remy, D. Thieffry, and L. Tichit. Dynamical modeling and analysis of large cellular regulatory networks. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 23(2):025114, 2013.

[8]   A. Biere, D. Le Berre, E. Lonca, and N. Manthey. Detecting cardinality constraints in CNF. In *Theory and Applications of Satisfiability Testing – SAT 2014*, volume 8561 of *Lecture Notes in Computer Science*, pages 285–301. Springer, 2014.

[9]   J. Chen. A new SAT encoding of the at-most-one constraint. In *Proc. International Workshop on Constraint Modelling and Reformulation (ModRef)*, 2010.

[10] A. B. Chitnis. The role of Notch in lateral inhibition and cell fate specification. *Molecular and Cellular Neuroscience*, 6(4):311–321, 1995.

[11] J. R. Collier, N. A. Monk, P. K. Maini, and J. H. Lewis. Pattern formation by lateral inhibition with feedback: a mathematical model of delta-notch intercellular signalling. *Journal of Theoretical Biology*, 183(4):429–446, 1996.

[12] V. Devloo, P. Hansen, and M. Labbé. Identification of all steady states in large networks by logical analysis. *Bulletin of Mathematical Biology*, 65(6):1025–51, Nov 2003.

[13] G. Didier, E. Remy, and C. Chaouiya. Mapping multivalued onto Boolean dynamics. *Journal of Theoretical Biology*, 270(1):177–84, 2011.

[14] E. Dubrova and M. Teslenko. A SAT-based algorithm for finding attractors in synchronous Boolean networks. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 8(5):1–7, 2011.

[15] A. Fauré, B. M. I. Vreede, E. Sucena, and C. Chaouiya. A discrete model of *Drosophila* eggshell patterning reveals cell-autonomous and juxtacrine effects. *PLoS Computational Biology*, 10(3):e1003527, 2014.

[16] A. M. Frisch and P. A. Giannaros. SAT encodings of the at-most-k constraint: Some old, some new, some fast, some slow. In *Proc. International Workshop on Constraint Modelling and Reformulation (ModRef)*, 2010.

[17] A. Garg, I. Xenarios, L. Mendoza, and G. DeMicheli. *An Efficient Method for Dynamic Analysis of Gene Regulatory Networks and in silico Gene Perturbation Experiments*, pages 62–76. Springer Berlin Heidelberg, Berlin, Heidelberg, 2007.

[18] K. Inoue. Logic programming for Boolean networks. In *22nd Intl. Joint Conference on Artificial Intelligence – IJCAI 2011*, pages 924–930. AAAI Press, 2011.

[19] W. Klieber and G. Kwon. Efficient CNF encoding for selecting 1 from N objects. In *Proc. Intl. Workshop on Constraints in Formal Verification*, 2007.

[20] I. Lynce and J. Marques-Silva. Breaking symmetries in SAT matrix models. In *Theory and Applications of Satisfiability Testing – SAT 2007*, volume 4501 of *Lecture Notes in Computer Science*, pages 22–27. Springer, 2007.

[21] N. D. Mendes, F. Lang, Y.-S. Le Cornec, R. Mateescu, G. Batt, and C. Chaouiya. Composition and abstraction of logical regulatory modules: application to multicellular systems. *Bioinformatics*, 29(6):749–57, 2013.

[22] A. Naldi, D. Berenguier, A. Fauré, F. Lopez, D. Thieffry, and C. Chaouiya. Logical modelling of regulatory networks with GINsim 2.3. *Biosystems*, 97(2):134–139, 2009.

[23] A. Naldi, D. Thieffry, and C. Chaouiya. Decision diagrams for the representation and analysis of logical models of genetic networks. In *Computational Methods in Systems Biology*, pages 233–247. Springer Berlin Heidelberg, 2007.

[24] V.-H. Nguyen and S. T. Mai. A new method to encode the at-most-one constraint into SAT. In *Proc. of the 6th Intl. Symposium on Information and Communication Technology – SoICT 2015*, pages 1–8. ACM Press, 2015.

[25] M. R. Owen, J. A. Sherratt, and H. J. Wearing. Lateral induction by juxtacrine signaling is a new mechanism for pattern formation. *Developmental Biology*, 217(1):54–61, 2000.

[26] L. Sánchez, C. Chaouiya, and D. Thieffry. Segmenting the fly embryo: logical analysis of the role of the segment polarity cross-regulatory module. *The International Journal of Developmental Biology*, 52(8):1059–75, 2008.

[27] T. Sang, F. Bacchus, P. Beame, H. A. Kautz, and T. Pitassi. Combining component caching and clause learning for effective model counting. In *Intl. Conf. on Theory and Applications of Satisfiability Testing – SAT 2004*, 2004.

[28] P. Varela, P. Monteiro, N. Mendes, A. Fauré, and C. Chaouiya. EpiLog, a computational tool for the logical modelling of epithelial pattern formation. In *European Conference on Computational Biology – ECCB 2014*, volume 5, page 1709. F1000Posters, 2014.