

---

# On Trace Languages Generated by Spiking Neural P Systems

Haiming Chen<sup>1</sup>, Mihai Ionescu<sup>2</sup>, Andrei Păun<sup>3</sup>, Gheorghe Păun<sup>4</sup>, Bianca Popa<sup>3</sup>

<sup>1</sup> Computer Science Laboratory, Institute of Software  
Chinese Academy of Sciences  
100080 Beijing, China  
[chm@ios.ac.cn](mailto:chm@ios.ac.cn)

<sup>2</sup> Research Group on Mathematical Linguistics  
Rovira i Virgili University  
Pl. Imperial Tàrraco 1, 43005 Tarragona, Spain  
[armandmihai.ionescu@urv.net](mailto:armandmihai.ionescu@urv.net)

<sup>3</sup> Department of Computer Science  
Louisiana Tech University, Ruston  
PO Box 10348, Louisiana, LA-71272 USA  
[apaun@latech.edu](mailto:apaun@latech.edu), [bdp010@latech.edu](mailto:bdp010@latech.edu)

<sup>4</sup> Institute of Mathematics of the Romanian Academy  
PO Box 1-764, 014700 Bucharest, Romania, and  
Research Group on Natural Computing  
Department of Computer Science and AI  
University of Sevilla  
Avda Reina Mercedes s/n, 41012 Sevilla, Spain  
[george.paun@imar.ro](mailto:george.paun@imar.ro), [gpaun@us.es](mailto:gpaun@us.es)

**Summary.** We extend to spiking neural P systems a notion investigated in the “standard” membrane systems: the language of the traces of a distinguished object. In our case, we distinguish a spike by “marking” it and we follow its path through the neurons of the system, thus obtaining a language. Several examples are discussed and some preliminary results about this way of associating a language with a spiking neural P system are given, together with a series of topics for further research. For instance, we show that each regular language is the morphic image of a trace language intersected with a very particular regular language, while each recursively enumerable language over the one-letter alphabet is the projection of a trace language.

## 1 Introduction

We continue here the study of the spiking neural P systems (in short, SN P systems) recently introduced in [7], by considering in this framework the idea of following the traces of an object through the system in order to get a language.

For usual P systems (with symport/antiport rules), this way of generating a language was introduced in [6] and then investigated in a series of papers (see, e.g., [3] and Section 4.4 of [9]).

In short, an SN P system consists of a set of neurons placed in the nodes of a graph and sending signals (spikes) along synapses (edges of the graph), under the control of firing rules. One also uses forgetting rules, which remove spikes from neurons. Therefore, the spikes are moved, created, destroyed, but never modified (there is only one type of objects in the system). This makes possible and natural to distinguish one of the spikes present already in the initial configuration of the system and to follow its path through the neurons, recording the labels of the neurons and thus obtaining a string. More precisely, we assume that one of the spikes has a “flag”, which remains with this spike as long as the spike is not consumed or forgotten, but the flag passes to the produced spike when the old holder is consumed; the flag disappears when the spike is forgotten or sent out of the system. More precise definitions will be given in Section 3. We only add here the fact that we consider only halting computations, so that the labels of neurons visited by the flag form a string; due to the non-determinism in the SN P system functioning, we get in this way a set of strings, hence a language, associated with the system.

Because the SN P systems are Turing complete as number computing devices (see [7] and [10]) and also complete modulo direct and inverse morphisms in the case when they are used as language generators (see [1]), it is expected that the SN P systems are powerful also as trace languages generators. This expectation is somewhat confirmed below: we both produce a series of examples of systems with an intricate functioning, we give a representation of regular languages (over any alphabet), and we also show that unary recursively enumerable languages have simple representations in terms of SN P systems trace languages. On the other hand, because of the specific way the trace language is generated, there are finite (even singleton) languages which cannot be obtained in this way. However, due to the preliminary stage of our research, we do not have yet a precise estimation of the size of the generated families of languages; many problems in this respect are formulated along the paper.

In all these investigations we try to keep the used system as simple as possible, with respect to the many descriptive complexity measures which can be considered for SN P systems: number of neurons, number of rules in each neuron, number of consumed or forgotten spikes in each rule, etc.

## 2 Formal Language Theory Prerequisites

We assume the reader to be familiar with basic language and automata theory, e.g., from [11] and [12], so that we introduce here only some notations and notions used later in the paper.

For an alphabet  $V$ ,  $V^*$  denotes the set of all finite strings of symbols from  $V$ ; the empty string is denoted by  $\lambda$ , and the set of all nonempty strings over  $V$  is

denoted by  $V^+$ . When  $V = \{a\}$  is a singleton, then we write simply  $a^*$  and  $a^+$  instead of  $\{a\}^*$ ,  $\{a\}^+$ .

A morphism  $h : V_1^* \rightarrow V_1^*$  such that  $h(a) \in \{a, \lambda\}$  for each  $a \in V_1$  is called projection, and a morphism  $h : V_1^* \rightarrow V_2^*$  such that  $h(a) \in V_2 \cup \{\lambda\}$  for each  $a \in V_1$  is called a weak coding; it is a coding if  $h(a) \in V_2$  for all  $a \in V_1$ .

If  $L_1, L_2 \subseteq V^*$  are two languages, the left and right quotients of  $L_1$  with respect to  $L_2$  are defined by  $L_2 \backslash L_1 = \{w \in V^* \mid xw \in L_1 \text{ for some } x \in L_2\}$ , and respectively  $L_1 / L_2 = \{w \in V^* \mid wx \in L_1 \text{ for some } x \in L_2\}$ . When  $L_2$  is a singleton,  $L_2 = \{x\}$ , we write  $\partial_x^l(L_1)$  for  $\{x\} \backslash L_1$  and  $\partial_x^r(L_1)$  for  $L_1 / \{x\}$ , and we call these operations left and right derivatives of  $L_1$  with respect to  $x$ .

A Chomsky grammar is given in the form  $G = (N, T, S, P)$ , where  $N$  is the nonterminal alphabet,  $T$  is the terminal alphabet,  $S \in N$  is the axiom, and  $P$  is the finite set of rules. For regular grammars, the rules are of the form  $A \rightarrow aB, A \rightarrow a$ , for some  $A, B \in N, a \in T$ .

We ignore here the empty string and this **convention** is assumed in the paper always when examining the generative power of any device.

We denote by *FIN*, *REG*, *CF*, *CS*, *RE* the families of finite, regular, context-free, context-sensitive, and recursively enumerable languages; by *MAT* we denote the family of languages generated by matrix grammars without appearance checking ([11], [2]).

### 3 Trace Languages Associated with Spiking Neural P Systems

We introduce first the SN P systems in their basic form, then we define the new way of using them, as (trace) language generators; for further details we refer to [7], [10].

A *spiking neural P system* (abbreviated as SN P system), of degree  $m \geq 1$ , is a construct of the form

$$\Pi = (O, \sigma_1, \dots, \sigma_m, \text{syn}),$$

where:

1.  $O = \{a\}$  is the singleton alphabet ( $a$  is called *spike*);
2.  $\sigma_1, \dots, \sigma_m$  are *neurons*, of the form

$$\sigma_i = (n_i, R_i), 1 \leq i \leq m,$$

where:

- a)  $n_i \geq 0$  is the *initial number of spikes* contained in  $\sigma_i$ ;
- b)  $R_i$  is a finite set of *rules* of the following two forms:
  - (1)  $E/a^c \rightarrow a; d$ , where  $E$  is a regular expression over  $a$ ,  $c \geq 1$ , and  $d \geq 0$ ;
  - (2)  $a^s \rightarrow \lambda$ , for some  $s \geq 1$ , with the restriction that for each rule  $E/a^c \rightarrow a; d$  of type (1) from  $R_i$ , we have  $a^s \notin L(E)$ ;

3.  $syn \subseteq \{1, 2, \dots, m\} \times \{0, 1, 2, \dots, m\}$  with  $i \neq j$  for each  $(i, j) \in syn$ ,  $1 \leq i, j \leq m$  (*synapses* between neurons), with 0 indicating the environment of the system.

Note that, in contrast to [7], we have not indicated here an output neuron, but we have allowed synapses  $(i, 0)$  for any neuron  $\sigma_i$ ; actually, in the trace case investigated here we do not really need such synapses, but we consider them just because such links with the environment are “realistic”.

The rules of type (1) are *firing* (we also say *spiking*) rules, and they are applied as follows. If the neuron  $\sigma_i$  contains  $k$  spikes, and  $a^k \in L(E)$ ,  $k \geq c$ , then the rule  $E/a^c \rightarrow a; d$  can be applied. The application of this rule means consuming (removing)  $c$  spikes (thus only  $k - c$  remain in  $\sigma_i$ ), the neuron is fired, and it produces a spike after  $d$  time units (a global clock is assumed, marking the time for the whole system, hence the functioning of the system is synchronized). If  $d = 0$ , then the spike is emitted immediately, if  $d = 1$ , then the spike is emitted in the next step, etc. If the rule is used in step  $t$  and  $d \geq 1$ , then in steps  $t, t+1, t+2, \dots, t+d-1$  the neuron is *closed* (this corresponds to the refractory period from neurobiology), so that it cannot receive new spikes (if a neuron has a synapse to a closed neuron and tries to send a spike along it, then that particular spike is lost). In the step  $t + d$ , the neuron spikes and becomes again open, so that it can receive spikes (which can be used starting with the step  $t + d + 1$ ), but it does not use any rule at this step (the neuron is busy with sending out the spike it has produced  $d$  steps before and stored up to now inside). A spike emitted by a neuron  $\sigma_i$  (is replicated and a copy of it) goes to each neuron  $\sigma_j$  such that  $(i, j) \in syn$ , as well as to the environment, if  $(i, 0) \in syn$ . (When a neuron  $\sigma_i$  emits a spike and there is no neuron to receive it, then a synapse  $(i, 0)$  is used as an explicit way to point out that the spike is “lost” in the environment.)

The rules of type (2) are *forgetting* rules and they are applied as follows: if the neuron  $\sigma_i$  contains exactly  $s$  spikes, then the rule  $a^s \rightarrow \lambda$  from  $R_i$  can be used, meaning that all  $s$  spikes are removed from  $\sigma_i$ .

If a rule  $E/a^c \rightarrow a; d$  of type (1) has  $E = a^c$ , then we will write it in the simplified form  $a^c \rightarrow a; d$ .

In each time unit, if a neuron  $\sigma_i$  can use one of its rules, then a rule from  $R_i$  must be used. Since two firing rules,  $E_1/a^{c_1} \rightarrow a; d_1$  and  $E_2/a^{c_2} \rightarrow a; d_2$ , can have  $L(E_1) \cap L(E_2) \neq \emptyset$ , it is possible that two or more rules can be applied in a neuron, and in that case, only one of them is chosen non-deterministically. By definition, if a firing rule is applicable, then no forgetting rule is applicable, and vice versa.

The initial configuration of the system is described by the numbers  $n_1, n_2, \dots, n_m$ , of spikes present in each neuron, with all neurons being open. During the computation, a configuration is described by both the number of spikes present in each neuron and by the state of the neuron, more precisely, by the number of steps from now on until it becomes open (this number is zero if the neuron is already open). Thus,  $\langle r_1/t_1, \dots, r_m/t_m \rangle$  is the configuration where neuron  $i = 1, 2, \dots, m$  contains  $r_i \geq 0$  spikes and it will be open after  $t_i \geq 0$  steps; with this notation, the initial configuration is  $C_0 = \langle n_1/0, \dots, n_m/0 \rangle$ .

Using the rules as described above, one can define transitions among configurations. A transition between two configurations  $C_1, C_2$  is denoted by  $C_1 \Longrightarrow C_2$ . Any sequence of transitions starting in the initial configuration is called a *computation*. A computation halts if it reaches a configuration where all neurons are open and no rule can be used.

In the spirit of spiking neurons, see, e.g., [8], as the result of a computation, in [7] and [10] one considers the distance between two consecutive spikes which exit a distinguished neuron of the system. Then, in [1] one considers as the result of a computation the so-called spike train of the computation, the sequence of symbols 0 and 1 obtained by associating 1 with a step when a spike exits the system and 0 otherwise. Languages over the binary alphabet are computed in this way.

Here we consider yet another idea for defining a language, taking into account the traces of a distinguished spike through the system.

Specifically, we distinguish one of the neurons of the system as the *input* one (thus, we add a further component,  $in$ , to the system description, with  $in \in \{1, 2, \dots, m\}$ ) and in the initial configuration of the system we “mark” one spike from this neuron – the intuition is that this spike has a “flag” – and we follow the path of this flag during the computation, *recording the labels of the neurons where the flag is present in the end of each step*. Actually, for neuron  $\sigma_i$  we consider the symbol  $b_i$  in the trace string. (When presenting the initial configuration of the system, the number  $n_{in}$ , of spikes present in the input neuron, is written in the form  $n'_{in}$ , to indicate that the marked spike is here.)

The previous definition contains many delicate points which need clarifications – and we use a simple example to do this.

Assume that in neuron  $\sigma_i$  we have three spikes, one of them marked; we write  $aaa'$  to represent them. Assume also that we have a spiking rule  $aaa/aa \rightarrow a;0$ . When applied, this rule consumes two spikes, one remains in the neuron and one spike is produced and sent along the synapses going out of neuron  $\sigma_i$ . Two cases already appear: the marked spike is consumed or not. If not consumed, it remains in the neuron. If consumed, then the flag passes to the produced spike. Now, if there are two or more synapses going out of neuron  $\sigma_i$ , then again we can have a branching: only one spike is marked, hence only on one of the synapses  $(i, j)$  we will have a marked spike and that synapse is non-deterministically chosen (on other synapses we send non-marked spikes). If  $\sigma_j$  is an open neuron, then the marked spike ends in this neuron. If  $\sigma_j$  is a closed neuron, then the marked spike is lost, and the same happens if  $j = 0$  (the marked spike exits in the environment). Anyway, if the marked spike is consumed, at the end of this step it is no longer present in neuron  $i$ ; it is in neuron  $\sigma_j$  if  $(i, j) \in syn$  and neuron  $\sigma_j$  is open, or it is removed from the system in other cases.

Therefore, if in the initial configuration of the system neuron  $\sigma_i$  contains the marked spike, then the trace can start either with  $b_i$  (if the marked spike is not consumed) or with  $b_j$  (if the marked spike was consumed and passed to neuron  $\sigma_j$ ); if the marked spike is consumed and lost, then we generate the empty string, which is ignored in our considerations. Similarly in later steps.

If the rule used is of the form  $aaa/aa \rightarrow a; d$ , for some  $d \geq 1$ , and the marked spike is consumed, then the newly marked spike remains in neuron  $\sigma_i$  for  $d$  steps, hence the trace starts/continues with  $b_i^d$ . Similarly, if no rule is used in neuron  $\sigma_i$  for  $k$  steps, then the trace records  $k$  copies of  $b_i$ .

If a forgetting rule is used in the neuron where the marked spike is placed, then the trace string stops (and no symbol is recorded for this step).

Therefore, when considering the possible branchings of the computation, we have to take into account the non-determinism not only in using the spiking rules, but also in consuming the marked spike and in sending it along one of the possible synapses.

The previous discussion has, hopefully, made clear what we mean by *recording the labels of the neurons where the flag is present in the end of each step*, and why we have chosen the end of a step and not the beginning: in the latter case, all traces would start with the same symbol, corresponding to the input neuron, which is a strong – and artificial – restriction.

In the next section, we will illustrate all these points by a series of examples.

Anyway, we take into account only halting computations: irrespective whether or not a marked spike is still present in the system, the computation should halt (note that it is possible that the marked spike is removed and the computation still continues for a while – but this time without adding further symbols to the trace string).

For an SN P system  $\Pi$  we denote by  $T(\Pi)$  the language of all strings describing the traces of the marked spike in all halting computations of  $\Pi$ . Then, we denote by  $TSNP_m(rule_k, cons_p, forg_q)$  the family of languages  $T(\Pi)$ , generated by systems  $\Pi$  with at most  $m$  neurons, each neuron having at most  $k$  rules, each of the spiking rules consuming at most  $p$  spikes, and each forgetting rule removing at most  $q$  spikes. As usual, a parameter  $m, k, p, q$  is replaced with  $*$  if it is not bounded.

## 4 Examples

We consider here several examples, both illuminating the previous definitions and relevant for the intricate work of SN P systems as language generators by means of traces; indications on the power of these devices are also obtained in this way.

We start with a system already having a complex behavior, the one whose initial configuration is given in Figure 1 (we denote it by  $\Pi_1$ ). This also gives us the opportunity to introduce the way to graphically represent an SN P system: as a directed graph, with the neurons as nodes and the synapses indicated by arrows; in each neuron we give the rules and the spikes present in the initial configuration, with the marked spike indicated by a prime; synapses of the form  $(i, 0)$  are indicated by arrows pointing to the environment.

We have three neurons, labeled with 1, 2, 3, with neuron  $\sigma_1$  being the input one. Each neuron contains three rules, and only neuron  $\sigma_1$  has two spiking rules, but the non-determinism of the system is considerable, due to the possible traces of the marked spike.

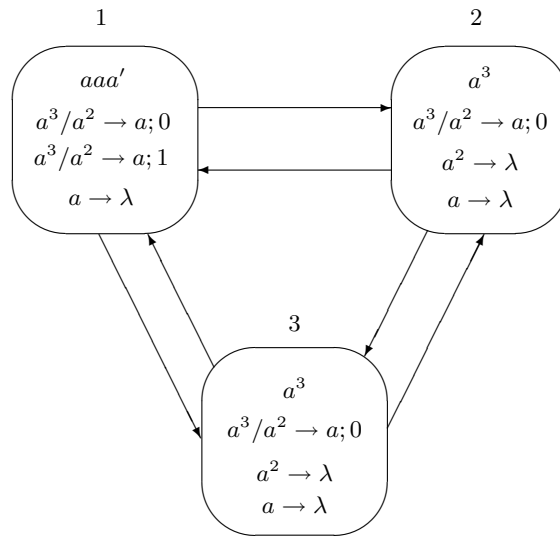


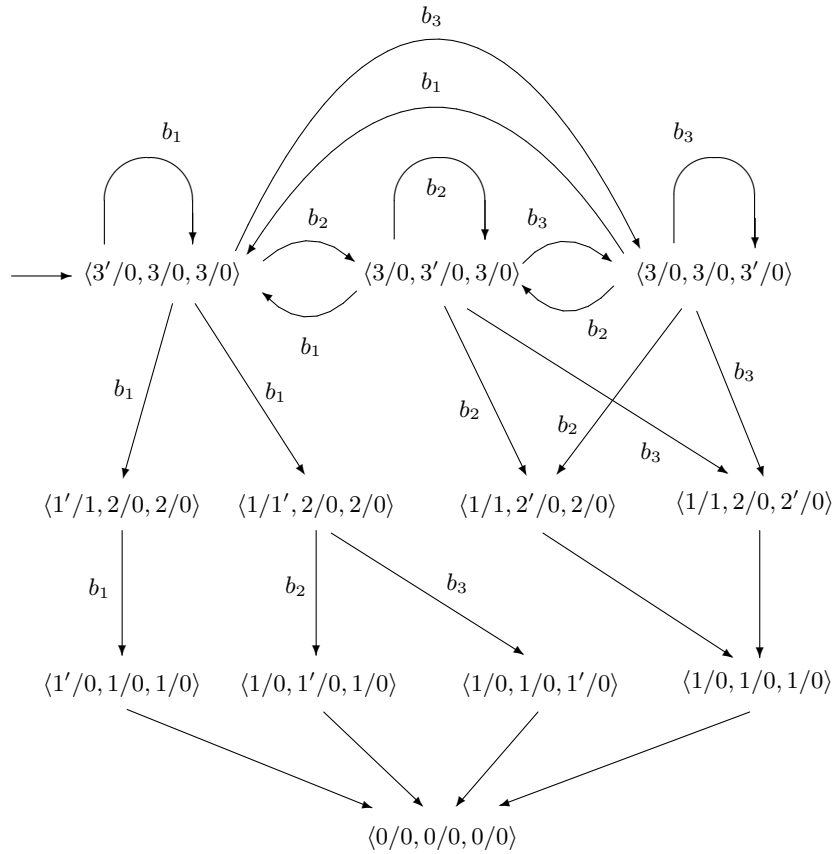
Fig. 1. The initial configuration of system  $II_1$

The evolution of the system  $II_1$  can be analyzed on a transition diagram as that from Figure 2: because the number of configurations reachable from the initial configuration is finite, we can place them in the nodes of a graph, and between two nodes/configurations we draw an arrow if a direct transition is possible between them.

It should be noted in Figure 2 an important detail: when presenting a configuration of the system where there is a marked spike, it is no longer sufficient to indicate only the number of spikes and the open status of neurons, but we also have to indicate the place of the marked spike (if the system still contains such a spike). This is done by priming either the number of spikes from the neuron where the marked spike is, or by priming the number of steps until the neuron is open, in the case when a marked spike was produced by means of a spiking rule with delay.

In Figure 2 we have also indicated on the arrows the symbol introduced in the trace string by that arrow (this is  $b_j$ , where  $\sigma_j$  is the neuron where the marked spike arrives in the end of this step). Thus, following the marked arrows, we can construct the language of all traces.

Let us follow on this diagram some of the possible traces. As long as neuron  $\sigma_1$  uses the rule  $a^3/a^2 \rightarrow a; 0$ , the marked spike circulates among neurons  $\sigma_1, \sigma_2, \sigma_3$  and the computation continues. Note that the marked spike can be consumed or not; in the first case it moves to one of the partner neurons, non-deterministically chosen, in the latter case it remains in the neuron where it is placed.



**Fig. 2.** The transition diagram of system  $\Pi_1$

When neuron  $\sigma_1$  uses the rule  $a^3/a^2 \rightarrow a; 1$ , then the computation passes to the halting phase – in the diagram from Figure 2, we leave the upper level and we pass to the next level of configurations. If the marked spike was in neuron  $\sigma_1$ , it is or not consumed; this is the case when reaching the configurations  $\langle 1'/1, 2/0, 2/0 \rangle$ ,  $\langle 1/1', 2/0, 2/0 \rangle$ : all neurons consume two spikes, but neurons  $\sigma_2$  and  $\sigma_3$  exchange one spike, hence they end the step with two spikes inside; neuron  $\sigma_1$  has only one spike inside (it is closed, does not accept spikes from neurons  $\sigma_2$  and  $\sigma_3$ ) and one ready to be emitted, and either one of them can be the marked one.

In each case, after two more steps the computation halts with no spike in the system.

Thus, the traces start with an arbitrary string  $x$  over  $\{b_1, b_2, b_3\}$ , and end with one of the strings  $b_1b_1, b_1b_2, b_1b_3, b_2, b_3$ , depending on the last symbol of  $x$ .



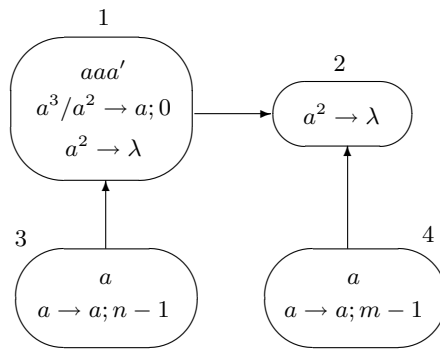


Fig. 3. The initial configuration of system  $\Pi_2$

We continue with two simpler examples, not devoid of interest: in search for counterexamples, we have tried several finite languages, and for many of them the result was that in all cases it was possible to find an SN P system to generate these languages. We continue to present these systems in a graphical form instead of giving them formally, because the figures are much easier to understand.

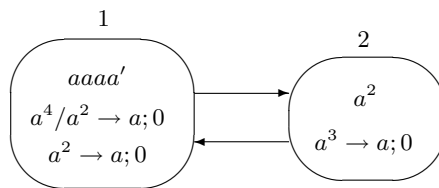


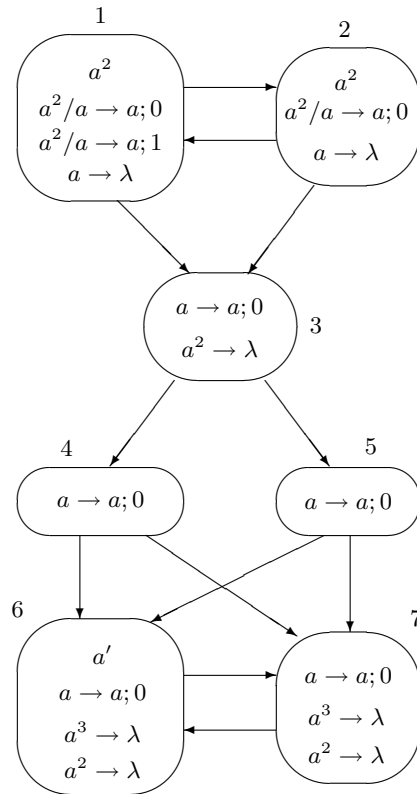
Fig. 4. The initial configuration of system  $\Pi_3$

The two examples are given in Figures 3, 4, and the respective systems are denoted by  $\Pi_2, \Pi_3$ , respectively.

The system  $\Pi_2$  from Figure 3 generates the language  $T(\Pi_2) = \{b_1^n, b_2^m\}$ , for  $n, m \geq 1$ . In the first step, neuron  $\sigma_1$  consumes or does not consume the marked spike, thus keeping it inside or sending it to neuron  $\sigma_2$ . One spike remains in neuron  $\sigma_1$  and one is placed in neuron  $\sigma_2$ . Simultaneously, neurons  $\sigma_3$  and  $\sigma_4$  fire, and they spike after  $n - 1$  and  $m - 1$  steps, respectively. Thus, in steps  $n$  and  $m$ , neurons  $\sigma_1$  and  $\sigma_2$ , respectively, receive one more spike, which is forgotten in the next step together with the spike existing there.

Note that  $n$  and  $m$  can be equal or different.

For the system in Figure 4 we have  $T(\Pi_3) = \{b_1 b_2, b_2 b_1\}$  – we leave the task to check this equality to the reader.



**Fig. 5.** The initial configuration of system  $\Pi_4$

We end this section with a slightly more complex system,  $\Pi_4$ , whose initial configuration is given in Figure 5. We have  $T(\Pi_4) = (b_7b_6)^+b_7 \cup (b_7b_6)^+b_7b_6$ .

The neurons  $\sigma_6, \sigma_7$  exchange the marked spike among them as long as they do not get “flooded” by neurons  $\sigma_4, \sigma_5$ , when a spike comes from neuron  $\sigma_3$ . In turn, this neuron spikes only in the step when neuron  $\sigma_1$  uses the rule  $a^2/a \rightarrow a; 1$  (if neuron  $\sigma_1$  uses the rule  $a^2/a \rightarrow a; 0$ , then the cooperation of neurons  $\sigma_1, \sigma_2$  continues, both of them returning to the initial state by exchanging spikes). When neurons  $\sigma_6, \sigma_7$  have inside two or three spikes, they forget them and the computation stops. Depending on the step when neuron  $\sigma_1$  uses the rule  $a^2/a \rightarrow a; 1$  (an odd or even step), the string ends with  $b_7$  or with  $b_7b_6$ .

## 5 The Generative Power of SN P Systems as Trace Language Generators

The following inclusions are direct consequences of the definitions:

**Lemma 1.**  $TSNP_m(rule_k, cons_p, forg_q) \subseteq TSNP_{m'}(rule_{k'}, cons_{p'}, forg_{q'}) \subseteq TSNP_*(rule_*, cons_*, forg_*) \subseteq RE$ , for all  $1 \leq m \leq m'$ ,  $1 \leq k \leq k'$ ,  $1 \leq p \leq p'$ ,  $1 \leq q \leq q'$ .

We pass now to investigating the relationship with the families of languages from Chomsky hierarchy, starting with a counterexample result.

**Theorem 1.** *There are singleton languages which are not in  $TSNP_*(rule_*, cons_*, forg_*)$ .*

*Proof.* Let us consider the language  $L = \{b_1b_2b_1b_3\}$ . In order to generate it, at least three neurons should be used, with labels 1, 2, 3. Moreover, the synapses (1, 2), (2, 1), (1, 3) are necessary. The existence of the synapses (1, 2), (2, 1) makes possible also the trace  $b_1b_2b_1b_2$ : the marked spike exists after the third step and it can go non-deterministically to each neuron  $\sigma_2$  and  $\sigma_3$ . Thus,  $T(\Pi)$  cannot be a singleton, for an SN P system  $\Pi$  such that  $b_1b_2b_1b_3 \in T(\Pi)$ .  $\square$

Clearly, this reasoning can be applied to any string of the form  $w = w_1b_ib_jw_2b_ib_kw_3$  with  $j \neq k$ , and to any language which contains a string like  $w$  but not also the string  $w_1b_ib_jw_2b_ib_jw_3$ .

Let us mention now a result already proved by the considerations related to the first example from the previous section (this introduces another complexity parameter, of a dynamical nature: the number of spikes present in the neurons during a computation).

**Theorem 2.** *The family of trace languages generated by SN P systems by means of computations with a bounded number of spikes present in their neurons is strictly included in the family of regular languages.*

*Proof.* The inclusion follows from the fact that the transition diagram associated with the computations of an SN P system which use a bounded number of spikes is finite and can be interpreted as the transition diagram of a finite automaton, as already done in Section 4. The fact that the inclusion is proper is a consequence of Theorem 1.  $\square$

However, modulo some simple squeezing operations, SN P systems with a bounded number of spikes inside can generate any regular language:

**Theorem 3.** *For each regular language  $L \subseteq V^*$  there is an SN P system  $\Pi$  such that each neuron from any computation of  $\Pi$  contains a bounded number of spikes, and  $L = h(\partial_c^r(T(\Pi)))$  for some coding  $h$ , and symbol  $c$  not in  $V$ ; actually,  $T(\Pi) \in TSNP_{m_L}(rule_{k_L}, cons_{p_L}, forg_0)$ , for some constants  $m_L, k_L, p_L$  depending on language  $L$ .*

*Proof.* Let us assume that  $V = \{b_1, b_2, \dots, b_n\}$ , and take a regular grammar  $G = (N, V, S, P)$  generating the language  $L$ . Without loss of the generality, we may assume that each rule  $A \rightarrow b_i B$  from  $P$  has  $A \neq B$ . If this is not the case with the initial set of rules, then we proceed as follows. Take a rule  $A \rightarrow b_i A$ ; we consider a new nonterminal,  $A'$ , and we replace the rule  $A \rightarrow b_i A$  with  $A \rightarrow b_i A'$ , then we also add the rule  $A' \rightarrow b_i A$  as well as all rules  $A' \rightarrow b_j B$  for each  $A \rightarrow b_j B \in P$ . Let us continue to denote by  $G$  the obtained grammar. It is clear that this change does not modify the language generated by  $G$ , but it diminishes by one the number of rules having the same nonterminal in both sides. Continuing in this way, we eliminate all rules of this form.

Assume that  $G$  contains  $k$  rules of the form  $A \rightarrow b_i B$  (from the previous discussion, we know that  $A \neq B$ ). We construct the SN P system  $\Pi$  as follows.

The set of neurons is the following:

1.  $\sigma_{\langle S \rangle} = (1', \{a \rightarrow a; 0\})$ ,
2.  $\sigma_{\langle i, B \rangle} = (0, \{a^j \rightarrow a; 0 \mid 1 \leq j \leq k-1\})$ , for each rule  $A \rightarrow b_i B$  of  $G$ ,
3.  $\sigma_{\langle i \rangle} = (0, \{a^j \rightarrow a; 0 \mid 1 \leq j \leq k-1\})$ , for each rule  $A \rightarrow b_i$  of  $G$ ,
4.  $\sigma_f = (0, \{a^j \rightarrow a; 0 \mid 1 \leq j \leq k-1\})$ ,
5.  $\sigma_{c_i} = (0, \{a \rightarrow a; 0\})$ , for  $i = 1, 2, \dots, k$ ,
6.  $\sigma_1 = (2, \{a^2/a \rightarrow a; 0, a^2/a \rightarrow a; 1\})$ ,
7.  $\sigma_2 = (2, \{a^2/a \rightarrow a; 0, a \rightarrow \lambda\})$ ,
8.  $\sigma_3 = (0, \{a \rightarrow a; 0, a^2 \rightarrow \lambda\})$ .

Among these neurons, we take the following synapses:

$$\begin{aligned} \text{syn} = & \{(\langle S \rangle, \langle i, A \rangle) \mid S \rightarrow b_i A \in P\} \\ & \cup \{(\langle i, A \rangle, \langle j, B \rangle) \mid A \rightarrow b_j B \in P, 1 \leq i \leq n\} \\ & \cup \{(\langle i, A \rangle, \langle j \rangle) \mid A \rightarrow b_j \in P, 1 \leq i \leq n\} \\ & \cup \{(\langle S \rangle, \langle j \rangle) \mid S \rightarrow b_j \in P\} \\ & \cup \{(\langle i \rangle, f) \mid 1 \leq i \leq n\} \\ & \cup \{(1, 2), (2, 1), (1, 3), (2, 3)\} \cup \{(3, c_i) \mid 1 \leq i \leq k\} \\ & \cup \{(c_i, \langle j, A \rangle), (c_i, \langle j \rangle) \mid 1 \leq i \leq k, 1 \leq j \leq n, A \in N\}. \end{aligned}$$

We start with a single spike in the system, marked, placed in neuron  $\sigma_{\langle S \rangle}$ ; the input neuron spikes in the first step.

The neurons  $\sigma_1, \sigma_2, \sigma_3, \sigma_f, \sigma_{c_1}, \dots, \sigma_{c_k}$  will be discussed separately below.

Among neurons with labels of the form  $\langle S \rangle, \langle i, A \rangle$ , and  $\langle j \rangle$ , we have synapses only if the labels of these neurons are linked by a rule of the grammar. If there are several rules of the form  $A \rightarrow b_{i_1} B_1, \dots, A \rightarrow b_{i_r} B_r$ , then the spike emitted by each neuron  $\sigma_{\langle j, A \rangle}, 1 \leq j \leq n$ , goes to all neurons  $\sigma_{\langle i_t, B_t \rangle}, 1 \leq t \leq r$  – and the marked spike can take any of these choices. Conversely, we can have several (actually, at most  $k-1$ ) synapses coming to the same neuron, if more rules have the same right hand member. Thus, in any neuron we can collect at most  $k-1$  spikes in a step. All of them are immediately consumed, hence never the number

of spikes from any neuron becomes greater than  $k - 1$ . Clearly, the marked spike can follow a derivation in  $G$ .

Because the grammar  $G$  can contain cycles (for instance, pairs of rules of the form  $A \rightarrow b_i B, B \rightarrow b_i A$ ), the system  $\Pi$  can contain pairs of self-sustaining neurons, hence the computation in  $\Pi$  could not stop, even if a derivation in  $G$  was correctly simulated (this is due to the fact that the spikes leaving a neuron are replicated to all neurons to which we have a synapse).

In order to prevent such “wrong” evolutions, we use the “halting module”, composed of the neurons  $\sigma_1, \sigma_2, \sigma_3, \sigma_f, \sigma_{c_1}, \dots, \sigma_{c_k}$ : neurons  $\sigma_1, \sigma_2$  exchange spikes for an arbitrary number of steps; when  $\sigma_1$  uses the rule  $a^2/a \rightarrow a; 1$ , their interplay stops, but neuron  $\sigma_3$  fires and sends a spike to all neurons  $\sigma_{c_i}, 1 \leq i \leq k$ . These neurons will send spikes to all neurons  $\sigma_{\langle j, A \rangle}, \sigma_{\langle j \rangle}$ , and  $\sigma_{\langle S \rangle}$  of the system, thus halting their evolution (they do not have rules for handling more than  $k - 1$  spikes). The computation halts.

What is not ensured yet by the construction of  $\Pi$  is the fact that the computation is not halted as above before finishing the derivation in  $G$  which we want to simulate, that is, with the marked spike present inside the system. This aspect is handled by the squeezing mechanism: we take the alphabet

$$U = \{b_{\langle i, A \rangle} \mid 1 \leq i \leq n, A \in N\} \cup \{b_{\langle i \rangle} \mid 1 \leq i \leq n\},$$

the symbol  $c = b_f$ , and the coding  $h : U^* \rightarrow V^*$  defined by

$$h(b_{\langle i, A \rangle}) = b_i, \quad 1 \leq i \leq n, A \in N, \\ h(b_{\langle i \rangle}) = b_i, \quad 1 \leq i \leq n.$$

The right derivative with respect to  $b_f$  selects from  $T(\Pi)$  only those traces for which the marked spike reaches the “final” neuron  $\sigma_f$ , which is accessible only from a “terminal” neuron  $\sigma_{\langle i \rangle}$ , hence the marked spike has followed a complete derivation in  $G$ . Then, the coding  $h$  renames the symbols, thus delivering exactly the strings of  $L(G)$ .

Clearly, we can determine the constants  $m_L, k_L, p_L$  as in the statement of the theorem, depending on the size of grammar  $G$ , and this observation completes the proof.  $\square$

A related result can be obtained by slightly changing the previous proof.

**Theorem 4.** *For each regular language  $L \subseteq V^*$  there is an SN P system  $\Pi$  such that each neuron from any computation of  $\Pi$  contains a bounded number of spikes, and  $L = h(T(\Pi) \cap U_1^* U_2)$ , for some coding  $h$  and alphabets  $U_1, U_2$ .*

*Proof.* We just repeat the construction above, removing neuron  $\sigma_f$  and all synapses involving it. Then, we take the alphabets

$$U_1 = \{b_{\langle i, A \rangle} \mid 1 \leq i \leq n, A \in N\}, \\ U_2 = \{b_{\langle i \rangle} \mid 1 \leq i \leq n\},$$

and the same coding  $h : (U_1 \cup U_2)^* \rightarrow V^*$  as above

Now, the intersection with  $U_1^*U_2$  selects from  $T(\Pi)$  only those traces for which the marked spike reaches a “terminal” neuron  $\sigma_{\langle i \rangle}$ , hence it follows a complete derivation in  $G$ .  $\square$

As expected, also non-regular languages can be generated – of course, by using systems whose computations are allowed to use an unbounded number of spikes in the neurons. We start by considering a particular example, the one of the system  $\Pi_5$  given in Figure 6.

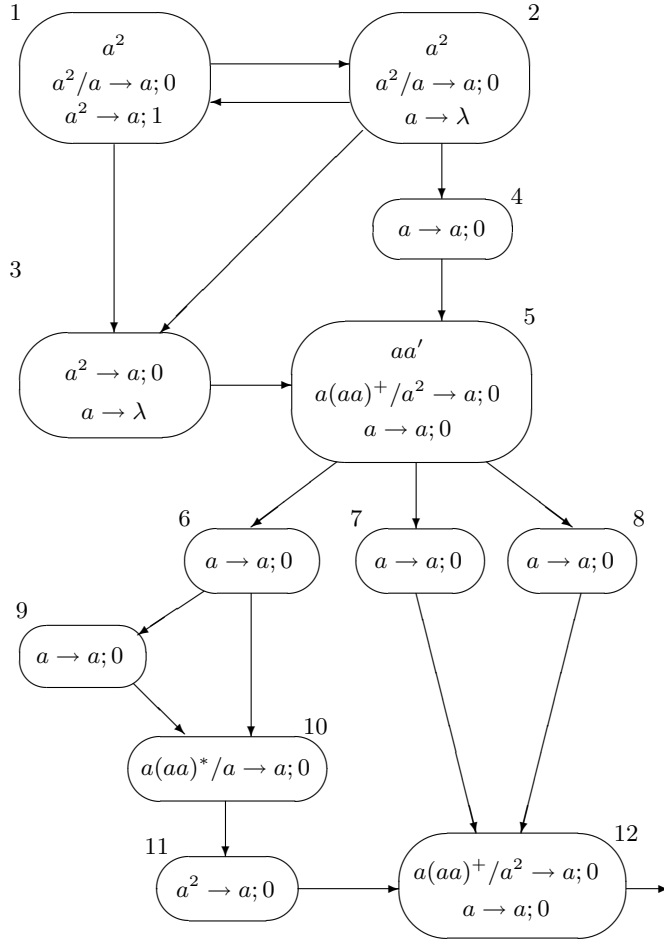


Fig. 6. An SN P system whose trace language is not regular

The neurons  $\sigma_1$  and  $\sigma_2$  play the same role as neurons  $\sigma_1$  and  $\sigma_2$  from Figure 5; as long as neuron  $\sigma_1$  uses the rule  $a^2/a \rightarrow a; 0$ , the computation continues with two spikes sent in each step to neuron  $\sigma_5$ . When neuron  $\sigma_1$  uses the rule  $a^2 \rightarrow a; 1$ , this phase is finished, neuron  $\sigma_5$  receives only one spike (from neuron  $\sigma_2$  via neuron  $\sigma_4$ ), and neurons  $\sigma_1, \sigma_2, \sigma_3$  forget all their spikes. Let us assume that we have accumulated  $2n + 1$  spikes in neuron  $\sigma_5$ , for some  $n \geq 1$ .

In the moment when neuron  $\sigma_5$  receives a single spike, its rules can be applied (they need an odd number of spikes to fire).

For each step when neuron  $\sigma_5$  fires (hence for  $n$  times), neuron  $\sigma_{12}$  receives two spikes from neuron  $\sigma_5$ , via neurons  $\sigma_7$  and  $\sigma_8$ . Simultaneously, one spike of neuron  $\sigma_5$  is sent to neuron  $\sigma_6$  and from here to neurons  $\sigma_9$  and  $\sigma_{10}$ . Neuron  $\sigma_{10}$  fires only when it has an odd number of spikes inside, and this happens first time two steps after the first moment when neuron  $\sigma_5$  fires and once again one step after the last spike of neuron  $\sigma_5$  (neuron  $\sigma_6$  sends nothing to neuron  $\sigma_{10}$ , but neuron  $\sigma_9$  does it). Therefore, neuron  $\sigma_{12}$  will hold an odd number of spikes – and can fire for the first time – three steps after the last spike of neuron  $\sigma_5$  (we need three steps to pass from neuron  $\sigma_9$  to neuron  $\sigma_{11}$  and from here to neuron  $\sigma_{12}$ ). Now, neuron  $\sigma_{12}$  starts to spike, and it will do it for  $n + 1$  times, and then the computation will stop.

Let us consider the language  $T(\Pi_5) \cap b_5^* b_7 b_{12}^*$ , which means that we select only those traces where the marked spike passes from neuron  $\sigma_5$  to neuron  $\sigma_7$  and from here to neuron  $\sigma_{12}$  (note that it can go to neuron  $\sigma_6$ , then to  $\sigma_9$  or  $\sigma_{10}$ , and in neuron  $\sigma_{10}$  it can remain forever).

The strings in this language are of the form  $b_5^{n+1+i} b_7 b_{12}^m$ , where  $i = 0$  if the marked spike leaves neuron  $\sigma_5$  at the first use of the rule  $a(aa)^+/a^2 \rightarrow a; 0$ , or greater, with the maximal value  $n$ , in the case when the marked spike leaves the neuron in the last moment, by using the rule  $a \rightarrow a; 0$ . In turn,  $m$  can be 3 in the case  $i = n$  and the marked spike leaves neuron  $\sigma_{12}$  with the first application of the rule  $a(aa)^+/a^2 \rightarrow a; 0$ , or greater, with the maximal value being  $2n + 5$  in the case  $i = 0$  and the marked spike leaves neuron  $\sigma_{12}$  in the last moment, when using the rule  $a \rightarrow a; 0$ .

This language is not regular (we cannot pump a substring of the suffix  $b_{12}^m$ ), hence the language  $T(\Pi_5)$  is not regular.

**Theorem 5.**  $TSNP_{12}(rule_2, cons_2, forg_1) - REG \neq \emptyset$ .

Actually, much more complex languages can be generated. First, the previous example can be generalized in order to generate non-context-free languages: in the same way as passing the contents of neuron  $\sigma_5$  to neuron  $\sigma_{12}$ , we can pass also the contents of neuron  $\sigma_{12}$  to another neuron, by repeating the block of neurons  $\sigma_6 - \sigma_{12}$  under neuron  $\sigma_{12}$ . In this way, strings with three correlated substrings would be generated and a language with such strings (even selected by an intersection with a regular language) is not context-free.

Still, we can go much further.

**Theorem 6.** *Every unary language  $L \in RE$  can be written in the form  $L = h(L') = (d_1^* \setminus L') \cap d_2^*$ , where  $L' \in TSNP_*(rule_2, cons_3, forg_3)$ , and  $h$  is a projection.*

*Proof.* This result is a consequence of the fact that SN P systems can simulate register machines. Specifically, as proved in [7], starting from a register machine  $M$  (we do not need precise definitions and notations; details about register machines can be found in many books), we construct an SN P system  $\Pi$  which halts its computation with  $2n$  spikes in a specified neuron  $\sigma_{out}$  if and only if  $n$  can be generated by the register machine  $M$ ; in the halting moment, a neuron  $\sigma_{l_h}$  of  $\Pi$  associated with the label of the halting instruction of  $M$  gets two spikes and fires. The neuron  $\sigma_{out}$  contains no rule used in the simulation of  $M$  (the corresponding register is only incremented, but never decremented – see the details of the construction from [7]).

Now, consider a language  $L \subseteq b_2^*$ ,  $L \in RE$ . There is a register machine  $M$  such that  $n \in N(M)$  if and only if  $b_2^n \in L$ . Starting from such a machine  $M$ , we construct the system  $\Pi$  as in [7], having the properties described above. We append to the system  $\Pi$  six more neurons, as indicated in Figure 7. There is a marked spike in neuron  $\sigma_1$ , and it will stay here during all the simulation of  $M$ . In the moment when neuron  $\sigma_{l_h}$  of  $\Pi$  spikes, its spike goes both to neuron  $\sigma_{out}$  and to neuron  $\sigma_1$ .

Neurons  $\sigma_3, \sigma_4, \sigma_5, \sigma_6$  play the same role as neurons  $\sigma_6, \sigma_9, \sigma_{10}, \sigma_{12}$  in the system from Figure 6, namely to send a spike to neuron  $\sigma_2$  only when neuron  $\sigma_{out}$  has finished its work (this happens after  $n$  steps of using the rule  $a(aa)^+ / a^2 \rightarrow a; 0$ , for  $2n$  being the contents of neuron  $\sigma_{out}$  in the moment when neuron  $\sigma_{l_h}$  spikes).

The marked spike leaves neuron  $\sigma_1$  four steps after using the rule  $a^2 \rightarrow a; 4$ , hence five steps after the spiking of neuron  $\sigma_{l_h}$ . This means that the marked spike waits in neuron  $\sigma_2$  exactly  $n$  steps. When the spike of neuron  $\sigma_6$  reaches neuron  $\sigma_2$ , the two spikes present here, the marked one included, are forgotten.

Thus, the traces of the marked spike are of the form  $b_1^r b_2^n$ , for some  $r \geq 1$  and  $n \in N(M)$ . By means of the left derivative with the regular language  $b_1^*$  we can remove prefixes of the form  $b_1^k$  and by means of the intersection with  $b_2^*$  we ensure that the maximal prefix of this form is removed. Similarly, the projection  $h : \{b_1, b_2\}^* \rightarrow \{b_1, b_2\}^*$  defined by  $h(b_1) = \lambda$ ,  $h(b_2) = b_2$ , removes all occurrences of  $b_1$ . Consequently,  $L = (b_1^* \setminus T(\Pi)) \cap b_2^* = h(T(\Pi))$ .

The system  $\Pi$  from [7] (Theorem 7.1 there) has the rule complexity described by  $rule_2, cons_3, forg_3$ ; one sees that the construction from Figure 7 does not increase these parameters.

In what concerns the number of neurons, we do not have a bound for the language  $L'$  from the previous theorem.

**Corollary 1.** *Each family  $TSNP_*(reg_k, cons_p, forg_q)$  with  $k \geq 2, p \geq 3, q \geq 3$ , is incomparable with each family of languages  $FL$  which contains the singleton languages, is closed under left derivative with regular languages and intersection*



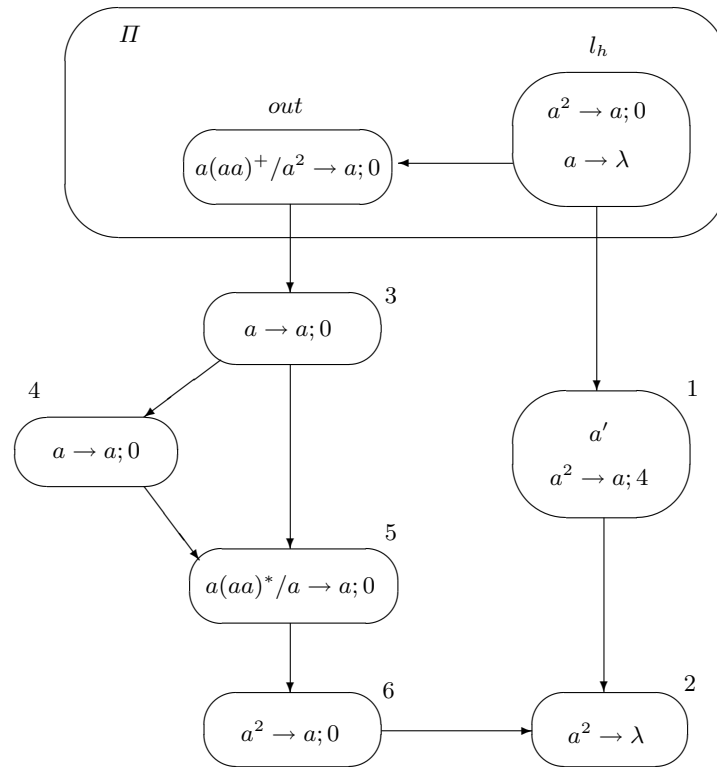


Fig. 7. The SN P system from the proof of Theorem 6

with regular languages, and does not contain all unary recursively enumerable languages.

Families as above are *FIN*, *REG*, *CF*, *CS*, *MAT* etc. (every one-letter language in *MAT* is regular, see [4]).

### 6 Final Remarks

We have considered here the possibility of generating a language by means of a spiking neural P system by following the traces of a distinguished spike during a halting computation. The power of such devices seems to be rather large – Theorem 6 – but we do not have a precise characterization of the obtained families of languages. Theorem 6 also raises the natural question whether RE languages over arbitrary alphabets can be represented starting from trace languages of SN P systems.

The main difficulty in handling the trace languages of SN P systems is, in the previous setup, the non-determinism of the marked spike evolution. A possible way to better control the marked spike is to consider spiking rules of the forms  $E'/a^c \rightarrow a;d$  and  $E/a'a^c \rightarrow a';d$ , with the meaning that the prime indicates whether or not the rule consumes the marked spike: this is not the case when the prime is attached to  $E$ , but this happens if the prime marks one of the consumed spikes, and hence also the produced spike. This removes the non-determinism induced by the use of the marked spike when applying the rules, but still non-determinism remains in what concerns the choice of synapses towards neighboring neurons (and this was the basis of the easy counterexample from Theorem 1). How also this non-determinism can be removed remains as a research topic.

## References

1. H. Chen, R. Freund, M. Ionescu, Gh. Păun, M.J. Pérez-Jiménez: On string languages generated by spiking neural P systems. In the present volume.
2. J. Dassow, Gh. Păun: *Regulated Rewriting in Formal Language Theory*. Springer-Verlag, Berlin, 1989.
3. P. Frisco, H.J. Hoogeboom: Simulating counter automata by P systems with symport/antiport. In *Membrane Computing. International Workshop, WMC 2002, Curtea de Argeş, Romania, August 2002. Revised Papers* (Gh. Păun, G. Rozenberg, A. Salomaa, C. Zandron, eds.), LNCS 2597, Springer-Verlag, Berlin, 2003, 288–301.
4. D. Hauschild, M. Jantzen: Petri nets algorithms in the theory of matrix grammars. *Acta Informatica*, 31 (1994), 719–728.
5. O.H. Ibarra, A. Păun, Gh. Păun, A. Rodríguez-Patón, P. Sosik, S. Woodworth: Normal forms for spiking neural P systems. In volume II of the present proceedings.
6. M. Ionescu, C. Martin-Vide, A. Păun, Gh. Păun: Membrane systems with symport/antiport: (unexpected) universality results. In *Proc. 8th International Meeting of DNA Based Computing* (M. Hagiya, A. Ohuchi, eds.), Japan, 2002, 151–160.
7. M. Ionescu, Gh. Păun, T. Yokomori: Spiking neural P systems. *Fundamenta Informaticae*, 71, 2-3 (2006), 279–308.
8. W. Maass: Computing with spikes. *Special Issue on Foundations of Information Processing of TELEMATIK*, 8, 1 (2002), 32–36.
9. Gh. Păun: *Membrane Computing – An Introduction*. Springer-Verlag, Berlin, 2002.
10. Gh. Păun, M.J. Pérez-Jiménez, G. Rozenberg: Spike trains in spiking neural P systems. *Intern. J. Found. Computer Sci.*, to appear (also available at [13]).
11. G. Rozenberg, A. Salomaa, eds.: *Handbook of Formal Languages*, 3 volumes. Springer-Verlag, Berlin, 1997.
12. A. Salomaa: *Formal Languages*. Academic Press, New York, 1973.
13. The P Systems Web Page: <http://psystems.disco.unimib.it>.