Spiking Neural P Systems with Extended Rules

Haiming Chen[1], Tseren-Onolt Ishdorj[3], Gheorghe Păun[2,3],
Mario J. Pérez-Jiménez[3]

[1] Computer Science Laboratory, Institute of Software
Chinese Academy of Sciences
100080 Beijing, China
chm@ios.ac.cn
[2] Institute of Mathematics of the Romanian Academy
PO Box 1-764, 014700 Bucharest, Romania
[3] Research Group on Natural Computing
Department of Computer Science and AI
University of Sevilla
Avda Reina Mercedes s/n, 41012 Sevilla, Spain
tseren@yahoo.com, gpaun@us.es, marper@us.es

Summary. We consider spiking neural P systems with spiking rules allowed to introduce zero, one, or more spikes at the same time. The computing power of the obtained systems is investigated, when considering them as number generating and as language generating devices. In the first case, a simpler proof of universality is obtained (universality is already known for the restricted rules), while in the latter case we find characterizations of finite and recursively enumerable languages (without using any squeezing mechanism, as it was necessary in the case of restricted rules). The relationships with regular languages are also investigated. In the end of the paper, a tool-kit for computing (some) operations with languages is provided.

1 Introduction

We combine here two ideas recently considered in the study of the spiking neural P systems (in short, SN P systems) introduced in [2], namely the *extended* rules from [4] and the *string generation* from [1].

For the reader's convenience, we shortly recall that 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 neuron is designated as the *output* neuron of the system and its spikes can exit into the environment, thus producing a *spike train*. Two main kinds of outputs can be associated with a computation in an SN P system: a set of numbers, obtained by considering the number of steps elapsed between consecutive spikes which exit the output neuron, and the string corresponding to the sequence of spikes which exit

the output neuron. This sequence is a binary one, with 0 associated with a step when no spike is emitted and 1 associated with a step when a spike is emitted.

The case of SN P systems as number generators was investigated in several papers, starting with [2], where it is proved that such systems are Turing complete (hence also universal, because the proof is constructive; universality in a rigorous framework was investigated in [4]). In turn, the string case is investigated in [1], where representations of finite, regular, and recursively enumerable languages were obtained, but also finite languages were found which cannot be generated in this way.

Here we consider an extension of the rules, already used in [4], namely we allow rules of the form $E/a^c \to a^p$, with the following meaning: if the content of the neuron is described by the regular expression E, then c spikes are consumed and p are produced and sent to the neurons to which there exist synapses leaving the neuron where the rule is applied (more precise definitions will be given in the next section). Thus, these rules cover and generalize at the same time both spiking rules and forgetting rules as considered so far in this area – with the mentioning that we do not also consider here a delay between firing and spiking, because in the proofs we never need such a delay.

As expected, this generalization allows much simpler constructions for the proof of universality in the case of considering SN P systems as number generators (we treat this issue in Section 4). More interesting is the case of strings produced by SN P systems with extended rules: we associate a symbol b_i to a step when the system sends i spikes into the environment, with two possible cases – b_0 is used as a separated symbol, or it is replaced by λ (sending no spike outside is interpreted as a step when the generated string is not grown). The first case is again restrictive: not all minimal linear languages can be obtained, but still results stronger than those from [1] can be proved in the new framework because of the possibility of removing spikes under the control of regular expressions – see Section 5. The freedom provided by the existence of steps when we have no output makes possible direct characterizations of finite and recursively enumerable languages (not only representations, modulo various operations with languages, as obtained in [1] for the standard binary case) – Section 6. In Section 7 we also present constructions of SN P systems for computing some usual operations with languages: union, concatenation, weak coding, intersection with regular languages.

2 Formal Language Theory Prerequisites

We assume the reader to be familiar with basic language and automata theory, e.g., from [6] and [7], 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 λ, 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 $\{a\}^*, \{a\}^+$. If $x = a_1 a_2 \dots a_n$, $a_i \in V$, $1 \leq i \leq n$, then $mi(x) = a_n \dots a_2 a_1$.

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

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 the language L_2 is a singleton, these operations are called left and right derivatives, and denoted by $\partial_x^l(L) = \{x\} \backslash L$ and $\partial_x^r(L) = L / \{x\}$, respectively.

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 $A \to aB, A \to a$, for some $A, B \in N, a \in T$.

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. The family of Turing computable sets of numbers is denoted by NRE (these sets are length sets of RE languages, hence the notation).

Let $V = \{b_1, b_2, \ldots, b_m\}$, for some $m \geq 1$. For a string $x \in V^*$, let us denote by $val_m(x)$ the value in base $m + 1$ of x (we use base $m + 1$ in order to consider the symbols $b_1, \ldots, b_m$ as digits $1, 2, \ldots, m$, thus avoiding the digit 0 in the left hand of the string). We extend this notation in the natural way to sets of strings.

All universality results of the paper are based on the notion of a register machine. Such a device – in the non-deterministic version – is a construct $M = (m, H, l_0, l_h, I)$, where m is the number of registers, H is the set of instruction labels, l_0 is the start label (labeling an ADD instruction), l_h is the halt label (assigned to instruction `HALT`), and I is the set of instructions; each label from H labels only one instruction from I, thus precisely identifying it. The instructions are of the following forms:

- $l_i : (\texttt{ADD}(r), l_j, l_k)$ (add 1 to register r and then go to one of the instructions with labels l_j, l_k non-deterministically chosen),
- $l_i : (\texttt{SUB}(r), l_j, l_k)$ (if register r is non-empty, then subtract 1 from it and go to the instruction with label l_j, otherwise go to the instruction with label l_k),
- $l_h : \texttt{HALT}$ (the halt instruction).

A register machine M generates a set $N(M)$ of numbers in the following way: we start with all registers empty (i.e., storing the number zero), we apply the instruction with label l_0 and we continue to apply instructions as indicated by the labels (and made possible by the contents of registers); if we reach the halt instruction, then the number n present in register 1 at that time is said to be generated by M. (Without loss of generality we may assume that in the halting configuration all other registers are empty; also, we may assume that register 1 is never subject of SUB instructions, but only of ADD instructions.) It is known (see, e.g., [3]) that register machines generate all sets of numbers which are Turing computable.

A register machine can also be used as a number accepting device: we introduce a number n is some register r_0, we start working with instruction with label l_0, and if the machine eventually halts, then n is accepted (we may also assume that all registers are empty in the halting configuration). Again, accepting register machines characterize NRE.

Furthermore, register machines can compute all Turing computable functions: we introduce the numbers $n_1, \ldots, n_k$ in some specified registers $r_1, \ldots, r_k$, we start with the instruction with label l_0, and when we stop (with the instruction with label l_h) the value of the function is placed in another specified register, r_t, with all registers different from r_t being empty.

In both the accepting and the computing case, the register machine can be *deterministic*, i.e., with the ADD instructions of the form $l_i : (\texttt{ADD}(r), l_j)$ (add 1 to register r and then go to the instruction with label l_j).

In the following sections, when comparing the power of two language generating/accepting devices the empty string λ is ignored.

3 Spiking Neural P Systems with Extended Rules

We directly introduce the type of SN P systems we investigate in this paper; the reader can find details about the standard definition in [2], [5], [1], etc.

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

$$\Pi = (O, \sigma_1, \ldots, \sigma_m, syn, i_0),$$

where:

1. $O = \{a\}$ is the singleton alphabet (a is called *spike*);
2. $\sigma_1, \ldots, \sigma_m$ 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 σ_i;
   b) R_i is a finite set of *rules* of the form $E/a^c \rightarrow a^p$, where E is a regular expression over a and $c \geq 1, p \geq 0$, with the restriction $c \geq p$;
3. $syn \subseteq \{1, 2, \ldots, m\} \times \{1, 2, \ldots, m\}$ with $i \neq j$ for each $(i, j) \in syn, 1 \leq i, j \leq m$ (*synapses* between neurons);
4. $i_0 \in \{1, 2, \ldots, m\}$ indicates the *output neuron* (σ_{i_0}) of the system.

A rule $E/a^c \rightarrow a^p$ is applied as follows. If the neuron σ_i contains k spikes, and $a^k \in L(E), k \geq c$, then the rule can *fire*, and its application means consuming (removing) c spikes (thus only $k - c$ remain in σ_i) and producing p spikes, which will exit immediately the neuron. A global clock is assumed, marking the time for the whole system, hence the functioning of the system is synchronized.

Note that we do not consider here a delay between firing and spiking (i.e., rules of the form $E/a^c \to a^p; d$, with $d \geq 0$), because we do not need this feature in the proofs below, but such a delay can be introduced in the usual way. (As a consequence, here the neurons are always open.)

If a rule $E/a^c \to a^p$ has $E = a^c$, then we will write it in the simplified form $a^c \to a^p$.

The spikes emitted by a neuron σ_i go to all neurons σ_j such that $(i, j) \in syn$, i.e., if σ_i has used a rule $E/a^c \to a^p$, then each neuron σ_j receives p spikes.

If several rules can be used at the same time, then the one to be applied is chosen non-deterministically.

During the computation, a configuration of the system is described by the number of spikes present in each neuron; thus, the initial configuration is described by the numbers $n_1, n_2, \ldots, n_m$.

Using the rules as described above, one can define transitions among configurations. Any sequence of transitions starting in the initial configuration is called a *computation*. A computation halts if it reaches a configuration where no rule can be used. With any computation (halting or not) we associate a *spike train*, the sequence of symbols 0 and 1 describing the behavior of the output neuron: if the output neuron spikes, then we write 1, otherwise we write 0 (note that at this stage we ignore the number of spikes emitted by the output neuron into the environment in each step, but this additional information will be considered below).

As the result of a computation, in [2] and [5] one considers the distance between two consecutive steps when there are spikes which exit the system, with many possible variants: taking the distance between the first two occurrences of 1 in the spike train, between all consecutive occurrences, considering only alternately the intervals between occurrences of 1, etc. For simplicity, we consider here only the first case mentioned above: we denote by $N_2(\Pi)$ the set of numbers generated by an SN P system in the form of the number of steps between the first two steps of a computation when spikes are emitted into environment, and by $Spik_2SN^eP_m(rule_k, cons_p, prod_q)$ the family of sets $N_2(\Pi)$ generated by SN P systems with at most m neurons, at most k rules in each neuron, consuming at most p and producing at most q spikes. Any of these parameters is replaced by $*$ if it is not bounded.

Following [1] we can also consider as the result of a computation the spike train itself, thus associating a language with an SN P system. Specifically, like in [1], we can consider the language $L_{bin}(\Pi)$ of all binary strings associated with halting computations in Π: the digit 1 is associated with a step when one or more spikes exit the output neuron, and 0 is associated with a step when no spike is emitted by the output neuron.

Because several spikes can exit at the same time, we can also work on an arbitrary alphabet: let us associate the symbol b_i with a step when the output neuron emits i spikes. We have two cases: interpreting b_0 (hence a step when no spike is emitted) as a symbol or as the empty string. In the first case we denote

the generated language by $L_{res}(\Pi)$ (with "res" coming from "restricted"), in the latter one we write $L_\lambda(\Pi)$.

The respective families are denoted by $L_\alpha SN^e P_m(rule_k, cons_p, prod_q)$, where $\alpha \in \{bin, res, \lambda\}$ and parameters m, k, p, q are as above.

4 Extended SN P Systems as Number Generators

Because non-extended SN P systems are already computationally universal, this result is directly valid also for extended systems. However, the construction on which the proof is based is much simpler in the extended case (in particular, it does not use the delay feature), that is why we briefly present it.

Theorem 1. $NRE = Spik_2 SN^e P_*(rule_5, cons_5, prod_2)$.

Proof. The proof of the similar result from [2] is based on constructing an SN P system Π which simulates a given register machine M. The idea is that each register r has associated a neuron σ_r, with the value n of the register represented by $2n$ spikes in neuron σ_r. Also, each label of M has a neuron in Π, which is "activated" when receiving two spikes. We do not recall other details from [2], and we pass directly to presenting modules for simulating the ADD and the SUB instructions of M, as well as an OUTPUT module, in the case of using extended rules: Figures 1, 2, and 3.

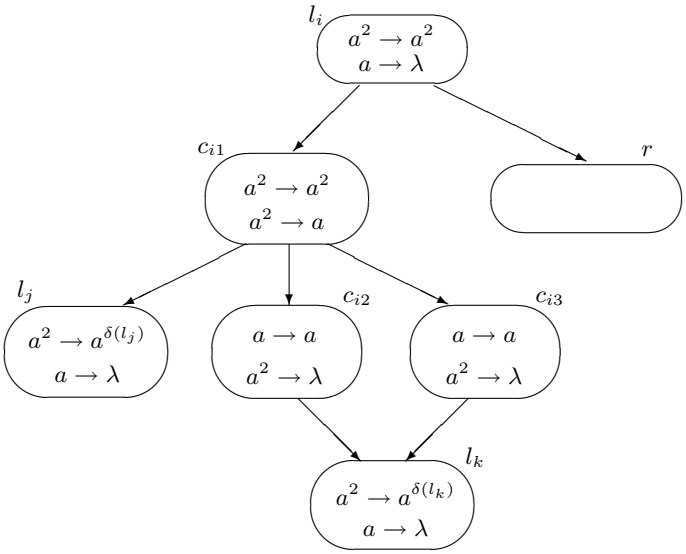

Fig. 1. Module ADD, for simulating an instruction $l_i : (\texttt{ADD}(r), l_j, l_k)$

Because the neurons associated with labels of ADD and SUB instructions have to produce different numbers of spikes, in the neurons associated with "output" labels of instructions we have written the rules in the form $a^2 \to a^{\delta(l)}$, with $\delta(l) = 1$ for l being the label of a SUB instruction and $\delta(l) = 2$ if l is the label of an ADD instruction.

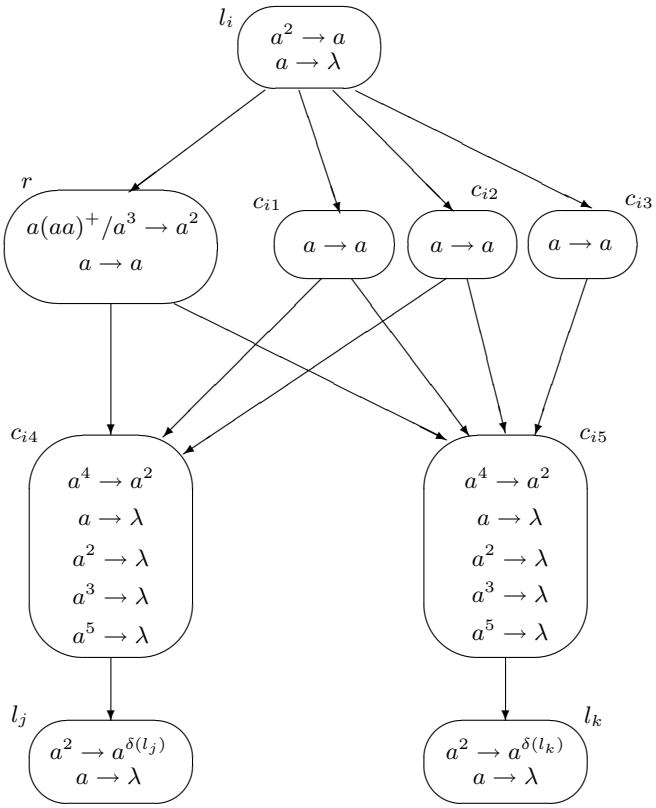

Fig. 2. Module SUB, for simulating an instruction $l_i : (\texttt{SUB}(r), l_j, l_k)$

Because l_i precisely identifies the instruction, the neurons $c_{i\alpha}$ are distinct for distinct instructions. However, an interference between SUB modules appears in the case of instructions SUB which operate on the same register r: synapses $(r, c_{is}), (r, c_{i's}), s = 4, 5$, exist for different instructions $l_i : (\texttt{SUB}(r), l_j, l_k)$, $l_{i'} : (\texttt{SUB}(r), l_{j'}, l_{k'})$. Neurons $\sigma_{c_{i's}}, \sigma_{c_{i'5}}$ receive 1 or 2 spikes from σ_r even when simulating the instruction with label l_i, but they are immediately forgotten (this is the role of rules $a \to \lambda$, $a^2 \to \lambda$ from neurons $\sigma_{c_{i4}}, \sigma_{c_{i5}}$).

The task of checking the functioning of the modules from Figures 1, 2, 3 is left to the reader. □

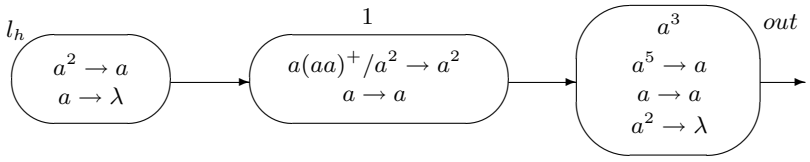

Fig. 3. Module OUTPUT

5 Languages in the Restricted Case

We pass now to considering SN P systems as language generators, starting with the restricted case, when the system outputs a symbol in each computation step.

In all considerations below, we work with the alphabet $V = \{b_1, b_2, \ldots, b_m\}$, for some $m \geq 1$. By a simple renaming of symbols, we may assume that any given language L is a language over V. When a symbol b_0 is also used, it is supposed that $b_0 \notin V$.

5.1 A Characterization of FIN

SN P systems with standard rules cannot generate all finite languages (see [1]), but extended rules help in this respect.

Lemma 1. $L_\alpha SN^e P_1(rule_*, cons_*, prod_*) \subseteq FIN$, $\alpha \in \{res, \lambda\}$.

Proof. In each step, the number of spikes present in a system with only one neuron decreases by at least one, hence any computation lasts at most as many steps as the number of spikes present in the system at the beginning. Thus, the generated strings have a bounded length. □

Lemma 2. $FIN \subseteq L_\alpha SN^e P_1(rule_*, cons_*, prod_*)$, $\alpha \in \{res, \lambda\}$.

Proof. Let $L = \{x_1, x_2, \ldots, x_n\} \subseteq V^*$, $n \geq 1$, be a finite language, and let $x_i = x_{i,1} \ldots x_{i,r_i}$ for $x_{i,j} \in V, 1 \leq i \leq n, 1 \leq j \leq r_i = |x_i|$. Denote $l = \max\{r_i \mid 1 \leq i \leq n\}$. For $b \in V$, define $index(b) = i$ if $b = b_i$. Define $\alpha_j = lm \sum_{i=1}^{j} |x_i|$, for all $1 \leq j \leq n$.

An SN P system that generates L is shown in Figure 4.

Initially, only a rule $a^{\alpha_n + lm}/a^{\alpha_n - \alpha_j + m} \to a^{index(x_{j,1})}$ can be used, and in this way we non-deterministically chose the string x_j to generate. This rule outputs the necessary number of spikes for $x_{j,1}$. Then, because $\alpha_j + (l-1)m$ spikes remain in the neuron, we have to continue with rules $a^{\alpha_j - t + 2 + (l-t+1)m}/a^m \to a^{index(x_{j,t})}$, for $t = 2$, and then for the respective $t = 3, 4, \ldots, r_j - 1$; in this way we introduce $x_{j,t}$, for all $t = 2, 3, \ldots, r_j - 1$. In the end, the rule $a^{\alpha_j - r_j + 2 + (l-r_j+1)m} \to a^{index(x_{j,r_j})}$ is used, which produces x_{j,r_j} and concludes the computation.

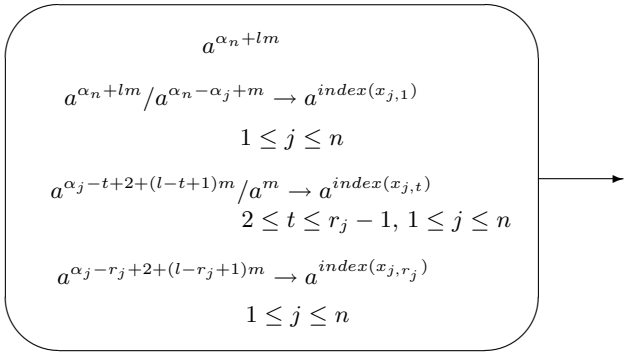

Fig. 4. An SN P system generating a finite language

It is easy to see that the rules which are used in the generation of a string x_j cannot be used in the generation of a string x_k with $k \neq j$. Also, in each rule the number of spikes consumed is not less than the number of spikes produced. The system Π never outputs zero spikes, hence $L_{res}(\Pi) = L_\lambda(\Pi) = L$. □

Theorem 2. $FIN = L_{res}SN^eP_1(rule_*, cons_*, prod_*) = L_\lambda SN^eP_1(rule_*, cons_*, prod_*)$.

This characterization is sharp in what concerns the number of neurons, because of the following result:

Proposition 1. $L_\alpha SN^eP_2(rule_2, cons_3, prod_3) - FIN \neq \emptyset$, $\alpha \in \{res, \lambda\}$.

Proof. The SN P system Π from Figure 5 generates the infinite language $L_{res}(\Pi) = L_\lambda(\Pi) = b_3^* b_1 \{b_1, b_3\}$. □

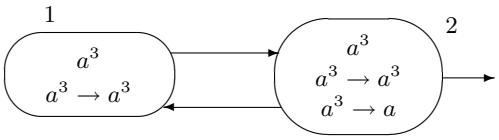

Fig. 5. An SN P system generating an infinite language

5.2 Representations of Regular Languages

Such representations are obtained in [1] starting from languages of the form $L_{bin}(\Pi)$, but in the extended SN P systems, regular languages can be represented in an easier and more direct way.

Theorem 3. *If* $L \subseteq V^*$, $L \in REG$, *then* $\{b_0\}L \in L_{res}SN^eP_4(rule_*, cons_*, prod_*)$.

Proof. Consider a regular grammar $G = (N, V, S, P)$ such that $L = L(G)$, where $N = \{A_1, A_2, \ldots, A_n\}$, $n \geq 1$, $S = A_n$, and the rules in P are of the forms $A_i \to b_k A_j$, $A_i \to b_k$, $1 \leq i, j \leq n$, $1 \leq k \leq m$.

Then $\{b_0\}L$ can be generated by an SN P system as shown in Figure 6.

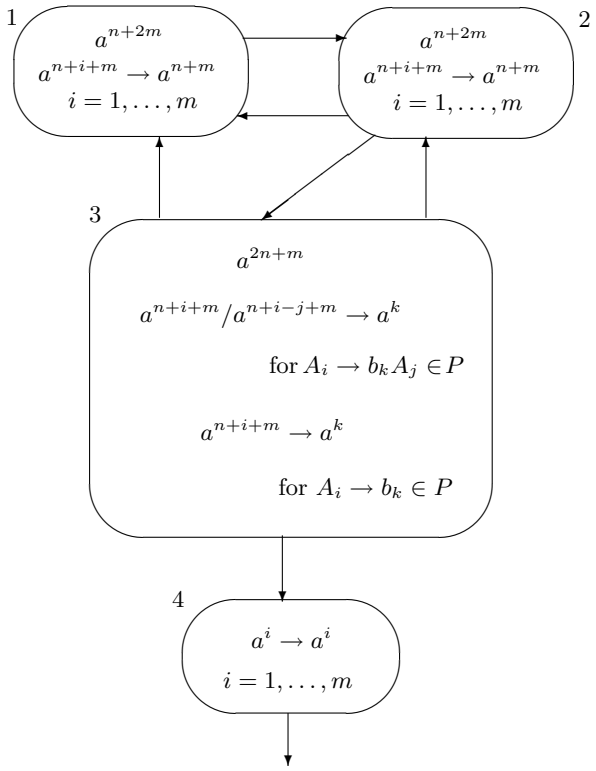

Fig. 6. The SN P system from the proof of Theorem 3

In each step, neurons σ_1 and σ_2 will send $n + m$ spikes to neuron σ_3, provided neuron σ_2 receives spikes from neuron σ_3. Neuron σ_3 fires in the first step by a rule $a^{2n+m}/a^{2n-j+m} \to a^k$ (or $a^{2n+m} \to a^k$) associated with a rule $A_n \to b_k A_j$ (or $A_n \to b_k$) from P, produces k spikes and receives $n + m$ spikes from neuron σ_2. In the meantime neuron σ_4 does not spike, hence it produces the symbol b_0, and receives spikes from neuron σ_3, therefore in the second step it generates the first symbol of the string.

Assume in some step t, the rule $a^{n+i+m}/a^{n+i-j+m} \to a^k$, for $A_i \to b_k A_j$, or $a^{n+i+m} \to a^k$, for $A_i \to b_k$, is used, for some $1 \leq i \leq n$, and $n+m$ spikes are received from neuron σ_2.

If the first rule is used, then k spikes are produced, $n+i-j+m$ spikes are consumed and j spikes remain in neuron σ_3. Then in step $t+1$, we have $n+j+m$ spikes in neuron σ_3, and a rule for $A_j \to b_k A_l$ or $A_j \to b_k$ can be used. In step $t+1$ neuron σ_3 also receives $n+m$ spikes from σ_2. In this way, the computation continues, unless the second rule is used.

If the second rule is used, then k spikes are produced, all spikes are consumed, and $n+m$ spikes are received in neuron σ_3. Then, in the next time step, neuron σ_3 receives $n+m$ spikes, but no rule can be used, so no spike is produced. At the same time, neuron σ_4 fires using spikes received from neuron σ_3 in the previous step, and then the computation halts.

In this way, all the strings in $\{b_0\}L$ can be generated. □

Corollary 1. *Every language $L \in REG, L \subseteq V^*$, can be written in the form $L = \partial^l_{b_0}(L')$ for some $L' \in L_{res}SN^eP_4(rule_*, cons_*, prod_*)$.*

One neuron in the previous representation can be saved, by adding the extra symbol in the right hand end of the string.

Theorem 4. *If $L \subseteq V^*$, $L \in REG$, then $L\{b_0\} \in L_{res}SN^eP_3(rule_*, cons_*, prod_*)$.*

Proof. The proof is based on a construction similar to the one from the proof of Theorem 3. Specifically, starting from a regular grammar G as above, we construct a system Π as in Figure 7, for which we have $L_{res}(\Pi) = L\{b_0\}$. We leave the task to check this assertion to the reader. □

Corollary 2. *Every language $L \in REG, L \subseteq V^*$, can be written in the form $L = \partial^r_{b_0}(L')$ for some $L' \in L_{res}SN^eP_3(rule_*, cons_*, prod_*)$.*

5.3 Going Beyond REG

We do not know whether the additional symbol b_0 can be avoided in the previous theorems (hence whether the regular languages can be directly generated by SN P systems in the restricted way), but such a result is not valid for the family of minimal linear languages (generated by linear grammars with only one nonterminal symbol).

Lemma 3. *The number of configurations reachable after n steps by an extended SN P system of degree m is bounded by a polynomial $g(n)$ of degree m.*

Proof. Let us consider an extended SN P system $\Pi = (O, \sigma_1, \ldots, \sigma_m, syn, i_0)$ of degree m, let n_0 be the total number of spikes present in the initial configuration of Π, and denote $\alpha = \max\{p \mid E/a^c \to a^p \in R_i, 1 \leq i \leq m\}$ (the maximal number

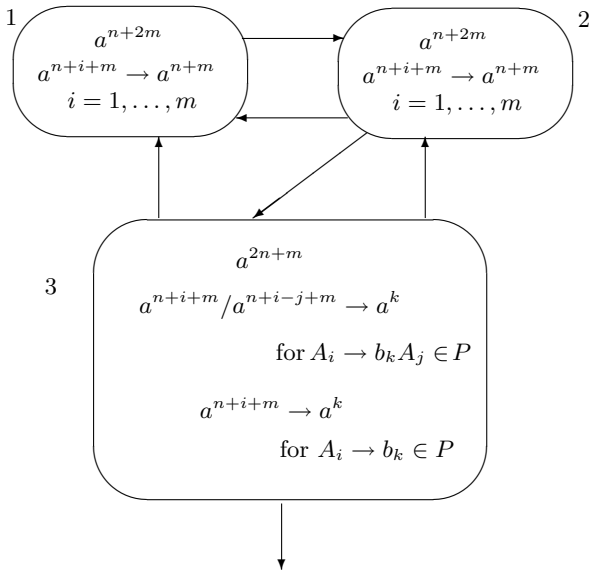

Fig. 7. The SN P system for the proof of Theorem 4

of spikes produced by any of the rules of Π). In each step of a computation, each neuron σ_i consumes some c spikes and produces $p \leq c$ spikes; these spikes are sent to all neurons σ_j such that $(i, j) \in syn$. There are at most $m - 1$ synapses $(i, j) \in syn$, hence the p spikes produced by neuron σ_i are replicated in at most $p(m - 1)$ spikes. We have $p(m - 1) \leq \alpha(m - 1)$. Each neuron can do the same, hence the maximal number of spikes produced in one step is at most $\alpha(m - 1)m$. In n consecutive steps, this means at most $\alpha(m-1)mn$ spikes. Adding the initial n_0 spikes, this means that after any computation of n steps we have at most $n_0 + \alpha(m-1)mn$ spikes in Π. These spikes can be distributed in the m neurons in less that $(n_0+\alpha(m-1)mn)^m$ different ways. This is a polynomial of degree m in n (α is a constant) which bounds from above the number of possible configurations obtained after computations of length n in Π. □

Theorem 5. *If $f : V^+ \longrightarrow V^+$ is an injective function, $card(V) \geq 2$, then there is no extended SN P system Π such that $L_f(V) = \{x\, f(x) \mid x \in V^+\} = L_{res}(\Pi)$.*

Proof. Assume that there is an extended SN P system Π of degree m such that $L_{res}(\Pi) = L_f(V)$ for some f and V as in the statement of the theorem. According to the previous lemma, there are only polynomially many configurations of Π which can be reached after n steps. However, there are $card(V)^n \geq 2^n$ strings of length n in V^+. Therefore, for large enough n there are two strings $w_1, w_2 \in V^+, w_1 \neq w_2$, such that after n steps the system Π reaches the same configuration when generating the strings $w_1\, f(w_1)$ and $w_2\, f(w_2)$, hence after step n the system can continue any of the two computations. This means that also

the strings $w_1 f(w_2)$ and $w_2 f(w_1)$ are in $L_{res}(\Pi)$. Due to the injectivity of f and the definition of $L_f(V)$ such strings are not in $L_f(V)$, hence the equality $L_f(V) = L_{res}(\Pi)$ is contradictory. □

Corollary 3. *The following languages are not in* $L_{res}SN^eP_*(rule_*, cons_*, prod_*)$ *(in all cases,* $card(V) = k \geq 2$*):*

$$L_1 = \{x\, mi(x) \mid x \in V^+\},$$
$$L_2 = \{xx \mid x \in V^+\},$$
$$L_3 = \{x\, c^{val_k(x)} \mid x \in V^+\}, c \notin V.$$

Note that language L_1 above is a non-regular minimal linear one, L_2 is context-sensitive non-context-free, and L_3 is non-semilinear. In all cases, we can also add a fixed tail of any length (e.g., considering $L_1' = \{x\, mi(x)z \mid x \in V^+\}$, where $z \in V^+$ is a given string), and the conclusion is the same – hence a result like that in Theorem 4 cannot be extended to minimal linear languages.

6 Languages in the Non-Restricted Case

As expected, the possibility of having intermediate steps when no output is produced is helpful, because this provides intervals for internal computations. In this way, we can get rid of the operations used in [1] and in the previous section when dealing with regular and with recursively enumerable languages.

6.1 Relationships with REG

Lemma 4. $L_\lambda SN^eP_2(rule_*, cons_*, prod_*) \subseteq REG$.

Proof. In a system with two neurons, the number of spikes from the system can remain the same after a step, but it cannot increase: the neurons can consume the same number of spikes as they produce, and they can send to each other the produced spikes. Therefore, the number of spikes in the system is bounded by the number of spikes present at the beginning. This means that the system can pass through a finite number of configurations and these configurations can control the evolution of the system like states in a finite automaton. Consequently, the generated language is regular (see similar reasonings, with more technical details, in [2], [1]). □

Lemma 5. $REG \subseteq L_\lambda SN^eP_3(rule_*, cons_*, prod_*)$.

Proof. For the SN P system Π constructed in the proof of Theorem 4 (Figure 7) we have $L_\lambda(\Pi) = L(G)$. □

This last inclusion is proper:

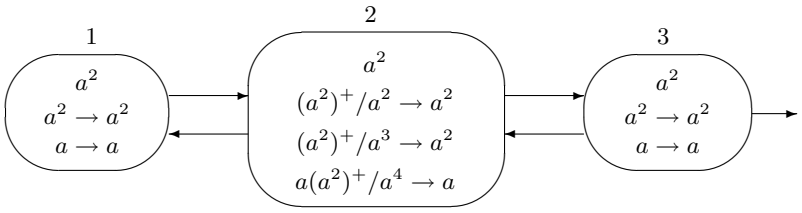

Fig. 8. An SN P system generating a non-regular language

Proposition 2. $L_\lambda SN^eP_3(rule_3, cons_4, prod_2) - REG \neq \emptyset$.

Proof. The SN P system Π from Figure 8 generates the language $L_\lambda(\Pi) = \{b_2^{n+1}b_1^n \mid n \geq 2\}$. Indeed, for a number of steps, neuron σ_2 consumes two spikes by using the rule $(a^2)^+/a^2 \to a^2$ and receives four from the other two neurons. After changing the parity of the number of spikes (by using the rule $(a^2)^+/a^3 \to a^2$), neuron σ_2 will continue by consuming four spikes (using the rule $a(a^2)^+/a^4 \to a$) and receiving only two. □

Corollary 4. $L_\lambda SN^eP_1(rule_*, cons_*, prod_*) \subset L_\lambda SN^eP_2(rule_*, cons_*, prod_*) \subset L_\lambda SN^eP_3(rule_*, cons_*, prod_*)$, *strict inclusions.*

6.2 Going Beyond CF

Actually, much more complex languages can be generated by extended SN P systems with three neurons.

Theorem 6. *The family* $L_\lambda SN^eP_3(rule_3, cons_6, prod_4)$ *contains non-semilinear languages.*

Proof. The system Π from Figure 9 generates the language

$$L_\lambda(\Pi) = \{b_4^2 b_2 b_4^{2^2} b_2 \dots b_4^{2^n} b_2 \mid n \geq 1\}.$$

We start with $2 + 4 \cdot 2^0$ spikes in neuron σ_1. When moved from neuron σ_1 to neuron σ_3, the number of spikes is doubled, because they pass both directly from σ_1 to σ_3, and through σ_2. When all spikes are moved to σ_3, the rule $a^2 \to a$ of σ_1 should be used. With a number of spikes of the form $4m + 1$, neuron σ_3 cannot fire, but in the next step one further spike comes from σ_2, hence the first rule of σ_3 can now be applied. Using this rule, all spikes of σ_3 are moved back to σ_1 – in the last step we use the rule $a^2 \to a^2$, which makes again the first rule of σ_1 applicable.

This process can be repeated any number of times. In each moment, after moving all but the last 6 spikes from neuron σ_1 to σ_3, we can also use the rule $a^6 \to a^3$ of σ_1, and this ends the computation: there is no spike in σ_1, neuron

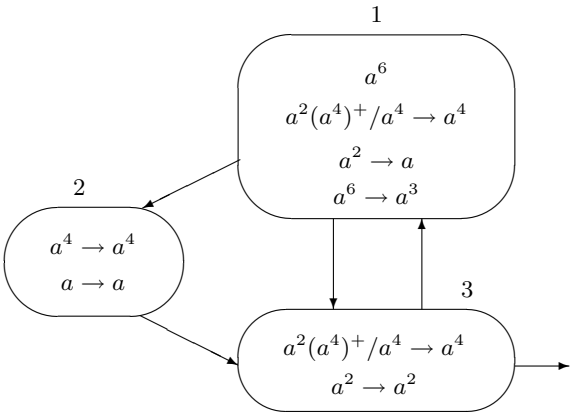

Fig. 9. An SN P system generating a non-semilinear language

σ_2 cannot work when having 3 spikes inside, and the same with σ_3 when having $4m + 3$ spikes.

Now, one sees that σ_3 is also the output neuron and that the number of times of using the first rule of σ_3 is doubled after each move of the contents of σ_3 to σ_1. □

In this proof we made use of the fact that no spike of the output neuron means no symbol introduced in the generated string. If we work in the restricted case, then symbols b_0 are shuffled in the string, hence the non-semilinearity of the generated language is preserved.

6.3 A Characterization of RE

If we do not bound the number of neurons, then a characterization of recursively enumerable languages is obtained.

Let us write m in front of a language family notation in order to denote the subfamily of languages over an alphabet with at most m symbols (e.g., $2RE$ denotes the family of recursively enumerable languages over alphabets with one or two symbols).

Lemma 6. *$mRE \subseteq mL_\lambda SN^eP_*(rule_{m'}, cons_m, prod_m)$, where $m' = \max(m, 2)$ and $m \geq 1$.*

Proof. We follow here the same idea as in the proof of Theorem 5.9 from [1], adapted to the case of extended rules.

Take an arbitrary language $L \subseteq V^*, L \in RE$. Obviously, $L \in RE$ if and only if $val_m(L) \in NRE$. In turn, a set of numbers is recursively enumerable if and only if it can be accepted by a deterministic register machine. Let M_1 be such a register machine, i.e., $N(M_1) = val_m(L)$.

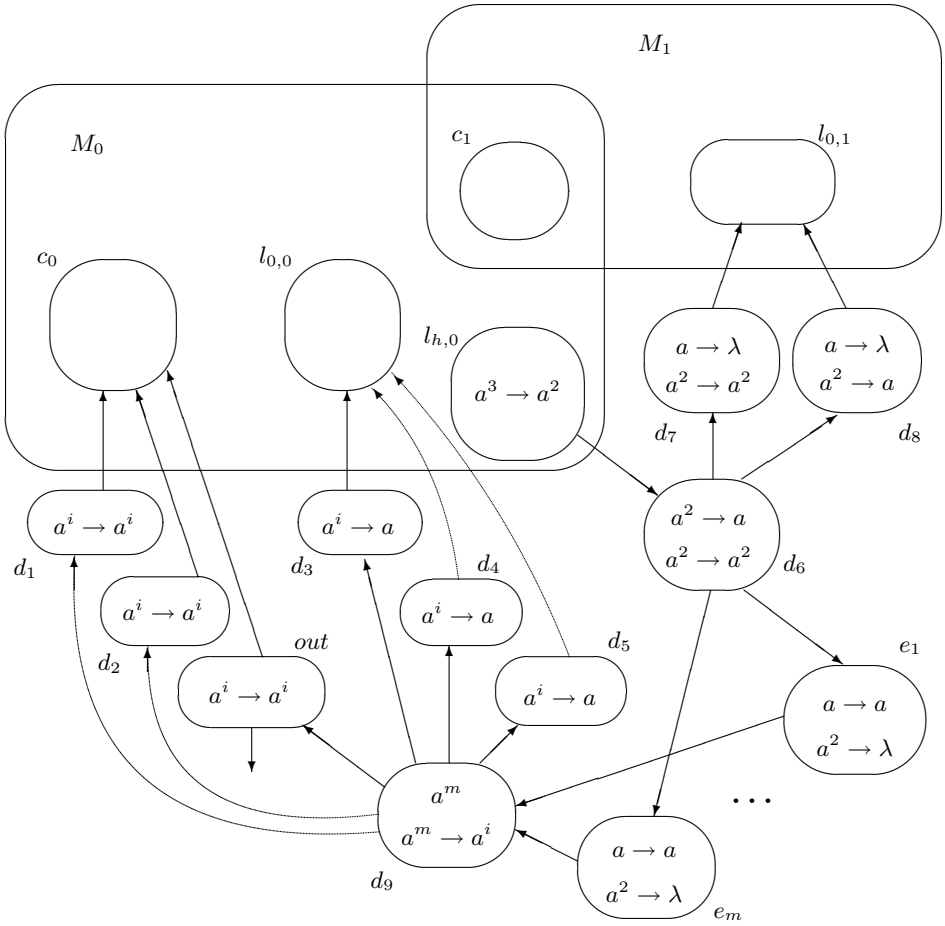

Fig. 10. The structure of the SN P system from the proof of Lemma 6

We construct an SN P system Π performing the following operations (σ_{c_0} and σ_{c_1} are two distinguished neurons of Π, which are empty in the initial configuration):

1. Output i spikes, for some $1 \leq i \leq m$, and at the same time introduce the number i in neuron σ_{c_0}; in the construction below, a number n is represented in a neuron by storing there $3n$ spikes, hence the previous task means introducing $3i$ spikes in neuron σ_{c_0}.
2. Multiply the number stored in neuron σ_{c_1} (initially, we have here number 0) by $m + 1$, then add the number from neuron σ_{c_0}; specifically, if neuron σ_{c_0} holds $3i$ spikes and neuron σ_{c_1} holds $3n$ spikes, $n \geq 0$, then we end this step

with $3(n(m+1)+i)$ spikes in neuron σ_{c_1} and no spike in neuron σ_{c_0}. In the meantime, the system outputs no spike.

3. Repeat from step 2, or, non-deterministically, stop the increase of spikes from neuron σ_{c_1} and pass to the next step.
4. After the last increase of the number of spikes from neuron σ_{c_1} we have here $val_m(x)$ for a string $x \in V^+$. Start now to simulate the work of the register machine M_1 in recognizing the number $val_m(x)$. The computation halts only if this number is accepted by M_1, hence the string x produced by the system is introduced in the generated language only if $val_m(x) \in N(M_1)$.

In constructing the system Π we use the fact that a register machine can be simulated by an SN P system. Then, the multiplication by $m+1$ of the contents of neuron σ_{c_1} followed by adding a number between 1 and m is done by a computing register machine (with the numbers stored in neurons $\sigma_{c_0}, \sigma_{c_1}$ introduced in two specified registers); we denote by M_0 this machine. Thus, in our construction, also for this operation we can rely on the general way of simulating a register machine by an SN P system. All other modules of the construction (introducing a number of spikes in neuron σ_{c_0}, sending out spikes, choosing non-deterministically to end the string to generate and switching to the checking phase, etc.) are explicitly presented below.

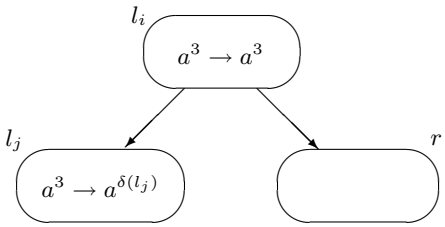

Fig. 11. Module ADD (simulating $l_i : (\texttt{ADD}(r), l_j)$)

The overall appearance of Π is given in Figure 10, where M_0 indicates the subsystem corresponding to the simulation of the register machine $M_0 = (m_0, H_0, l_{0,0}, l_{h,0}, I_0)$ and M_1 indicates the subsystem which simulates the register machine $M_1 = (m_1, H_1, l_{0,1}, l_{h,1}, I_1)$. Of course, we assume $H_0 \cap H_1 = \emptyset$. In all cases, $i \in \{1, 2, \ldots, m\}$.

We start with spikes only in neuron σ_{d_9}. We spike in the first step, non-deterministically choosing the number i of spikes to produce, hence the first letter b_i of the generated string. Simultaneously, i spikes are sent out by the output neuron, $3i$ spikes are sent to neuron σ_{c_0}, and three spikes are sent to neuron $\sigma_{l_{0,0}}$, thus triggering the start of a computation in M_0. The subsystem corresponding to the register machine M_0 starts to work, multiplying the value of σ_{c_1} with $n+1$

and adding i. When this process halts, neuron $\sigma_{l_{h,0}}$ is activated, and in this way two spikes are sent to neuron σ_{d_6}.

This is the neuron which non-deterministically chooses whether the string should be continued or we pass to the second phase of the computation, of checking whether the produced string is accepted. In the first case, neuron σ_{d_6} uses the rule $a^2 \to a$, which makes neurons $\sigma_{e_1}, \ldots, \sigma_{e_m}$ spike; these neurons send m spikes to neuron σ_{d_9}, like in the beginning of the computation. In the latter case, one uses the rule $a^2 \to a^2$, which activates the neuron $\sigma_{l_{0,1}}$ by sending three spikes to it, thus starting the simulation of the register machine M_1. The computation stops if and only if $val_m(x)$ is accepted by M_1.

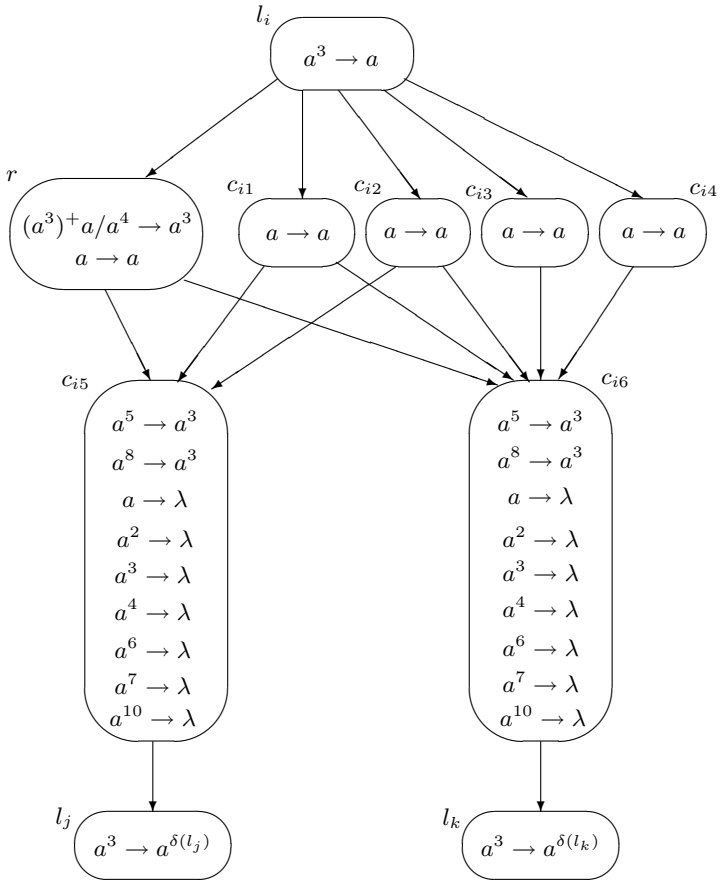

Fig. 12. Module SUB (simulating $l_i : (\texttt{SUB}(r), l_j, l_k)$) for machine M_0

In order to complete the proof we need to show how the two register machines are simulated, using the common neuron σ_{c_1} but without mixing the computations. To this aim, we consider the modules ADD and SUB from Figures 11, 12, and 13. Like in Section 4, neurons are associated with each label of the machine (they fire if they have three spikes inside) and with each register (with $3t$ spikes representing the number t from the register); there also are additional neurons with primed labels – it is important to note that all these additional neurons have distinct labels.

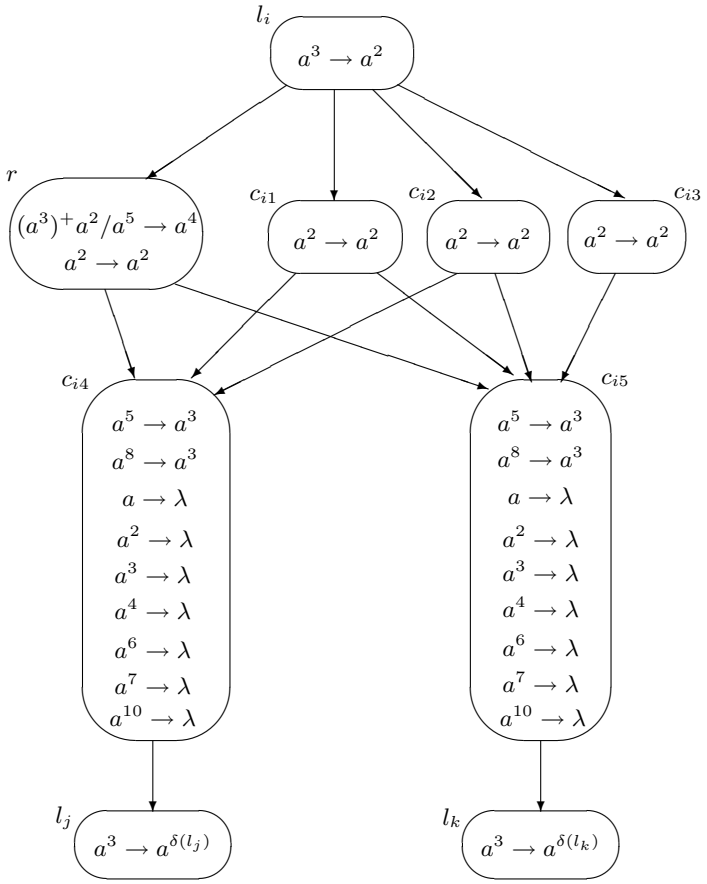

Fig. 13. Module SUB (simulating $l_i : (\texttt{SUB}(r), l_j, l_k)$) for machine M_1

The simulation of an ADD instruction is easy, we just add three spikes to the respective neuron; no rule is needed in the neuron – Figure 11. The SUB instructions of machines M_0, M_1 are simulated by modules as in Figures 12 and

13, respectively. Note that the rules for M_0 fire for a content of the neuron σ_r described by the regular expression $(a^3)^+a$ and the rules for M_1 fire for a content of the neuron σ_r described by the regular expression $(a^3)^+a^2$. To this aim we use the rule $a^3 \to a^2$ in σ_{l_i} instead of $a^3 \to a$, while in σ_r we use the rule $(a^3)^+a^2/a^5 \to a^4$ instead of $(a^3)a/a^4 \to a^3$. This ensures the fact that the rules of M_0 are not used instead of those of M_1 or vice versa. In neurons associated with different labels of M_0, M_1 we have to use different rules, depending on the type of instruction simulated, that is why in Figures 11, 12, and 13 we have written again some rules in the form $a^3 \to a^{\delta(l)}$, as in Figures 1 and 2. Specifically, $\delta(l) = 3$ if l labels an ADD instruction, $\delta(l) = 1$ or $\delta(l) = 2$ if l labels a SUB instruction of M_0 or of M_1, respectively, and, as one sees in Figure 10, we also take $\delta(l_{h,0}) = 2$.

With these explanations, the reader can check that the system Π works as requested, hence $L_\lambda(\Pi) = L$. □

Theorem 7. *$RE = L_\lambda SN^eP_*(rule_*, cons_*, prod_*)$.*

In the proof of Lemma 6, if the moments when the output neuron emits no spike are associated with the symbol b_0, then the generated strings will be shuffled with occurrences of b_0. Therefore, L is a projection of the generated language.

Corollary 5. *Every language $L \in RE, L \subseteq V^*$, can be written in the form $L = h(L')$ for some $L' \in L_{res}SN^eP_*(rule_*, cons_*, prod_*)$, where h is a projection on $V \cup \{b_0\}$ which removes the symbol b_0.*

7 A Tool-Kit for Handling Languages

In this section we present some constructions for performing operations with languages generated by SN P systems with extended rules.

For instance, starting with two SN P systems Π_1, Π_2, we look for a system Π which generates the language $L_\lambda(\Pi_1) \diamond L_\lambda(\Pi_2)$, where $\diamond$ is a binary operation with languages.

For the union of languages, such a system Π is easy to be constructed (as already done in [2]): we start with the systems Π_1, Π_2 without any spike inside and we consider a module which non-deterministically activates one of these systems, by introducing in their neurons as many spikes as we have in the initial configurations of Π_1 and Π_2.

Not so simple is the case of concatenation, which, however, can be handled as in Figure 14. We start with system Π_1 as it is (with the neurons loaded with the necessary spikes), and with system Π_2 without any spike inside.

We have in Figure 14 three sub-systems/modules with specific tasks to solve. For instance, neurons $\sigma_{d_5}, \sigma_{d_6}, \sigma_{d_7}$ non-deterministically choose a moment when the string generated by system Π_1 is assumed completed. After using rule $a^2 \to a$ in σ_{d_6}, neuron σ_{d_5} fires, this activates neurons $\sigma_{c_1}, \ldots, \sigma_{c_n}$, and these neurons both "flood" neuron σ_{d_4} with $m + 1$ spikes and activate the neurons of system

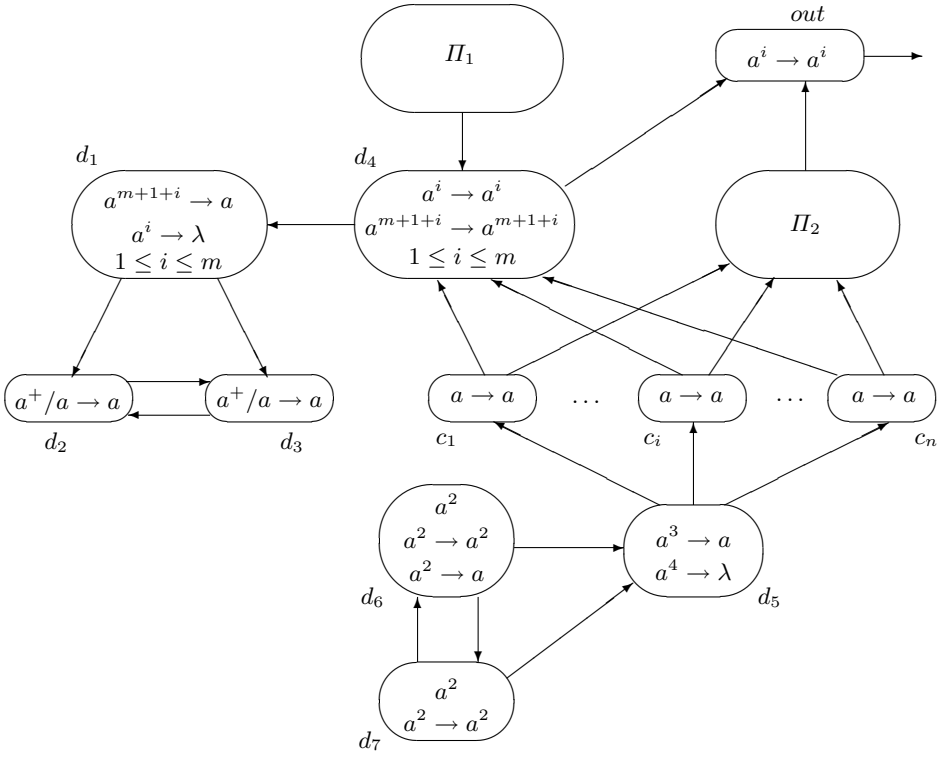

Fig. 14. Computing the concatenation of two languages

Π_2, introducing as many spikes as Π_2 has in its initial configuration. Specifically, we have $n = \max\{m+1, spin(\Pi_2)\}$, where m is the cardinality of the alphabet we work with, and $spin(\Pi_2)$ is the maximum of the number of spikes present in any neuron of Π_2 in the initial configuration. Then we have synapses (c_i, d_4) for $1 \le i \le m+1$, and (c_i, k), for σ_k a neuron in Π_2, for $1 \le i \le n_k$, where n_k is the number of spikes present in σ_k in the initial configuration of Π_2.

The pair of neurons $\sigma_{d_4}, \sigma_{out}$ takes care of the output of the whole system, first passing the output of Π_1 to σ_{out} and then taking the output of Π_2 and sending it out. If σ_{d_4} receives any further spike from Π_1 after neurons $\sigma_{d_5}, \sigma_{d_6}, \sigma_{d_7}$ have "decided" that the work of Π_1 is finished, then σ_{d_4} fires (note that it cannot fire for exactly $m+1$ spikes), this makes σ_{d_1} fire, and then the computation will never finish, because of the pair of neurons $\sigma_{d_2}, \sigma_{d_3}$.

Thus, the computation ends if and only if after sending out a complete string generated by Π_1 we also send out a string generated by Π_2, hence we generate the concatenation of strings produced by the two systems.

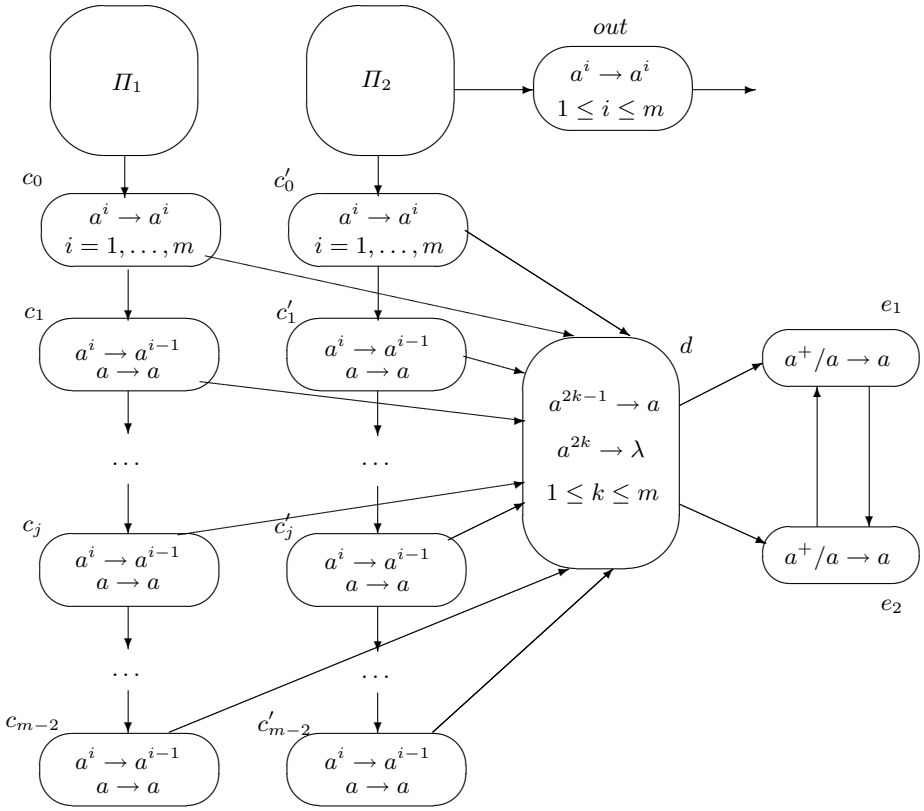

Fig. 15. Computing the intersection with a regular language

Consider now an arbitrary SN P system Π_1 and an SN P system Π_2 simulating a regular grammar G, like in Figure 7, with the following changes: chain rules $A_i \to A_i$ are added to grammar G for all nonterminals A_i; then, we assume that the number of rules (n in the construction) is strictly bigger than the number of symbols (m) – if this is not the case, then we simply duplicate some rules. The system looks now as in Figure 16 (k can be 0 only for chain rules $A_i \to A_i$, where $b_0 = \lambda$). Thus, after simulating a rule $A_i \to b_k$, neurons σ_1, σ_2 are "flooded" and have to stop. The grammar G – and hence also Π_2 – outputs a terminal symbol after an arbitrary number of steps of using chain rules $A_i \to A_i$, hence steps when nothing exits the system. This makes possible the synchronization of Π_1 and Π_2 in the sense that they output spikes in the same steps. What remains to do is to compare the number of spikes emitted by the two systems, so that we can select the strings from the intersection $L_\lambda(\Pi_1) \cap L_\lambda(\Pi_2)$.

This is ensured as suggested in Figure 15 (in order to keep the figure smaller, we have not indicated the range of parameter i, but it is as follows: in all neurons

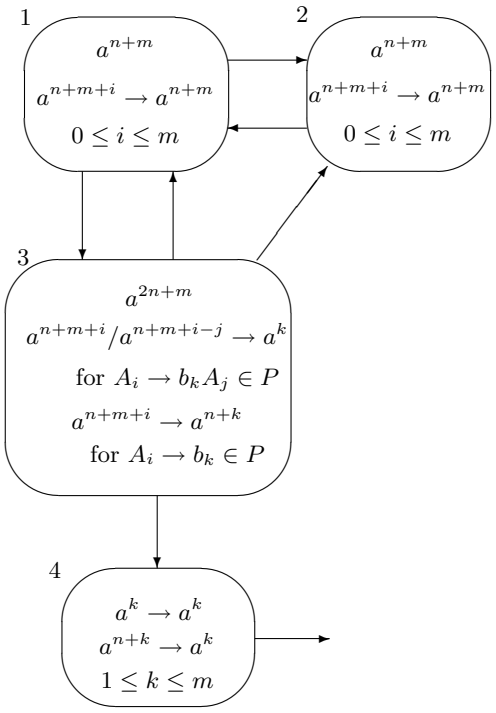

Fig. 16. Simulating a regular grammars having chain rules

σ_{c_j} and $\sigma_{c'_j}$, $1 \le j \le m-2$, we have $2 \le i \le m$). If the two systems Π_1 and Π_2 do not spike at the same time or one sends out $r \ge 1$ spikes and the other one $s \ge 1$ spikes for $r \ne s$, then the neurons $\sigma_{e_1}, \sigma_{e_2}$ will get activated and the computation never stops: the spikes emitted by the two systems circulate from top down along the chains of neurons $\sigma_{c_0}, \sigma_{c_1}, \ldots, \sigma_{c_{m-2}}$ and $\sigma_{c'_0}, \sigma_{c'_1}, \ldots, \sigma_{c'_{m-2}}$, and if we do not obtain exactly one spike at the same time in the two columns, then the neuron σ_d fires and activates the neurons $\sigma_{e_1}, \sigma_{e_2}$.

We do not know how to compute – in an elegant way – morphisms, but the particular case of weak codings can be handled as in Figure 17. The difficulty is to have $h(b_i) = b_j$ with $i < j$, and to this aim the "spike supplier" pair of neurons $\sigma_{c_1}, \sigma_{c_2}$ is considered. In each step, they send $m+1$ spikes to neuron σ_{c_3}. If this neuron receives nothing at the same time from the system Π, then the $m+1$ spikes are forgotten. If i spikes come from system Π, $1 \le i \le m$, then, using the $m+1+i$ spikes, neuron σ_{c_3} can send $j+1$ spikes to the output neuron, which emits the right number of spikes to the environment.

At any time, the neurons $\sigma_{c_1}, \sigma_{c_2}$ can stop their work; if this happens prematurely (before having the system Π halted), then neuron σ_{c_3} will emit only

one spike, and this triggers the "never halting module", composed of the neurons $\sigma_{c_4}, \sigma_{c_5}, \sigma_{c_6}$, which will continue to work forever.

The reader can check that the system produces indeed the language $h(L_\lambda(\Pi))$, for a weak coding h which moves some b_i into b_k, and erases other symbols b_j.

We do not know how to compute arbitrary morphisms or the other AFL operations, Kleene + and inverse morphisms.

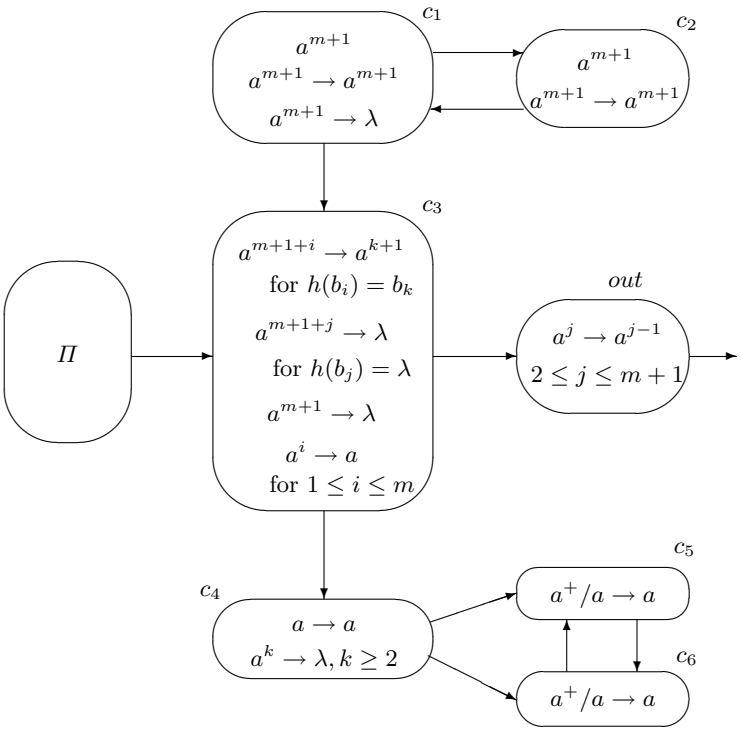

Fig. 17. Computing a weak coding

A possible way to address these problems is to reduce them to another problem, that of introducing delays of arbitrarily many steps in between any two steps of computations in an arbitrary SN P system Π (in the same way as the chain rules introduce such "dummy steps" in the work of a regular grammar). If such a slowing-down of a system would be possible, then we can both compute arbitrary morphisms and the intersection of languages generated by two arbitrary SN P systems (not only with one of them generating a regular language, as above).

Another open problem of interest (but difficult, we believe) is to find an SN P system, as small as possible in the number of neurons, generating a Dyck language

(over at least two pairs of parentheses). If such a systems would be found, then a representation of context-free languages would be obtained, using the Chomsky-Schützenberger characterization of these languages as the weak coding of the intersection of a Dyck language with a regular language.

8 Final Remarks

We have investigated here the power of SN P systems with extended rules (rules allowing to introduce several spikes at the same time) both as number generators and as language generators. In the first case we have provided a simpler proof of a known universality result, in the latter case we have proved characterizations of finite and recursively enumerable languages, and representations of regular languages.

Finding characterizations (or at least representations) of other families of languages from Chomsky hierarchy and Lindenmayer area remains as a research topic. It is also of interest to investigate the possible hierarchy on the number of neurons, extending the result from Corollary 4.

Acknowledgements

The work of the first author was supported by the National Natural Science Foundation of China under Grants numbers 60573013 and 60421001. The work of the last two authors was supported by Project TIN2005-09345-C04-01 of the Ministry of Education and Science of Spain, cofinanced by FEDER funds.

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. M. Ionescu, Gh. Păun, T. Yokomori: Spiking neural P systems. *Fundamenta Informaticae*, 71, 2-3 (2006), 279–308.
3. M. Minsky: *Computation – Finite and Infinite Machines*. Prentice Hall, Englewood Cliffs, NJ, 1967.
4. A. Păun, Gh. Păun: Small universal spiking neural P systems. In volume II of the present proceedings.
5. 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 [8]).
6. G. Rozenberg, A. Salomaa, eds.: *Handbook of Formal Languages*, 3 volumes. Springer-Verlag, Berlin, 1997.
7. A. Salomaa: *Formal Languages*. Academic Press, New York, 1973.
8. The P Systems Web Page: `http://psystems.disco.unimib.it`.