

Monodirectional P systems

Alberto Leporati · Luca Manzoni · Giancarlo Mauri
Antonio E. Porreca · Claudio Zandron

the date of receipt and acceptance should be inserted later

Abstract We investigate the influence that the flow of information in membrane systems has on their computational complexity. In particular, we analyse the behaviour of P systems with active membranes where communication only happens from a membrane towards its parent, and never in the opposite direction. We prove that these “monodirectional P systems” are, when working in polynomial time and under standard complexity-theoretic assumptions, much less powerful than unrestricted ones: indeed, they characterise classes of problems defined by polynomial-time Turing machines with **NP** oracles, rather than the whole class **PSPACE** of problems solvable in polynomial space.

1 Introduction

P systems with active membranes working in polynomial time are known to be able to solve all **PSPACE**-complete problems (Alhazov et al 2003); this exploits membrane structures of polynomial depth and a bidirectional flow of information (in terms of moving objects or changing charges), both from a parent membrane to its children, and in the opposite direction.

When restricting the depth of the membrane structures of a family of P systems to a constant amount, it is still possible to solve problems in the counting hierarchy **CH**, defined in terms of polynomial-time Turing machines with oracles for counting problems (Leporati et al 2015). In the proof of this result, it has been noticed that send-in communication

This work was partially supported by Università degli Studi di Milano-Bicocca, FA 2014: “Complessità computazionale nei sistemi a membrane”.

Dipartimento di Informatica, Sistemistica e Comunicazione
Università degli Studi di Milano-Bicocca
Viale Sarca 336/14, 20126 Milano, Italy
E-mail: {leporati,luca.manzoni,mauri.porreca,zandron}@disco.unimib.it

rules of the form $a []_h^\alpha \rightarrow [b]_h^\beta$ allow us to check whether the amount of objects located in a membrane exceeds a (possibly exponential) threshold in polynomial time.

It is then natural to ask whether that feature is actually necessary in order to obtain the power of counting in polynomial time. In this paper we prove (under the standard complexity-theoretic assumption that $\mathbf{P}^{\mathbf{NP}} \neq \mathbf{P}^{\#\mathbf{P}}$) that this is actually the case: P systems with monodirectional communication, where the information flows only towards the outermost membrane, are limited to $\mathbf{P}^{\mathbf{NP}}$, the class of problems efficiently solved by Turing machines with **NP** oracles. This happens even when allowing polynomially deep membrane structures, a weak form of non-elementary membrane division, or dissolution (which, in this case, turns out to be as powerful as weak non-elementary division). The $\mathbf{P}^{\mathbf{NP}}$ upper bound is actually reached when dissolution or weak non-elementary division are allowed; if neither is available, then the computation power decreases to $\mathbf{P}_{||}^{\mathbf{NP}}$, where the queries must all be fixed in advance, rather than asked adaptively.

Let $\mathcal{M} = \mathcal{AM}(-i, -n, +wn)$ be the class of *monodirectional P systems with active membranes*, without send-in rules; we also remove the usual (“strong”) non-elementary division rules, of the form

$$[[]_{h_1}^+ \cdots []_{h_m}^+ []_{h_{m+1}}^- \cdots []_{h_n}^-]_h^\alpha \rightarrow [[]_{h_1}^\delta \cdots []_{h_m}^\delta]_h^\beta [[]_{h_{m+1}}^\zeta \cdots []_{h_n}^\zeta]_h^\gamma$$

since they also provide a way for membrane h to share information with its children by changing their charge. We replace these rules by “weak” non-elementary division rules (Zandron et al 2008) of the form $[a]_h^\alpha \rightarrow [b]_h^\beta [c]_h^\gamma$, which allow the creation of complex membrane structures (such as complete binary trees) without exchanging information with the children membranes.

Let $\mathcal{M}(-d)$, $\mathcal{M}(-wn)$, and $\mathcal{M}(-d, -wn)$ denote monodirectional P systems without dissolution, without weak non-elementary division, and without both kinds of rules, respec-

tively. For each class \mathcal{D} of P systems, let $\mathbf{PMC}_{\mathcal{D}}$ and $\mathbf{PMC}_{\mathcal{D}}^*$ be the classes of problems solvable by uniform and semi-uniform families of P systems of class \mathcal{D} . Then, the main results of this paper can be summarised as follows:

- The whole class $\mathbf{PMC}_{\mathcal{M}}^{[*]}$, as well as $\mathbf{PMC}_{\mathcal{M}(-d)}^{[*]}$ and $\mathbf{PMC}_{\mathcal{M}(-wn)}^{[*]}$, are equivalent to $\mathbf{P}^{\mathbf{NP}}$. Here $[*]$ denotes optional semi-uniformity.
- The class $\mathbf{PMC}_{\mathcal{M}(-d,-wn)}^{[*]}$ is equivalent to $\mathbf{P}_{\parallel}^{\mathbf{NP}}$.

The rest of the paper is structured as follows: in Section 2 we recall the technical notions related to P systems with active membranes and their complexity classes; in Section 3 we prove some basic limitations of monodirectional P systems; in Section 4 we exploit these results to prove upper bounds to the complexity classes for monodirectional P systems; in Section 5 we provide the corresponding lower bounds by simulating Turing machines with NP oracles; in Section 6 some results of the preceding sections are improved; finally, in Section 7 we present some open problems and directions for future research.

2 Basic notions

We begin by formally defining monodirectional P systems as a variant of P systems with active membranes (Păun 2001).

Definition 1 A *monodirectional P system with active membranes* of initial degree $d \geq 1$ is a tuple

$$\Pi = (\Gamma, \Lambda, \mu, w_{h_1}, \dots, w_{h_d}, R)$$

where:

- Γ is an alphabet, i.e., a finite non-empty set of symbols, usually called *objects*;
- Λ is a finite set of labels for the membranes;
- μ is a membrane structure (i.e., a rooted *unordered* tree, usually represented by nested brackets) consisting of d membranes labelled by elements of Λ in a one-to-one way;
- w_{h_1}, \dots, w_{h_d} , with $h_1, \dots, h_d \in \Lambda$, are strings over Γ , describing the initial multisets of objects placed in the d regions of μ ;
- R is a finite set of rules.

Each membrane possesses, besides its label and position in μ , another attribute called *electrical charge*, which can be either neutral (0), positive (+) or negative (−) and is always neutral before the beginning of the computation.

The rules in R are of the following types:

- (a) *Object evolution rules*, of the form $[a \rightarrow w]_h^\alpha$
They can be applied inside a membrane labelled by h , having charge α and containing an occurrence of the

object a ; the object a is rewritten into the multiset w (i.e., a is removed from the multiset in h and replaced by the objects in w).

- (c) *Send-out communication rules*, of the form $[a]_h^\alpha \rightarrow []_h^\beta b$
They can be applied to a membrane labelled by h , having charge α and containing an occurrence of the object a ; the object a is sent out from h to the outside region becoming b and, simultaneously, the charge of h becomes β .
- (d) *Dissolution rules*, of the form $[a]_h^\alpha \rightarrow b$
They can be applied to a membrane labelled by h , having charge α and containing an occurrence of the object a ; the membrane is dissolved and its contents are left in the surrounding region unaltered, except that an occurrence of a becomes b .
- (e) *Elementary division rules*, of the form $[a]_h^\alpha \rightarrow [b]_h^\beta [c]_h^\gamma$
They can be applied to a membrane labelled by h , having charge α , containing an occurrence of the object a but having no other membrane inside (an *elementary membrane*); the membrane is divided into two membranes having label h and charges β and γ ; the object a is replaced, respectively, by b and c , while the other objects of the multiset are replicated in both membranes.
- (f*) *Weak non-elementary division rules* (Zandron et al 2008), of the form $[a]_h^\alpha \rightarrow [b]_h^\beta [c]_h^\gamma$
They can be applied to a membrane labelled by h , having charge α , and containing an occurrence of the object a , even if it contains further membranes; the membrane is divided into two membranes having label h and charges β and γ ; the object a is replaced, respectively, by b and c , while the rest of the contents (including whole membrane substructures) is replicated in both membranes.

The instantaneous *configuration* of a membrane consists of its label h , its charge α , and the multiset w of objects it contains at a given time. It is denoted by $[w]_h^\alpha$. The (*full*) *configuration* \mathcal{C} of a P system Π at a given time is a rooted, unordered tree. The root is a node corresponding to the external environment of Π , and has a single subtree corresponding to the current membrane structure of Π . Furthermore, the root is labelled by the multiset located in the environment, and the remaining nodes by the configurations $[w]_h^\alpha$ of the corresponding membranes.

A computation step changes the current configuration according to the following set of principles:

- Each object and membrane can be subject to at most one rule per step, except for object evolution rules: inside each membrane, several evolution rules can be applied simultaneously.
- The application of rules is *maximally parallel*: each object appearing on the left-hand side of evolution, communication, dissolution or division rules must be subject to exactly one of them (unless the current charge of the membrane prohibits it). Analogously, each membrane

can only be subject to one communication, dissolution, or division rule (types (c)–(f')) per computation step. In other words, the only objects and membranes that do not evolve are those associated with no rule, or only to rules that are not applicable due to the electrical charges.

- When several conflicting rules can be applied at the same time, a nondeterministic choice is performed; this implies that, in general, multiple possible configurations can be reached after a computation step.
- In each computation step, all the chosen rules are applied simultaneously (in an atomic way). However, in order to clarify the operational semantics, each computation step is conventionally described as a sequence of micro-steps as follows. First, all evolution rules are applied inside the elementary membranes, followed by all communication, dissolution and division rules involving the membranes themselves; this process is then repeated to the membranes containing them, and so on towards the root (outermost membrane). In other words, the membranes evolve only after their internal configuration has been updated. For instance, before a membrane division occurs, all chosen object evolution rules must be applied inside it; this way, the objects that are duplicated during the division are already the final ones.
- The outermost membrane cannot be divided or dissolved, and any object sent out from it cannot re-enter the system again.

A *halting computation* of the P system Π is a finite sequence $\mathcal{C} = (\mathcal{C}_0, \dots, \mathcal{C}_k)$ of configurations, where \mathcal{C}_0 is the initial configuration, every \mathcal{C}_{i+1} is reachable from \mathcal{C}_i via a single computation step, and no rules of Π are applicable in \mathcal{C}_k . A *non-halting computation* $\mathcal{C} = (\mathcal{C}_i : i \in \mathbb{N})$ consists of infinitely many configurations, again starting from the initial one and generated by successive computation steps, where the applicable rules are never exhausted.

P systems can be used as language *recognisers* by employing two distinguished objects *yes* and *no*: we assume that all computations are halting, and that either object *yes* or object *no* (but not both) is sent out from the outermost membrane, and only in the last computation step, in order to signal acceptance or rejection, respectively. If all computations starting from the same initial configuration are accepting, or all are rejecting, the P system is said to be *confluent*. If this is not necessarily the case, then we have a *non-confluent* P system, and the overall result is established as for nondeterministic Turing machines: it is acceptance iff an accepting computation exists. All P systems in this paper are assumed to be confluent.

In order to solve decision problems (or, equivalently, decide languages), we use *families* of recogniser P systems $\Pi = \{\Pi_x : x \in \Sigma^*\}$. Each input x is associated with a P system Π_x deciding the membership of x in a language $L \subseteq \Sigma^*$ by accepting or rejecting. The mapping $x \mapsto \Pi_x$ must be effi-

ciently computable for inputs of any length, as discussed in detail in Murphy and Woods (2011).

Definition 2 A family of P systems $\Pi = \{\Pi_x : x \in \Sigma^*\}$ is (*polynomial-time*) *uniform* if the mapping $x \mapsto \Pi_x$ can be computed by two polynomial-time deterministic Turing machines E and F as follows:

- $F(1^n) = \Pi_n$, where n is the length of the input x and Π_n is a common P system for all inputs of length n with a distinguished input membrane.
- $E(x) = w_x$, where w_x is a multiset encoding the specific input x .
- Finally, Π_x is simply Π_n with w_x added to its input membrane.

The family Π is said to be (*polynomial-time*) *semi-uniform* if there exists a single deterministic polynomial-time Turing machine H such that $H(x) = \Pi_x$ for each $x \in \Sigma^*$.

Any explicit encoding of Π_x is allowed as output of the construction, as long as the number of membranes and objects represented by it does not exceed the length of the whole description, and the rules are listed one by one. This restriction is enforced in order to mimic a (hypothetical) realistic process of construction of the P systems, where membranes and objects are presumably placed in a constant amount during each construction step, and require actual physical space proportional to their number; see also Murphy and Woods (2011) for further details on the encoding of P systems.

The classes of decision problems solved by uniform and semi-uniform families of monodirectional P systems working in polynomial time are denoted by $\mathbf{PMC}_{\mathcal{M}}$ and $\mathbf{PMC}_{\mathcal{M}}^*$, respectively. When dissolution, weak non-elementary division, or both kinds of rules are not allowed, the corresponding classes are denoted by $\mathbf{PMC}_{\mathcal{M}(-d)}^{[*]}$, $\mathbf{PMC}_{\mathcal{M}(-wn)}^{[*]}$, and $\mathbf{PMC}_{\mathcal{M}(-d,-wn)}^{[*]}$, respectively, where $[*]$ denotes optional semi-uniformity.

Chapter 17 of Papadimitriou's book (1993) provides the technical details on the complexity classes $\mathbf{P}^{\mathbf{NP}}$ and $\mathbf{P}_{\parallel}^{\mathbf{NP}}$, defined in terms of Turing machines with **NP** oracles.

3 Properties of monodirectional P systems

The first results that we prove show how the lack of inbound communication substantially restricts the range of behaviours exhibited during the computations of monodirectional P systems.

Definition 3 Let Π be a P system, and let \mathcal{C} and \mathcal{D} be configurations of Π . We say that \mathcal{D} is a *restriction* of \mathcal{C} , in symbols $\mathcal{D} \sqsubseteq \mathcal{C}$, if the membrane structures of the two configurations are identical (i.e., they have the same shape, labelling, and charges) and each multiset of objects of \mathcal{D} is a submultiset of that located in the corresponding region of \mathcal{C} .

The following proposition shows that, while a recogniser P system working in time t might create exponentially many objects per region during its computation, only a polynomial amount (with respect to t) of them in each region does actually play a useful role if the system is monodirectional: indeed, the final result of the computation can be identified by just keeping track of a number of objects per region equal to the number of steps yet to be carried out.

Lemma 1 *Let Π be a monodirectional recogniser P system, and let $\mathcal{C} = (\mathcal{C}_0, \dots, \mathcal{C}_t)$, with $t \geq 1$, be a halting computation of Π . Then, there exists a sequence of configurations $(\mathcal{D}_0, \dots, \mathcal{D}_t)$ such that*

- (i) *we have $\mathcal{D}_i \sqsubseteq \mathcal{C}_i$ for $0 \leq i \leq t$, and each multiset of \mathcal{D}_i has at most $t - i$ objects;*
- (ii) *for all $i < t$ there exists a configuration \mathcal{E}_{i+1} such that \mathcal{E}_{i+1} is reachable in one step from \mathcal{D}_i ($\mathcal{D}_i \rightarrow \mathcal{E}_{i+1}$ for brevity) and $\mathcal{D}_{i+1} \sqsubseteq \mathcal{E}_{i+1}$;*
- (iii) *a send-out rule of the form $[a]_h^\alpha \rightarrow []_h^\beta$ yes (resp., $[a]_h^\alpha \rightarrow []_h^\beta$ no) is applied to the outermost membrane during the transition step $\mathcal{D}_{t-1} \rightarrow \mathcal{E}_t$ if and only if \mathcal{C} is an accepting (resp., rejecting) computation.*

Proof. By induction on t . If $t = 1$, then the environment of \mathcal{C}_1 contains *yes* or *no*, which have been sent out during the computation step $\mathcal{C}_0 \rightarrow \mathcal{C}_1$ by a rule $[a]_h^\alpha \rightarrow []_h^\beta$ *yes* or $[a]_h^\alpha \rightarrow []_h^\beta$ *no*. Let $\mathcal{D}_0 \sqsubseteq \mathcal{C}_0$ be obtained by keeping only the objects on the left-hand side of send-out, dissolution, and division rules applied during $\mathcal{C}_0 \rightarrow \mathcal{C}_1$ (we call these rules “blocking”, since at most one of them can be applied inside each membrane at each step). At most one object per region is kept, given the lack of send-in rules. Let $\mathcal{D}_1 \sqsubseteq \mathcal{C}_1$ be obtained by deleting all objects. Then:

- (i) we have $\mathcal{D}_0 \sqsubseteq \mathcal{C}_0$ and $\mathcal{D}_1 \sqsubseteq \mathcal{C}_1$ by construction, and all multisets of \mathcal{D}_0 and \mathcal{D}_1 have at most 1 and exactly 0 objects, respectively;
- (ii) let the transition $\mathcal{D}_0 \rightarrow \mathcal{E}_1$ be computed by applying all blocking rules applied during the step $\mathcal{C}_0 \rightarrow \mathcal{C}_1$, which are all enabled by construction; then $\mathcal{E}_1 \sqsubseteq \mathcal{C}_1$ and, since $\mathcal{D}_1 \sqsubseteq \mathcal{C}_1$ and \mathcal{D}_1 contains no objects, necessarily $\mathcal{D}_1 \sqsubseteq \mathcal{E}_1$;
- (iii) the computation \mathcal{C} is accepting if and only if the rule $[a]_h^\alpha \rightarrow []_h^\beta$ *yes* is applied from \mathcal{C}_0 , and the latter is equivalent by construction to that rule being applicable from \mathcal{D}_0 (the reasoning is similar if \mathcal{C} is rejecting).

This proves the base case. Now let $\mathcal{C} = (\mathcal{C}_{-1}, \mathcal{C}_0, \dots, \mathcal{C}_t)$ be a halting computation of length $t + 1$. The sub-computation $(\mathcal{C}_0, \dots, \mathcal{C}_t)$ is also halting, and by induction hypothesis there exists a sequence of configurations $(\mathcal{D}_0, \dots, \mathcal{D}_t)$ satisfying properties (i)–(iii). Construct the configuration \mathcal{D}_{-1} as follows: first of all, keep all objects from \mathcal{C}_{-1} that appear on

the left-hand side of blocking rules applied during the computation step $\mathcal{C}_{-1} \rightarrow \mathcal{C}_0$; this requires at most one object per region, and guarantees that the membrane structure’s shape and charges can be updated correctly (i.e., the same as \mathcal{C}_0 and \mathcal{D}_0).

We must also ensure that all objects of \mathcal{D}_0 can be generated from \mathcal{D}_{-1} during the transition $\mathcal{D}_{-1} \rightarrow \mathcal{E}_0$. Once the blocking rules to be applied have been chosen, any object a located inside a membrane of \mathcal{D}_0 can be traced back to a single object in \mathcal{D}_{-1} . Either a appears on the right-hand side of one of those blocking rules, or it appears on the right-hand side of an object evolution rule applied in the step $\mathcal{C}_{-1} \rightarrow \mathcal{C}_0$, or it does not appear explicitly in any rule applied in that step; in the latter case, it is either carried on unchanged from \mathcal{D}_{-1} (possibly from another region, if membrane dissolution occurred), or is created by duplicating the content of a membrane by applying a division rule (triggered by a different object). As a consequence, at most t objects per region of \mathcal{D}_{-1} , possibly in conjunction with a single object per region involved in blocking rules, suffice in order to generate the t objects per region of \mathcal{D}_0 . As a consequence,

- (i) we have $\mathcal{D}_{-1} \sqsubseteq \mathcal{C}_{-1}$ by construction, and \mathcal{D}_{-1} contains at most $t + 1$ objects per region;
- (ii) by applying all blocking rules and as many evolution rules as possible from the computation step $\mathcal{C}_{-1} \rightarrow \mathcal{C}_0$ in \mathcal{D}_{-1} , we obtain a configuration \mathcal{E}_0 with the same membrane structure as \mathcal{D}_0 and, as mentioned above, containing all objects from \mathcal{D}_0 (and possibly other objects generated by evolution rules).

Since (iii) holds by induction hypothesis, this completes the proof. \square

Notice that this lemma does not give us an efficient algorithm for choosing *which* objects are important for each step of the computation; it only proves that a small (i.e., polynomial-sized) multiset per region exists. However, it is easy to find such an algorithm by slightly relaxing the conditions: instead of limiting the cardinality of the multisets to $t - i$, we limit the number of occurrences of *each symbol* to that value, and simply delete the occurrences in excess separately for each symbol. This gives us the larger cardinality bound $|\Gamma| \cdot (t - i)$ per region, which is polynomial whenever the number of computation steps of the system is, and still allows us to simulate the overall behaviour of the P system.

Lemma 1 fails for P systems with send-in rules because some configurations where each multiset is small nonetheless require a previous configuration with a region containing exponentially many objects. This is the case, for instance, for P systems solving counting problems, where the number of assignments satisfying a Boolean formula is checked against a threshold by means of send-in rules (Leporati et al 2015). Those assignments are represented in the P system by a potentially exponential number of objects located in the

same region, which are sent into exponentially many children membranes in parallel (i.e., at most one object enters each child membrane), and cannot always be reduced to a polynomial amount without changing the accepting behaviour of the P system.

Another property of monodirectional P systems is the existence of computations where membranes having the same labels always have children (and, recursively, all the descendants) with the same configuration. This property will be useful when simulating confluent recogniser monodirectional P systems in Section 4.

Lemma 2 *Let Π be a monodirectional P system. Then there exists a computation $C = (C_0, \dots, C_t)$ of Π where, in each configuration C_i , the following holds: any two subconfigurations¹ of C_i having membranes with the same label as roots are identical, except possibly for the multiset and charge of the root membranes themselves.*

Proof. By induction on i . The statement trivially holds for the initial configuration of Π , since the membranes are injectively labelled.

When a division rule is applied to a membrane h , two subconfigurations with root h are created; this is the only way to generate multiple membranes sharing the same label. The two resulting subconfigurations may only differ with respect to the contents and charges of the root membranes, since the internal membranes have evolved before the division of h occurs: recall that the rules are applied, from a logical standpoint, in a bottom-up way (Păun 2001).

On the other hand, if two subconfigurations with identically labelled root membranes already exist in a configuration C_i , then we can assume that the property holds by induction hypothesis. We can then nondeterministically choose which rules to apply in the subconfiguration having the first membrane as root, excluding the root itself; since the other subconfiguration is identical (except possibly for the root), the same multiset of rules can also be applied to it, thus preserving the property in the next configuration of the system. \square

While Lemma 2 somehow “compresses” each level of the configuration of monodirectional P systems, it does not, however, reduce the number of distinct membranes per level to a polynomial number. Indeed, the standard membrane computing technique of generating all (exponentially many) possible assignments to a set of variables does not require send-in rules (Zandron et al 2001), and can be carried out in parallel on all levels of the membrane structure.

Lemma 2 also fails for P systems with send-in rules. The reason is that two identical subconfigurations can be

¹ We define a subconfiguration of C_i as a subtree (a root node together with all its descendants) of the membrane structure of C_i , including labels, multisets, and charges of the membranes.

made different by having a single object located immediately outside, and nondeterministically sending it into one of the root membranes of the two subtrees; the evolution of the two branches of the system might then diverge completely.

4 Simulation of monodirectional P systems

It is a well-known result in membrane computing that P systems with active membranes can be simulated in polynomial time by deterministic Turing machines if no membrane division rules are allowed (Zandron et al 2001). More specifically, the portion of the system that is not subject to membrane division can be simulated deterministically with a polynomial slowdown, while the output of the dividing membranes can be obtained by querying an appropriate oracle. It was recently proved that, for standard (bidirectional) P systems where only elementary membranes can divide, an oracle for a #P function is necessary and sufficient (Leporati et al 2014).

In what follows we prove that an NP-oracle is sufficient for the simulation of monodirectional P systems. In particular, the oracle will solve the following problem.

Lemma 3 *Given the initial configuration of an elementary membrane with label h of a monodirectional P system (possibly without dissolution rules), an object type $a \in \Gamma$, and two integers $k, t \in \mathbb{N}$ in unary notation, it is NP-complete to decide whether the set of membranes with label h existing at time t emits (via send-out or dissolution rules) at least k copies of object a at that time step.*

Proof. The problem is NP-hard, since it is possible to simulate an arbitrary polynomial-time, nondeterministic Turing machine M by using a single membrane with elementary division (without using send-in rules) and obtain the same result as M by checking if the resulting membranes send out at least one ($k = 1$) “acceptance object” at a specific time step (Leporati et al 2015). Clearly, the problem remains NP-hard if we allow dissolution rules in addition to send-out rules.

Conversely, the problem can be solved by a nondeterministic, polynomial-time Turing machine M as follows. Simulate t computation steps of the membrane explicitly, by keeping track of its charge and multiset, as in any standard simulation (Zandron et al 2001). If the membrane divides, then M keeps track of all the resulting membranes, *until the number exceeds k* . If that happens, then k copies of the membrane are chosen nondeterministically among those being simulated (which are at most $2k$ after any simulated step, if all membranes divide), and the remaining ones are discarded. Since there is no incoming communication, any instance of the membrane can be simulated correctly, as its behaviour does not depend on the behaviour of its siblings. If one of the simulated membranes dissolves before t steps, one of

the k “slots” is released and can be reused in case of a further membrane division.

After having simulated t steps as described, the machine M accepts if and only if at least k copies of a are emitted (sent out, or released by dissolution) in the last step by the membranes being simulated. At most k membranes need to be simulated in order to check whether at least k copies of the object are emitted and, by exploiting nondeterminism, we are guaranteed that the correct subset of membranes is chosen by at least one computation of M . Since k and t are polynomial with respect to the size of the input, the result follows. \square

The values of t and k are given in unary since, otherwise, the number of steps or the number of membranes to simulate could be exponential with respect to the size of the input, and the problem would not be solvable in polynomial time.

As a consequence of Lemma 3, monodirectional P systems without non-elementary division can be simulated in polynomial time with access to an NP oracle.

Theorem 1 $\text{PMC}_{\mathcal{M}(-\text{wn})}^* \subseteq \mathbf{P}^{\text{NP}}$.

Proof. The rules applied to non-elementary membranes (i.e., non-dividing membranes) can be simulated directly in deterministic polynomial time by a Turing machine M (Leporati et al 2014); this includes the outermost membrane, which ultimately sends out the result object. In order to update the configurations of the non-elementary membranes correctly, the objects emitted from elementary membranes (which potentially divide) have to be added to their multisets.

Suppose the P systems of the family being simulated work in polynomial time $p(n)$. By Lemma 1, the final result of the computation can be correctly determined by keeping track of at most $p(n)$ copies of each object per region. Hence, we can update the configurations by using an oracle for the problem of Lemma 3. At time step t , we make multiple queries for each label h of an elementary membrane and for each object type $a \in \Gamma$: by performing a binary search on k over the range $[0, p(n)]$, we can find the exact number of copies of a emitted by membranes with label h at time t , or discover that this number is at least $p(n)$ (and, in that case, we only add $p(n)$ objects to the multiset). This completes the proof. \square

Monodirectional P systems without non-elementary division become weaker if dissolution is also disallowed: now a membrane cannot *become* elementary during the computation, and thus the evolution of each dividing membrane is always independent of the rest of the system. This allows us to perform all queries in parallel, rather than sequentially (in an adaptive way).

Theorem 2 $\text{PMC}_{\mathcal{M}(-\text{d}, -\text{wn})}^* \subseteq \mathbf{P}_{\parallel}^{\text{NP}}$.

Proof. If dissolution rules are not allowed, being elementary is a static property of the membranes, i.e., a membrane is elementary for the whole computation if and only if it is elementary in the initial configuration. By observing that each query is completely independent of the others (i.e., each query involves a different membrane, time step and object) and also independent of the configurations of the non-dividing membranes (due to the lack of send-in rules), we can perform them in parallel even before starting to simulate the P system. This proves the inclusion in $\mathbf{P}_{\parallel}^{\text{NP}}$. \square

Now let us consider monodirectional P systems with non-elementary membrane division. For this kind of systems, the behaviour of a dividing membrane is, of course, dependent on the behaviour of its children and, recursively, of all its descendants. In order to simulate the behaviour of the children by using oracles, we define a more general query problem, where we assume that the behaviour of the descendants of the membrane mentioned in the query has already been established.

First of all, notice that the lack of send-in rules allows us to extend the notion of transition step $\mathcal{C} \rightarrow \mathcal{D}$ between configurations to labelled subforests² \mathcal{E} of \mathcal{C} and \mathcal{F} of \mathcal{D} as $\mathcal{E} \rightarrow \mathcal{F}$; the only differences from the standard definition are that \mathcal{E} is not necessarily a single tree, and that its outermost membranes may divide and dissolve.

Definition 4 Let Π be a monodirectional P system, let \mathcal{C} be a configuration of Π , and let $h \in \Lambda$ be a membrane label. A subforest \mathcal{S} of \mathcal{C} is called a *label-subforest induced by h* , or *h -subforest* for brevity, if one of the following conditions hold:

- \mathcal{C} is the initial configuration of Π , and \mathcal{S} consists of a single tree rooted in the (unique) membrane h ,
- \mathcal{C} is a possible configuration of Π at time $t + 1$ with $\mathcal{C}' \rightarrow \mathcal{C}$, and there exists an h -subforest \mathcal{S}' in \mathcal{C}' such that $\mathcal{S}' \rightarrow \mathcal{S}$.

The notion of h -subforest can be viewed as a generalisation of the equivalence classes of membranes in P systems without charges defined by Murphy and Woods (2007).

Lemma 4 Let Π be a monodirectional P system. Then there exists a computation of Π where, at each time step and for each membrane label $h \in \Lambda$, all h -subforests are identical.

Proof. Multiple h -subforests can only be created by division of an ancestor of h ; but then, by Lemma 2, there exists a computation of Π where the resulting h -subforests are identical. \square

² We define a subforest F' of a forest F to be any subgraph such that, whenever F' includes a vertex v , it also includes all the descendants of v .

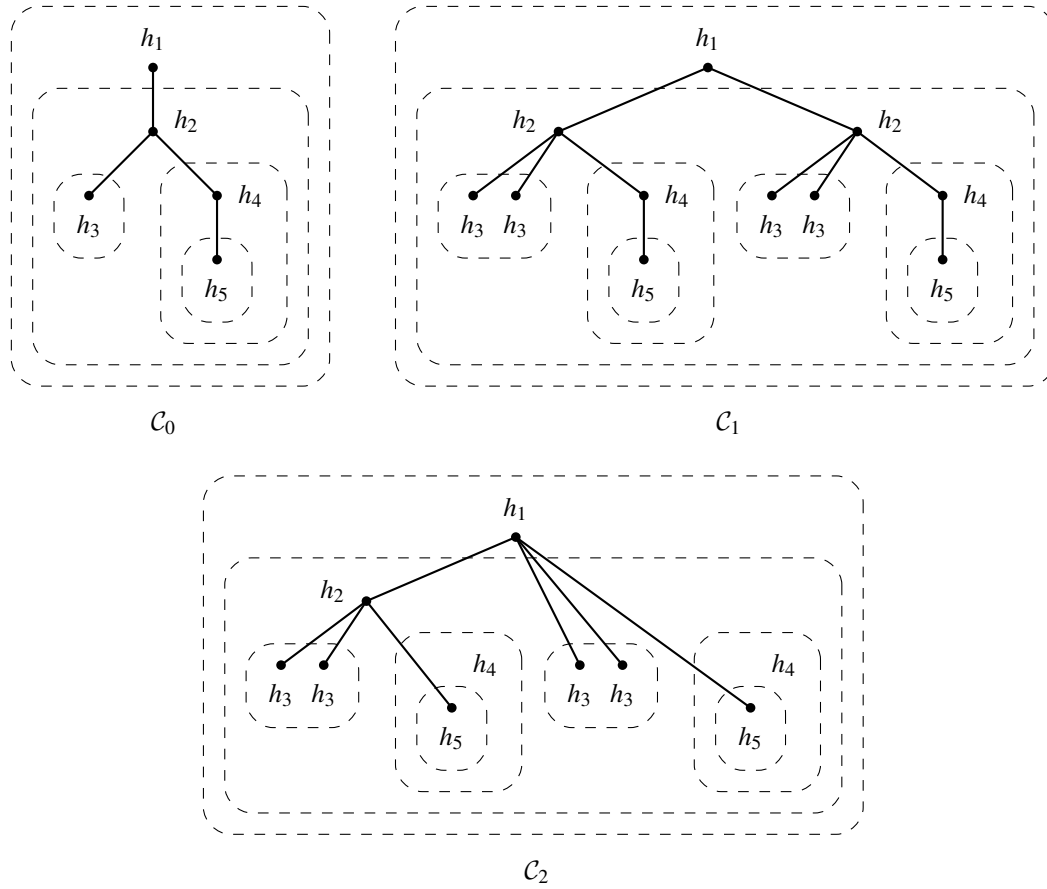


Fig. 1 Evolution of a membrane structure and its label-subforests, which are enclosed by dashed rectangles.

Example 1 Fig. 1 shows the evolution of the membrane structure of a monodirectional P system together with its label-subforests. The label-subforests in the initial configuration \mathcal{C}_0 coincide with all downward-closed subtrees. In the computation step $\mathcal{C}_0 \rightarrow \mathcal{C}_1$ both h_2 and h_3 divide; the division of the former causes the duplication of the h_3 - and h_4 -subforests (and, indirectly, of the h_5 -subforest); the division of an ancestor membrane is the only way to have more than one label-subforest. By Lemma 4, we can always assume that multiple label-subforests induced by the same label are identical. In the computation step $\mathcal{C}_1 \rightarrow \mathcal{C}_2$, the rightmost membrane having label h_2 and both instances of h_4 dissolve. Notice that this does *not* cause the disappearance of the two h_4 -subforests: in the general case, the membranes h_4 might contain label-subforests induced by different labels, and we still need to refer to them as a single entity (that is, the h_4 -subforest), without the need to describe the internal structure, even when h_4 ceases to exist.

As can be observed from Fig. 1, a subforest can be identified as an h -subforest by checking whether it can be generated from the downward-closed subtree rooted in h in the initial configuration.

A computation that ensures that all h -subforests are identical for all $h \in \Lambda$ can be obtained by imposing a total ordering (a priority) on the set of rules of the P system, and applying inside each membrane the rules with higher priority whenever possible. In the following, we assume that a priority order (e.g., the lexicographic order) has been fixed; there is no loss of generality in doing that, since we only focus on *confluent* P systems in this paper. We define the multiset of objects emitted by a label-subforest as the union of the multisets emitted by its outermost membranes.

Lemma 5 *Given the initial configuration of a membrane with label h of a monodirectional P system, an object $a \in \Gamma$, two integers $k, t \in \mathbb{N}$ in unary notation, and a table T of the objects emitted during computation steps $1, \dots, t$ by the label-subforests immediately contained in h , it is **NP-complete** to decide whether each h -subforest emits at least k copies of object a at time t .*

Proof. The problem is **NP-hard**, since the set of elementary membranes with label h of Lemma 3 is an example of an h -subforest; that problem is thus a special case (limited to label-subforests of height 0) of the current one.

To prove membership in **NP** we also use an algorithm similar to the proof of Lemma 3: simulate up to k instances of membrane h , nondeterministically choosing which ones to keep when a membrane division occurs. However, besides simulating the rules directly involving the membranes with label h , we need to update their configuration by adding, at each computation step, the objects emitted by the label-subforests they contain. This is trivial, since the required data is supplied as the input table T . Here we exploit Lemma 4, and simulate a computation where all label-subforests contained in multiple instances of h are identical, and always emit the same objects.

The other main difference from the proof of Lemma 3 is that we do not release one of the k slots when one instance of membrane h dissolves, since its children may still emit objects, and those count in determining the output of the h -subforest. Rather, if an instance of h currently being simulated dissolved during steps $1, \dots, t$, then we add the outputs at time t of the label-subforests immediately contained in h to the result of the computation; those outputs are obtained from table T .

The statement of this lemma then follows from an argument completely analogous to that presented in the proof of Lemma 3: there exists a sequence of nondeterministic choices leading to the simulation of k instances of h sending out at least k objects if and only if at least k objects are actually sent out by the P system being simulated. \square

We can finally show that monodirectional P systems using non-elementary division (and dissolution) also do not exceed the upper bound **P^{NP}**.

Theorem 3 $\text{PMC}_{\mathcal{M}}^* \subseteq \text{P}^{\text{NP}}$.

Proof. We use an algorithm similar to the one described in the proof of Theorem 1. However, instead of using the oracle to compute the output of the elementary membranes, we use it to compute the output of the label-subforests; this, as proved in Lemma 5, is **NP**-complete.

First, we perform all queries involving all h -subforests of height 0, that is, those where h is an elementary membrane. Since these label-subforests do not contain further label-subforests, the corresponding parameter T of the query is always the empty table. By collecting all query results across all meaningful values of parameters h, a, k, t , we obtain a table T_0 of the number of objects emitted during the computation by label-subforests of height 0 (up to a threshold of $p(n)$, corresponding to the computation time of the family of P systems being simulated).

The table T_0 can then be used as parameter to queries involving label-subforests of height 1, which allows us to compute a corresponding table T_1 . In general, the table T_i is used as a parameter for the queries related to label-subforests of height $i + 1$. The procedure is repeated until non-divisible

membranes are reached; these can be simulated directly by using the table obtained by collecting the results of the queries involving the label-subforests immediately contained in them.

Notice that the queries involving h -subforests of a given height can always be asked in parallel (across all values of h, a, k, t); the queries must be asked sequentially only when involving different heights, since the tables T related to inner label-subforests must be computed in advance. \square

5 Simulation of **P^{NP}** machines

In order to prove the converse inclusions between complexity classes, we describe a simulation of any Turing machine M with an **NP** oracle by means of monodirectional P systems, as an adaptation of the simulation by Leporati et al (2015). Let Q be the set of states of M ; we assume, without loss of generality, a binary alphabet $\{0, 1\}$ for M . Finally, we denote by $\delta: Q \times \Sigma \rightarrow Q \times \Sigma \times \{\leftarrow, \triangleright\}$ the transition function of M .

Suppose that the configuration of M at a certain time step is the following: the tape contains the string $x = x_1 \cdots x_m$, the state of the machine is q , and the tape head is located on cell i . This configuration is encoded as a multiset located in a single membrane h of the P system, as follows. There is one object 1_{j-i} for each $1 \leq j \leq m$ such that $x_j = 1$; that is, each 1 in the string x is represented as an object indexed by its position in x , shifted by i ; the 0s of x are not represented by an object, but rather by the absence of the corresponding 1. The object 1_0 (resp., its absence) represents a 1 (resp., a 0) located under the tape head; the indices will be updated (increased or decreased) when simulating a tape head movement. Finally, the state q of M is encoded as an object q with the same name. Further objects, not part of the encoding of the configuration of M , may also appear for simulation purposes.

A transition step of M is simulated by 7 steps of the P system. We assume that the membrane h containing the encoding of the configuration of M also contains the object \ominus .

Step 1. The object \ominus is sent out (as the “junk” object $\#$) in order to change the charge of h to negative:

$$[\ominus]_h^0 \rightarrow []_h^- \# \quad (1)$$

Step 2. When h is negative, the object 1_0 is sent out, if appearing, in order to change the charge to positive. If 1_0 does not appear, the membrane remains negative.

$$[1_0]_h^- \rightarrow []_h^+ \# \quad (2)$$

The remaining tape-objects are primed:

$$[1_i \rightarrow 1'_i]_h^- \quad \text{for } i \neq 0 \quad (3)$$

The state-object q is also primed, and produces the object \odot :

$$[q \rightarrow q' \odot]_h^- \quad (4)$$

Step 3. The system can now observe the charge of h and establish whether 1_0 appeared (i.e., whether the symbol under the tape head was 1) or not (i.e., the symbol was 0); this corresponds to a positive or negative charge, respectively. The object q' is rewritten accordingly:

$$[q' \rightarrow (q, 1)]_h^+ \quad [q' \rightarrow (q, 0)]_h^- \quad (5)$$

At the same time, the neutral charge of h is restored by \odot :

$$[\odot]_h^\alpha \rightarrow []_h^0 \# \quad \text{for } \alpha \in \{+, -\} \quad (6)$$

Step 4. For the sake of example, suppose the transition function of M on state q is defined by $\delta(q, 0) = (r, 1, \triangleright)$ and $\delta(q, 1) = (s, 0, \triangleleft)$; the other cases are similar. The object $(q, 0)$ or $(q, 1)$ is rewritten accordingly:

$$[(q, 0) \rightarrow (r, 1, \triangleright)]_h^0 \quad [(q, 1) \rightarrow (s, 0, \triangleleft)]_h^0 \quad (7)$$

Simultaneously, the tape-objects are primed again:

$$[1'_i \rightarrow 1''_i]_h^0 \quad \text{for } i \neq 0 \quad (8)$$

Step 5. Now the triple generated in the previous step is “unpacked” into its components, which include an object that will be eventually rewritten into the new state-object, the object $1''_0$ (or nothing), and an object to be used to change the charge according to the direction of the movement of the tape head:

$$[(r, 1, \triangleleft) \rightarrow \hat{r} 1''_0 \oplus]_h^0 \quad \text{for } r \in Q \quad (9)$$

$$[(r, 1, \triangleright) \rightarrow \hat{r} 1''_0 \ominus]_h^0 \quad \text{for } r \in Q \quad (10)$$

$$[(r, 0, \triangleleft) \rightarrow \hat{r} \oplus]_h^0 \quad \text{for } r \in Q \quad (11)$$

$$[(r, 0, \triangleright) \rightarrow \hat{r} \ominus]_h^0 \quad \text{for } r \in Q \quad (12)$$

Step 6. The object \oplus , if appearing, changes the charge of the membrane to positive:

$$[\oplus]_h^0 \rightarrow []_h^+ \# \quad (13)$$

If \ominus appears, it behaves similarly, according to rule (1). Simultaneously, the object \hat{r} is primed and produces \odot :

$$[\hat{r} \rightarrow \hat{r}' \odot]_h^0 \quad \text{for } r \in Q \quad (14)$$

Step 7. Now the charge of h is negative if the tape head is moving right, and the indices of the tape-objects have to be decremented, or positive if the tape head is moving left, and the indices must be incremented; the primes are also removed:

$$[1''_i \rightarrow 1_{i-1}]_h^- \quad [1''_i \rightarrow 1_{i+1}]_h^+ \quad \text{for } -(m-1) \leq i \leq m-1 \quad (15)$$

The object \hat{r}' is now rewritten into the state-object r , and produces the \ominus object to be used in Step 1 of the simulation of the next step of M :

$$[\hat{r}' \rightarrow r \ominus]_h^\alpha \quad \text{for } r \in Q \text{ non final and } \alpha \in \{+, -\} \quad (16)$$

Finally, the neutral charge of h is restored by \odot through rule (6). The configuration of the membrane now encodes the next configuration of M , and the system can begin simulating the next computation step. The process is depicted in Fig. 2.

When $r \in Q$ is a final state (accepting or rejecting), instead of applying rule (16) the system rewrites the object \hat{r}' as *yes* or *no*:

$$[\hat{r}' \rightarrow \text{yes}]_h^\alpha \quad \text{for } r \in Q \text{ accepting and } \alpha \in \{+, -\} \quad (17)$$

$$[\hat{r}' \rightarrow \text{no}]_h^\alpha \quad \text{for } r \in Q \text{ rejecting and } \alpha \in \{+, -\} \quad (18)$$

The object *yes* or *no* is then sent out as the result of the computation of the P system in the next step:

$$[\text{yes}]_h^0 \rightarrow []_h^0 \text{yes} \quad [\text{no}]_h^0 \rightarrow []_h^0 \text{no} \quad (19)$$

It is easy to see that this simulation provides us with a uniform family of P systems $\Pi_M = \{\Pi_x : x \in \{0, 1\}^*\}$, each consisting of a single membrane h and simulating the deterministic Turing machine M on all possible inputs.

5.1 Simulating oracle queries

If membrane h is not the outermost membrane of the system, then we can use division rules to simulate nondeterminism with parallelism. Suppose, for the sake of example, that the transition function of M describes nondeterministic binary choices such as $\delta(q, 0) = \{(r, 1, \triangleright), (s, 0, \triangleleft)\}$. Then, instead of the rules (7), we define the elementary division rule

$$[(q, 0)]_h^0 \rightarrow [(r, 1, \triangleright)]_h^0 [(s, 0, \triangleleft)]_h^0 \quad (20)$$

The two resulting copies of membrane h can then evolve in parallel according to the two possible choices.

This construction allows us to simulate polynomial-time deterministic Turing machines M with an **NP** oracle. In this section, we use the following conventions: the machine M simulates a work tape and a query tape with a single tape, by using the odd and even positions, respectively. When making a query, M writes the query string in the even positions of its tape, then enters a query state $q_?$. The oracle answers by erasing the query string (i.e., overwriting it with zeros), except for the first cell, where it writes 0 or 1 according to the result. The machine M then resumes its computation in state $q_!$, with the tape head located on the answer.

The oracle can be simulated by a polynomial-time nondeterministic Turing machine M' deciding the oracle language. We can assume, without loss of generality, that this machine has $q_?$ (the same as the query state of M) as its initial state, uses only the even positions of the tape, and ends its computation in the post-query configuration described above. We assume that M' performs a series of nondeterministic choices leading to acceptance, if an accepting computation exists at all.

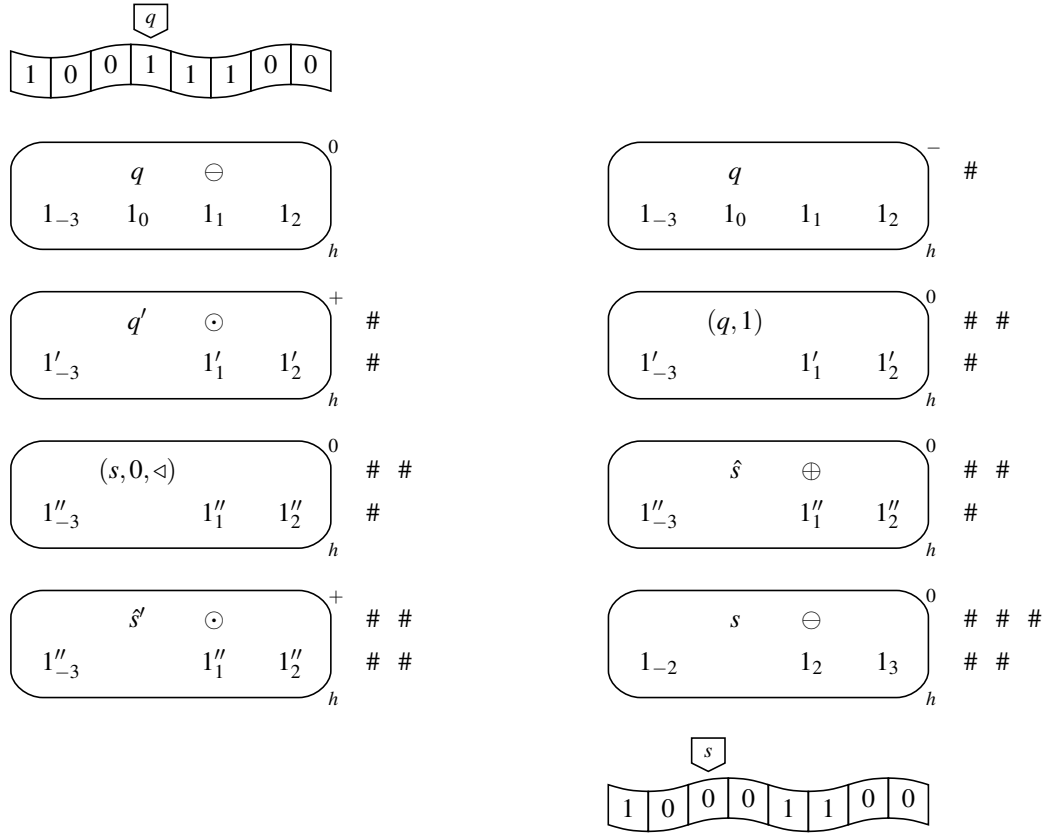


Fig. 2 Two successive Turing machine configurations, and the configurations of the P system simulating the transition step (in left-to-right, top-to-bottom order).

This combination of M and M' can be simulated by linearly nested membranes of a P system, one membrane for each query to be asked. The computation begins inside the innermost membrane, where we place a multiset encoding the initial configuration of M on its input x ; whenever a query is performed, the computation moves one level higher in the membrane structure. In the following description we refer to all nested membranes as h , for brevity; the labels can be made unique, and the rules replicated for each label, with a polynomial-time preprocessing. The P system simulates the computation steps of M as described above, until M enters the query state $q?$. Now the system pauses the simulation of M . Instead of producing $q?$ and \ominus , as in rule (16), the system produces $q?$ and $\tilde{q}_{1,t}$, where t is the maximum number of steps required by M' on query strings written by M . This number can be bounded above by considering the polynomial running time of M' on the longest possible query string, which is at most as long as the running time of M on its input x . The object $\tilde{q}_{1,t}$ is sent out from h as $q_{1,t}$, setting its charge to negative as \ominus does, and upon reaching the parent

membrane it begins counting down:

$$[\tilde{q}_{1,t}]_h^0 \rightarrow []_h^- q_{1,t} \quad (21)$$

$$[q_{1,j} \rightarrow q_{1,j-1}]_h^0 \quad \text{for } 1 \leq j \leq t \quad (22)$$

In the internal membrane, the nondeterministic Turing machine M' is now simulated. Since M' is allowed to make nondeterministic choices, in general there will be a number of membranes simulating M' after the first simulated step. When one of these membranes is simulating the last step of a computation of M' , the object \hat{q}'_1 is produced by rule (14): then, instead of having a rule of type (16), the object \hat{q}'_1 is used to dissolve the membrane and release the tape-objects to the parent membrane:

$$[\hat{q}'_1]_h^\alpha \rightarrow \# \quad \text{for } \alpha \in \{+, -\} \quad (23)$$

After t steps, all membranes simulating M' have completed the simulation, and have released their contents to the parent membrane. This membrane now contains:

- the object $q_{1,0}$;

- objects 1_i corresponding to the 1s contained in the odd positions of the tape of M (which are left unchanged by the simulation of M'); each of these objects has a multiplicity equal to the number of computations of M' on the previous query string;
- zero or more occurrences of 1_1 , one for each accepting computation of M' on the query string; in particular, there is at least one occurrence of 1_1 if and only if the query string is accepted by the oracle. Notice that this object has index 1 even if it is on the first even position of the tape, since index 0 is reserved to the tape cell under the head (tape cell 1).

Before resuming the simulation of M , the system needs to eliminate any duplicate copies of objects 1_i . First of all, the object $q_{!,0}$ is rewritten into $q_!$, the next state of M :

$$[q_{!,0} \rightarrow q_!]_h^0 \quad (24)$$

We then change the behaviour of M in such a way that, before continuing its original computation after receiving the answer to the oracle query, it sweeps its entire tape left-to-right and back to the first cell. This behaviour, in conjunction with the following extra rule of the P system:

$$[1_0 \rightarrow \varepsilon]_h^+ \quad (25)$$

erases any duplicate of 1_i for all i . Indeed, if a copy of 1_0 appears when h is positive, then another copy has been sent out in the previous step by rule (2); rule (25) eliminates such duplicates.

When the tape head of M moves back to the leftmost cell, the machine can resume its original behaviour, and the encoding of the configuration of M in the P system is now correct according to the description given at the beginning of this section.

Further queries by M are simulated analogously, by exploiting another level of the membrane structure. Notice that simulating a query actually “consumes” one level of the membrane structure, due to the dissolution rule (23). For this reason, the initial membrane structure of the P system simulating M consists of an outermost membrane, containing as many nested membranes as the number of queries performed by M .

Theorem 4 *A deterministic polynomial-time Turing machine which asks $p(n)$ queries to an **NP** oracle on inputs of length n can be simulated by a uniform family of monodirectional P systems of depth $p(n)$ without non-elementary division rules.*

Proof. The family of P systems $\Pi = \{\Pi_x : x \in \{0, 1\}^*\}$ simulating M on input x can be constructed uniformly in polynomial time, since only the initial multiset depends on the actual string x , while the set of rules and the membrane structure only depend on $|x|$. We only need to make sure that the

indices of the tape-objects are large enough to ensure that both the tape of M and the tape of M' can be represented at the same time. \square

Corollary 1 $\mathbf{P}^{\mathbf{NP}} \subseteq \mathbf{PMC}_{\mathcal{M}(-\text{wn})}$. \square

Instead of using membrane dissolution as in rule (23), we can use the object $\hat{q}_!$ to produce \oplus :

$$[\hat{q}_! \rightarrow \oplus]_h^\alpha \quad \text{for } \alpha \in \{+, -\} \quad (26)$$

which ensures that the charge of h is positive instead of negative two steps later. The tape-objects are then sent out, one at a time, by using the following rules:

$$[1_i]_h^+ \rightarrow []_h^+ 1_i \quad \text{for } -(m-1) \leq i \leq m-1 \quad (27)$$

The timer t of the object $\tilde{q}_{!,t}$ has to be increased appropriately, in order to take into account the time needed to send out all the tape-objects. However, since the membrane where the simulation of M is non-elementary after the first query, rule (20) is now a weak non-elementary division rule. As a consequence, we have:

Theorem 5 *A deterministic polynomial-time Turing machine which asks $p(n)$ queries to an **NP** oracle on inputs of length n can be simulated by a uniform family of monodirectional P systems of depth $p(n)$ without dissolution rules.* \square

Corollary 2 $\mathbf{P}^{\mathbf{NP}} \subseteq \mathbf{PMC}_{\mathcal{M}(-\text{d})}$. \square

In order to prove the converse of Theorem 2, we introduce an auxiliary complexity class, a variant of the class of optimisation problems **OptP** (Krentel 1988).

Definition 5 Let **OrP** be the class of functions over binary strings $f: \{0, 1\}^* \rightarrow \{0, 1\}^*$ having a polynomial-time non-deterministic Turing machine M such that $f(x) = \bigvee M(x)$ for all $x \in \{0, 1\}^*$, where $M(x)$ denotes the set of possible output strings of M on input x , and \bigvee denotes bitwise disjunction of strings; here we assume that the bitwise disjunction of strings of different lengths is performed by padding the shortest ones with zeros.

The purpose of the class **OrP** is to capture a polynomial number of parallel **NP** queries with a single query to a function over binary strings.

Proposition 1 $\mathbf{P}_{\parallel}^{\mathbf{NP}} = \mathbf{P}^{\mathbf{OrP}[1]}$.

Proof. A polynomial number of parallel queries y_1, \dots, y_m to an oracle for $L \in \mathbf{NP}$ can be replaced by a single query to an oracle for the function $f(y_1, \dots, y_m) = z_1 \cdots z_m$, where $z_i = 1$ if and only if $y_i \in L$. Let M be an **NP** machine deciding L , and let M' be the following nondeterministic machine: on input y_1, \dots, y_m simulate M on each y_i and record the corresponding output bit z_i ; finally, output $z_1 \cdots z_m$. For all i

with $1 \leq i \leq m$, if y_i is accepted by the oracle, then there exists a computation of M' such that $z_i = 1$: thus, by taking the bitwise disjunction of all possible output strings of M' , we obtain the i -th bit of $f(y_1, \dots, y_m)$; this proves that $f \in \mathbf{OrP}$. Notice that this proof requires the query strings y_1, \dots, y_m to be fixed in advance, i.e., the queries cannot be performed adaptively.

Vice versa, a single query to an oracle for $f \in \mathbf{OrP}$ with query string y can be replaced by the following polynomial number of parallel queries, one for each $1 \leq i \leq |f(y)|$: “is the i -th bit of $f(y)$ a 1?”. These queries are in \mathbf{NP} , since they can be answered by simulating an \mathbf{OrP} machine M for f and selecting only its i -th output bit; the answer will be positive if and only if there exists a computation of M having a 1 as the i -th output bit, which (by definition of \mathbf{OrP}) is equivalent to the i -th bit of $f(y)$ being 1. \square

Simulating an \mathbf{OrP} query by means of a P system is completely analogous to simulating an \mathbf{NP} query, except that, instead of a single output bit, we have a polynomial number of them. These binary strings are automatically combined by bitwise disjunction when the tape-objects are sent out of the membrane simulating the nondeterministic Turing machine. Furthermore, since a single \mathbf{OrP} query suffices to capture $\mathbf{P}_{\parallel}^{\mathbf{NP}}$, we obtain the following results:

Theorem 6 *A deterministic polynomial-time Turing machine which asks a polynomial number of parallel queries to an \mathbf{NP} oracle on inputs of length n can be simulated by a uniform family of monodirectional P systems of depth 1 without dissolution (and, necessarily, without non-elementary division).* \square

Corollary 3 $\mathbf{P}_{\parallel}^{\mathbf{NP}} \subseteq \mathbf{PMC}_{\mathcal{M}(-d, -wn)}$. \square

6 Further results

The depth of the P systems of Theorems 4 and 5 can be asymptotically reduced by exploiting the equivalence of a logarithmic number of adaptive queries and a polynomial number of parallel queries (Papadimitriou 1993, Theorem 17.7), formally $\mathbf{P}_{\parallel}^{\mathbf{NP}} = \mathbf{P}^{\mathbf{NP}[\log n]}$. Suppose a deterministic polynomial-time Turing machine performs $p(n)$ sequential \mathbf{NP} queries, and divide these queries into $\Theta(p(n)/\log n)$ blocks of $\Theta(\log n)$ queries. Each block can then be replaced by a polynomial number of parallel \mathbf{NP} queries or, by Proposition 1, by a single \mathbf{OrP} query. Hence, $p(n)$ sequential \mathbf{NP} queries can be simulated by $\Theta(p(n)/\log n)$ sequential \mathbf{OrP} queries, and each of the latter can be simulated by one level of depth in a P system:

Corollary 4 *A deterministic polynomial-time Turing machine which asks $p(n)$ queries to an \mathbf{NP} oracle on inputs*

of length n can be simulated by a uniform family of monodirectional P systems of depth $\Theta(p(n)/\log n)$ without non-elementary division rules (or without division rules). \square

Theorem 3 can be sharpened by making the intra-level query parallelism explicit with \mathbf{OrP} queries:

Corollary 5 *Let Π be a family of semi-uniform polynomial-time monodirectional P systems of depth $f(n)$. Then Π can be simulated by a polynomial-time deterministic Turing machine with $f(n)$ queries to an \mathbf{OrP} oracle.* \square

We can also prove that monodirectional families of P systems of any constant depth, even with dissolution and non-elementary division rules (denoted, in symbols, by $\mathcal{M}(\mathcal{O}(1))$), are always equivalent to families of depth one without dissolution and without non-elementary division (in symbols $\mathcal{M}(1, -d, -wn)$), and thus only able to simulate parallel \mathbf{NP} queries.

Theorem 7 $\mathbf{PMC}_{\mathcal{M}(\mathcal{O}(1))}^{[*]} = \mathbf{PMC}_{\mathcal{M}(1, -d, -wn)}^{[*]} = \mathbf{P}_{\parallel}^{\mathbf{NP}}$.

Proof. By Theorem 6, we know that $\mathbf{P}_{\parallel}^{\mathbf{NP}} \subseteq \mathbf{PMC}_{\mathcal{M}(\mathcal{O}(1))}$, even when limited to depth 1. Furthermore, the inclusion $\mathbf{PMC}_{\mathcal{M}(\mathcal{O}(1))} \subseteq \mathbf{PMC}_{\mathcal{M}(\mathcal{O}(1))}^*$ holds by definition. We can prove the inclusion $\mathbf{PMC}_{\mathcal{M}(\mathcal{O}(1))}^* \subseteq \mathbf{P}_{\parallel}^{\mathbf{NP}}$ as follows. By Theorem 3, a family of P systems of constant depth k can be simulated in polynomial time by asking k sets (one per level) of $p(n)$ parallel queries, for some polynomial p . Each set of $p(n)$ parallel queries can be converted into $\Theta(\log n)$ sequential queries (Papadimitriou 1993, Theorem 17.7), for a total of $k \cdot \Theta(\log n)$ sequential queries. These can be converted back into a polynomial number of parallel queries. \square

Finally, observe that Theorem 3 also trivially holds for monodirectional P systems *without charges*. This implies a better upper bound than previously known (Leporati et al 2014) for a monodirectional variant of the P conjecture (Păun 2005, Problem F), which states that P systems without charges and without non-elementary division characterise \mathbf{P} .

7 Conclusions

In this paper we confirmed the importance of the direction of the information flow in P systems with active membranes with respect to their computing power. Indeed, when working in polynomial time and using only outward-bound communication, the corresponding complexity class decreases from \mathbf{PSPACE} (Alhazov et al 2003) to $\mathbf{P}^{\mathbf{NP}}$ and, when non-elementary division and dissolution rules are disallowed, from $\mathbf{P}^{\#\mathbf{P}}$ (Leporati et al 2015) to $\mathbf{P}_{\parallel}^{\mathbf{NP}}$. It is interesting to notice that, unlike with other restrictions such as removing membrane division (Zandron et al 2001) or charges and dissolution (Gutiérrez-Naranjo et al 2006), the resulting P systems are still more powerful than \mathbf{P} (unless, of course, $\mathbf{P} = \mathbf{NP}$).

The role of strong non-elementary division (which is replaced in this paper by weak non-elementary division) in the absence of send-in rules is still unclear. Even if it provides a way to convey information from a parent membrane to its children, we do not know whether this is sufficient to altogether replace send-in communication while maintaining a polynomial run-time.

Finally, it would be interesting to investigate monodirectional P systems where the information flow is reversed, i.e., send-out communication and dissolution rules (as well as strong non-elementary division rules) are disallowed. A first issue to overcome is choosing an appropriate acceptance condition for the P systems, to replace sending out *yes* or *no* from the outermost membrane. The acceptance condition most similar “in spirit” to the original one is probably accepting (resp., rejecting) by having at least one *yes* (resp., *no*) object appear, either anywhere in the system, or inside a distinguished (and possibly dividing) membrane, during the last computation step; we also add the restriction that *yes* and *no* can never appear together, since giving the priority to one of them would allow us to solve NP-complete (or coNP-complete) problems “for free”. Such monodirectional P systems appear to be very weak when working in polynomial time; indeed, even though exponentially many membranes can still be created by division, they can never communicate. Is P actually an upper bound to the class of problems they can solve?

References

- Alhazov A, Martín-Vide C, Pan L (2003) Solving a PSPACE-complete problem by recognizing P systems with restricted active membranes. *Fundamenta Informaticae* 58(2):67–77, URL <http://iospress.metapress.com/content/99n72anvn6bk14mm/>
- Gutiérrez-Naranjo MA, Pérez-Jiménez MJ, Riscos-Nuñez A, Romero-Campero FJ (2006) Computational efficiency of dissolution rules in membrane systems. *International Journal of Computer Mathematics* 83(7):593–611, URL <http://dx.doi.org/10.1080/00207160601065413>
- Krentel MW (1988) The complexity of optimization problems. *Journal of Computer and System Sciences* 36:490–509, URL [http://dx.doi.org/10.1016/0022-0000\(88\)90039-6](http://dx.doi.org/10.1016/0022-0000(88)90039-6)
- Leporati A, Manzoni L, Mauri G, Porreca AE, Zandron C (2014) Simulating elementary active membranes, with an application to the P conjecture. In: Gheorghe M, Rozenberg G, Sosík P, Zandron C (eds) *Membrane Computing, 15th International Conference, CMC 2014, Lecture Notes in Computer Science*, vol 8961, Springer, pp 284–299, URL http://dx.doi.org/10.1007/978-3-319-14370-5_18
- Leporati A, Manzoni L, Mauri G, Porreca AE, Zandron C (2015) Membrane division, oracles, and the counting hierarchy. *Fundamenta Informaticae* 138(1–2):97–111, URL <http://dx.doi.org/10.3233/FI-2015-1201>
- Murphy N, Woods D (2007) Active membrane systems without charges and using only symmetric elementary division characterise P. In: Eleftherakis G, Kefalas P, Păun Gh, Rozenberg G, Salomaa A (eds) *Membrane Computing, 8th International Workshop, WMC 2007, Lecture Notes in Computer Science*, vol 4860, pp 367–384, URL http://dx.doi.org/10.1007/978-3-540-77312-2_23
- Murphy N, Woods D (2011) The computational power of membrane systems under tight uniformity conditions. *Natural Computing* 10(1):613–632, URL <http://dx.doi.org/10.1007/s11047-010-9244-7>
- Papadimitriou CH (1993) *Computational Complexity*. Addison-Wesley
- Păun Gh (2001) P systems with active membranes: Attacking NP-complete problems. *Journal of Automata, Languages and Combinatorics* 6(1):75–90
- Păun Gh (2005) Further twenty six open problems in membrane computing. In: Gutiérrez-Naranjo MA, Riscos-Nuñez A, Romero-Campero FJ, Sburlan D (eds) *Proceedings of the Third Brainstorming Week on Membrane Computing*, Félix Editora, pp 249–262, URL <http://www.gcn.us.es/3BWMC/Volumen.htm>
- Zandron C, Ferretti C, Mauri G (2001) Solving NP-complete problems using P systems with active membranes. In: Antoniou I, Calude CS, Dinneen MJ (eds) *Unconventional Models of Computation, UMC’2K, Proceedings of the Second International Conference*, Springer, pp 289–301, URL http://dx.doi.org/10.1007/978-1-4471-0313-4_21
- Zandron C, Leporati A, Ferretti C, Mauri G, Pérez-Jiménez MJ (2008) On the computational efficiency of polarizationless recognizer P systems with strong division and dissolution. *Fundamenta Informaticae* 87:79–91, URL <http://iospress.metapress.com/content/95t4416hp5w165m2/>