

# Stochastic Simulation of Biological Systems with Dynamical Compartment Structure

Cristian Versari      Nadia Busi

Università di Bologna, Dipartimento di Scienze dell'Informazione  
Mura Anteo Zamboni 7, 40127 Bologna, Italy  
{versari,busi}@cs.unibo.it

## Abstract

The Gillespie stochastic simulation algorithm represents one of the main physical abstractions exploited for the simulation of biological systems modeled by means of concurrent calculi. While the faithful modelling of bio-systems often requires multi-compartment semantics, the original Gillespie algorithm considers only one fixed-size volume. In this paper we introduce an extended formalisation of the above algorithm which preserves the original model but allows the stochastic simulation in presence of multiple compartments with dynamical structure and variable sizes. The presented algorithm can be then used as basis for simulating systems expressed in an extended version of the stochastic  $\pi$ -Calculus, the  $S\pi@$  language, obtained by means of polyadic synchronisation. Despite of its conservativeness,  $S\pi@$  is showed to allow flexible modelling of multiple compartments with dynamical structure and to provide increased biological faithfulness.

## 1 Introduction

The Gillespie stochastic simulation algorithm (SSA for short) represents one of the main physical abstractions exploited for the simulation of discrete, stochastic processes which model chemical and biological systems [21, 19, 20, 16]. The original algorithm presented in [9] allows the simulation of systems composed of a single adiabatic compartment of fixed volume, containing a uniform mixture of chemical species interacting and producing new chemical reactants. While the abstraction of single, adiabatic, fixed-volume compartment is well suited for an isolated, gas-phase system as assumed in [9], the simulation of biological systems requires to enrich the model with some non-trivial details. As denoted by many recent approaches (e.g. [20, 3, 18]) the multi-compartment structure appears as a natural and indispensable requirement for the faithful representation of biological systems.

The most intuitive way to adapt the SSA to multiple compartments [5] is to consider each compartment as an isolated system evolving in parallel with the others, so that each compartment is implemented as an instance of the SSA. Even if correct, the above approach is based on strong assumptions which severely limit its application.

In fact, the distinctive behaviour of biological systems is the continuous interaction of their con-

stituting elements. A compartmentalised model of a bio-systems cannot dispense with the capability of interaction between adjacent compartments by means of interaction of their respective elements or exchange of material: a refined model should also be able of representing a dynamical compartment structure. When the simulation of compartments is obtained by independent instances of the SSA under its original assumptions, the interaction between compartments must obey the hypothesis of volume invariance but, for a liquid-state system like biological ones usually are, this means that the exchanged elements must be inappreciable with respect to the dimension of the whole compartments. This approximation can be tolerated until very few molecules walk through during the time of the simulation, but becomes unreasonable when the aim is the modelling of compartments with dynamical structure (i.e. the system is subject to complex structural changes like for example the movement, merging, splitting of compartments), a necessary step towards a satisfactory representation of biological systems.

Besides the limit of static compartment structure, the above approach presents another relevant drawback. Consider the system in Fig. 1 (a), composed of two compartments with different volume. According to the model of the SSA, the molecules  $m_1$  and  $m_2$  have different probability of collision

when floating in compartment  $C_1$  instead of  $C_2$ . In fact, since  $C_2$  is characterised by a larger volume, the probability to collide is lower than in  $C_1$ . Since the SSA requires to specify the reaction rate (more precisely, the “reaction probability per unit time”) for each reaction, the same kind of reaction (like for example the one between  $m_1$  and  $m_2$ ) is characterised by a different rate for each compartment. In other words, in presence of  $K$  compartments, the same reaction rule (performed by the same kind of reactants, with the same physical properties) will have  $K$  different rates depending on the compartment the reaction is localised into, hence  $K$  different representations, one for each instance of the SSA.

In this paper we follow a different approach to the extension of the SSA to the multi-compartment model: here we show that by introducing in the model the informations pertaining the volumes, the SSA can be transparently adapted to the exact representation of multiple compartments. Furthermore, we show that the extended model is consistent with the kinetic hypotheses of the original one, but closer to the biological scenario and simpler w.r.t. to the approach discussed above, since allows to give unique description of elements and reactions while keeping consistent their rates in function of the enclosing compartment.

Then we show how the extended SSA can be smoothly used as basis for the simulation of systems modeled in a stochastic version of the  $\pi$ -Calculus added with polyadic synchronisation. This extension, we called  $S\pi@$ , exhibits a significant increase of expressiveness and faithfulness towards biological modellings, despite of its conservativeness w.r.t. both the  $\pi$ -Calculus and the original Gillespie model.

We consider the reader familiar with the Gillespie algorithm [9] and the SpiM simulator [16].

**Structure of the paper.** In Sect. 2 the extension of the original Gillespie stochastic model to the multi-compartment context is presented. In Sect. 3 the  $S\pi@$  language is formalised and the multi-compartment version of the SSA is coded. In Sect. 4 some chemical and biological modelling examples are shown and in Sect. 6 the pros and cons of the approach and future improvements are discussed.

## 2 Multi-Compartment Gillespie Stochastic Model

The original SSA considers a gas-phase system of fixed volume  $V$ , filled (in the simplest case) with

two (distinct) molecular species  $S_1$  and  $S_2$  which can undergo the reaction  $R_1$  (that is react through the *reaction channel*  $R_1$ ). By simple kinetic deductions, the probability  $c dt$  of collision of two molecules  $m_1$  and  $m_2$  (modeled as hard spheres of radii  $d_1$  and  $d_2$ ) in the infinitesimal time interval  $(t, t + dt)$  is calculated as

$$c dt = \overline{dV_{coll}/V} = V^{-1} \pi d_{12}^2 \overline{v_{12}} dt$$

where  $dV_{coll}$  is the average “collision volume”,  $d_{12} = d_1 + d_2$  and  $\overline{v_{12}}$  is the average relative velocity of the two molecules. In the same way, the probability  $c_1 dt$  of *reactive collision* of two molecules can be calculated as

$$c_1 dt = c c_r dt = V^{-1} \pi d_{12}^2 \overline{v_{12}} c_r dt$$

where  $c_r$  represents the probability that a given collision is actually reactive. In general, there exists a constant  $r$  which depends only on the physical properties of the two molecules and the temperature of the system, such that

$$c_1 dt = V^{-1} r dt$$

Furthermore, if  $X_1$  and  $X_2$  are the number of molecules of type  $S_1$  and  $S_2$  respectively, the probability  $a_1 dt$  that an  $R_1$  reaction will occur somewhere inside  $V$  in the infinitesimal time interval  $(t, t + dt)$  is

$$a_1 dt = X_1 X_2 c_1 dt = V^{-1} X_1 X_2 r dt \quad (1)$$

The value  $a_1$  captures all the informations pertaining the *concentration* of the reactants  $S_1$  and  $S_2$ . In the case of a single compartment of fixed volume,  $a_1$  depends only on the quantity of reactants inside  $V$ , while in the case of multiple compartments or variable volume, expression (1) allows to calculate the effective reaction rates in function of the number of involved molecules inside each compartment and the volume of the compartment itself.

The effect of compartments is to *separate* the enclosed elements, that is to *prevent the interaction between elements placed in different compartments*. If  $m_{11}, m_{21}$  are two molecules of species  $S_1$  and  $S_2$  inside compartment  $C_1$ , and  $m_{12}, m_{22}$  are two molecules of species  $S_1$  and  $S_2$  inside compartment  $C_2$ , even if  $m_{11}$  may interact with  $m_{22}$  by an  $R_1$  reaction, their collision is prevented because of the separation granted by compartment boundaries. Even if the pairs  $m_{11}, m_{21}$  and  $m_{12}, m_{22}$  may undergo the same reaction of type  $R_1$ , the reaction  $R_1$  of elements inside compartment  $C_1$  is completely independent of the same reaction  $R_1$  of elements inside compartment  $C_2$ . Therefore, in a multi-compartment environment each *reaction*

*channel* should be denoted not only by the *type* of reaction but also by the *compartment* the reaction happens in.

By following these intuitions, the SSA can be transparently adapted to simulate the multi-compartment model (we call multi-compartment Gillespie model, MCGM for short) without affecting its original kinetic hypotheses.

If  $S_1, \dots, S_N$  are the chemical species and  $R_1, \dots, R_M$  are the reaction types,  $X_{11}, \dots, X_{NK}$  are the number of molecules of each of the  $N$  species inside each of the  $K$  compartments,  $R_{11}, \dots, R_{MK}$  are the reaction channels (in the number of  $M \cdot K$ , one for each reaction type inside each compartment) then we can introduce, in place of  $P(\tau, \mu) d\tau$  [9], the expression  $P(\tau, (\mu, c)) d\tau$  which denotes *the probability that, given the state  $(X_{11}, \dots, X_{NK})$  at time  $t$ , the next reaction in  $V$  will occur in the infinitesimal time interval  $(t+\tau, t+\tau+d\tau)$ , and will be a reaction through the channel  $R_{\mu c}$* . Then, according to expression (1), we define the *propensity function* of the reaction  $R_{\mu c}$ :

$$a_{\mu c} dt = V_c^{-1} h_{\mu c} r_{\mu} dt \quad (2)$$

is *the probability that an  $R_{\mu c}$  reaction (that is an  $R_{\mu}$  reaction inside compartment  $c$ ) will occur in the system in  $(t, t+dt)$ , given that the system is in the state  $(X_{11}, \dots, X_{NK})$  at time  $t$ , where*

- $r_{\mu} dt$  is the average probability that a particular molecular pair will react according to  $R_{\mu}$  in the infinitesimal time interval  $dt$  inside a compartment whose volume is equal to 1 in the given volume measure unit
- $h_{\mu c}$  is the number of distinct  $R_{\mu c}$  molecular reactant combinations available in the state  $(X_{11}, \dots, X_{NK})$  (that is the molecular reactant combinations pertaining reaction  $R_{\mu}$  inside compartment  $c$ )
- $V_c$  is the volume of compartment  $c$ .

According to [9], for a reaction of the kind  $S_1 + S_2 \rightarrow S_3$ ,  $h_{\mu c}$  is the product  $X_{1c}X_{2c}$ , while for a reaction  $2S_1 \rightarrow S_2$ ,  $h_{\mu c} = X_{1c}(X_{1c} - 1)/2$ .

Finally, by the same considerations stated in [9], the following expression can be easily derived:

$$P(\tau, (\mu, c)) = \begin{cases} a_{\mu c} \exp(-a_0 \tau) & 0 \leq \tau < +\infty, \\ & \mu = 1, \dots, M \\ 0 & \text{otherwise} \end{cases} \quad (3)$$

where  $a_{\mu c} = V_c^{-1} h_{\mu c} r_{\mu}$  and  $a_0 = \sum_{\nu=1}^M \sum_{\gamma=1}^K a_{\nu \gamma}$ .

### 3 The Stochastic $\pi@$ language

The MCGM is a transparent generalisation of the original Gillespie model, parameterised w.r.t. the compartments the system is composed of. The price of this generalisation is the insertion, in the implementation of the model, of the informations pertaining the volume size of each compartment, which can be accomplished in different ways. The most immediate method would be the definition of a function  $\text{Vol} : \mathcal{C} \rightarrow \mathbb{R}$  returning the volume size of each compartment, but this would not allow to model variable volumes.

The approach chosen here is to allow (but not require) the specification of the volume  $\bar{V}_i$  for each element  $i$ , which represents *the (average) increment of volume of a compartment  $c$  needed for including the additional element  $i$* . In the case of constant volume,  $\bar{V}_i = 0$  for each  $i$ . In fact, the additional volume required to include any further element is 0 if the volume is constant. Conversely, at constant and homogeneous temperature and pressure, for a single chemical species  $S$  occupying volume  $V$ ,  $V_i$  can be calculated as

$$\bar{V}_i = \frac{V}{n} = \frac{m}{n} \cdot \frac{1}{D} = \frac{w_S}{D}$$

where  $n$  is the number of molecules of  $S$  inside  $V$ ,  $m$  is the total mass,  $D$  is the density of  $S$ ,  $w_S$  its molecular weight.

$\bar{V}_i$  is a function depending on the kind of the element  $i$  but also on the compartment  $c$ , since the chemical composition of  $c$  may vary the average distance between the inner elements, because of atomic-level forces like van der Waals interactions and hydrogen bonds. Under the assumption (tolerable if the discussed variation is reasonably small or the chemical composition of the compartments is almost the same) of constant  $\bar{V}_i$  for each  $i$ , the volume  $V_c$  of each compartment  $c$  can be easily calculated as the sum of  $\bar{V}_i$  for each  $i$  inside  $c$ , even in the case of exchange of molecules or reorganisations in the compartment structure.

In the next section, the syntax and semantics of  $S\pi@$  is defined according to the above assumptions. In Sect. 4 some useful examples are presented and it is shown how the approach chosen for  $S\pi@$  is both general and simple, since it allows to represent the function  $\text{Vol} : \mathcal{C} \rightarrow \mathbb{R}$  discussed above as well as the single-compartment situation but also  $K$  multiple compartments with variable volumes, without increasing the computational complexity of the algorithm, complicating the notation or forcing the introduction of unessential informations in the model.

### 3.1 Syntax and Semantics

The  $\pi@$  language [23] is a conservative extension of the  $\pi$ -Calculus [14, 15, 13] provided with polyadic synchronisation [2] and different levels of priority for actions [6]. The  $S\pi@$  language presented here can be considered in first approximation as the stochastic version of a core  $\pi@$  limited to two levels of priority and two names for each channel. The capability of giving infinite rates to reactions replaces the two priority levels of this core  $\pi@$ , while the two names denoting each action assume different meaning, since the first represents the type of (chemical) reaction, while the second the compartment where the reaction takes place.

**Definition 1** Let  $\mathcal{N}, \mathcal{C}$  be distinct sets of names on finite alphabet, with  $m, n$  ranging over  $\mathcal{N}$ ,  $a, b$  over  $\mathcal{C}$  and  $x, y$  over  $\mathcal{X} = \mathcal{N} \cup \mathcal{C}$ . Let also  $v$  range over  $\mathbb{R}$  within the interval  $[0, +\infty[$ . The syntax of the  $S\pi@$  language is defined as

$$\begin{aligned} P & ::= \mathbf{0} \mid \sum_{i \in I} \pi_i.P_i \mid P \mid Q \mid \\ & \quad !\pi.P \mid (\nu x)P \\ \pi & ::= n@a:v(\tilde{x}) \mid \bar{n}@a:v(\tilde{x}) \end{aligned}$$

where  $\tilde{x}$  represents zero or more names  $x_1, \dots, x_i$  ranging over  $\mathcal{X}$ .

$\mathbf{0}$  is the null process, capable of doing nothing.  $\sum_{i \in I} \pi_i.P_i$ , written also  $\pi_1.P_1 + \pi_2.P_2$  in the case  $|I| = 2$ , represents the guarded choice between different actions.  $P \mid Q$  means that  $P$  and  $Q$  are two processes executing in parallel.  $!\pi.P$  is the guarded replication.  $(\nu x)P$  allows the scope restriction of the name  $x$ : the restriction of compartment names allows the creation of new compartments, while the restriction of reaction names is used in several ways, like for representing bindings between different elements. The expressions  $n@a:v(\tilde{x})$  and  $\bar{n}@a:v(\tilde{x})$  represent respectively the polyadic input and output capabilities of a process, where

- $n$  is the kind of reaction the process is ready to perform: in Expr. 2 it corresponds to the index  $\mu$  denoting the reaction  $R_\mu$ ;
- $a$  is the compartment where the reaction may take place;
- $v$  corresponds to the average volume  $\bar{V}_i$  of each element  $i$  discussed above and represents the volume occupied inside compartment  $a$  by the process ready to perform the input or output action.

For sake of simplicity, the set  $\mathcal{N}$  of reaction names is kept distinct from compartment names  $\mathcal{C}$ . In order to retain syntactical and semantical compatibility with SPiM, the additional informations of compartment and volume are kept optional, that is the grammar is implicitly added with the rules

$$\begin{aligned} \pi & ::= n@a(\tilde{x}) \mid \bar{n}@a(\tilde{x}) \mid n:v(\tilde{x}) \mid \\ & \quad \bar{n}:v(\tilde{x}) \mid n(\tilde{x}) \mid \bar{n}(\tilde{x}) \end{aligned}$$

which allow to omit volume and compartment informations if not needed. If omitted, the volume  $v$  is considered 0, while the compartment  $c$  is given a default, distinct value. For sake of simplicity, the additional grammar rules are not considered in the following definitions but their inclusion is straightforward.

**Definition 2** The congruence relation  $\equiv$  is defined as the least congruence satisfying alpha conversion, the commutative monoidal laws with respect to both  $(\mid, \mathbf{0})$  and  $(+, \mathbf{0})$  and the following axioms:

$$\begin{aligned} (\nu x)P \mid Q & \equiv (\nu x)(P \mid Q) & \text{if } x \notin \text{fn}(Q) \\ (\nu x)P & \equiv P & \text{if } x \notin \text{fn}(P) \\ !\pi.P & \equiv \pi.(!\pi.P \mid P) \end{aligned}$$

where the function  $\text{fn}$  is defined as

$$\begin{aligned} \text{fn}(n@a:v(\tilde{x})) & \stackrel{\text{def}}{=} \{n, a\} \\ \text{fn}(\bar{n}@a:v(\tilde{x})) & \stackrel{\text{def}}{=} \{n, a, \tilde{x}\} \\ \text{fn}(\mathbf{0}) & \stackrel{\text{def}}{=} \emptyset \\ \text{fn}((\nu x)P) & \stackrel{\text{def}}{=} \text{fn}(P) \setminus \{x\} \\ \text{fn}(\pi.P) & \stackrel{\text{def}}{=} \text{fn}(\pi) \cup \text{fn}(P) \\ \text{fn}(\sum_{i \in I} \pi_i.P_i) & \stackrel{\text{def}}{=} \bigcup_i \text{fn}(\pi_i.P_i) \\ \text{fn}(P \mid Q) & \stackrel{\text{def}}{=} \text{fn}(P) \cup \text{fn}(Q) \\ \text{fn}(!\pi.P) & \stackrel{\text{def}}{=} \text{fn}(\pi.P) \end{aligned}$$

**Definition 3**  $S\pi@$  semantics is given in terms of the following reduction system:

$$\begin{aligned} (C) & \frac{}{(n@a:v_1(\tilde{x}).P + M) \mid (\bar{n}@a:v_2(\tilde{y}).Q + N)} \\ & \quad \xrightarrow{\text{rate}(n)} P\{\tilde{y}/\tilde{x}\} \mid Q \\ (R) & \frac{P \xrightarrow{r} P'}{(\nu x)P \xrightarrow{r} (\nu x)P'} \quad (P) \frac{P \xrightarrow{r} P'}{P \mid Q \xrightarrow{r} P' \mid Q} \\ (E) & \frac{P \equiv Q \quad P \xrightarrow{r} P' \quad P' \equiv Q'}{Q \xrightarrow{r} Q'} \end{aligned}$$

The rule (C) allows the communication of the names  $\tilde{x}$  from process  $P$  to  $Q$ , where they are

properly substituted to names  $\tilde{y}$ . The function  $\text{rate} : \mathcal{N} \rightarrow (\mathbb{R} \cup +\infty)$  is an external function which permits to associate the correct rate to each reaction, where the rate corresponds to the value  $r_\mu$  of Expr. 2. Rules  $(R), (P), (E)$  allow the transition of processes in presence of restriction, parallel operator or by exploiting structural equivalence.

**Definition 4** A  $S\pi@$  system  $S$  is said to be in standard form if

$$S = (\nu \tilde{x})(P_1 \mid \cdots \mid P_j \mid !P_{j+1} \mid \cdots \mid !P_k)$$

and each  $P_i$  is a non-empty sum.

**Proposition 5** For every  $S\pi@$  system  $S$ , there exists a system  $S'$  such that  $S \equiv S'$  and  $S'$  is in standard form.

In order to calculate the value  $h_{\mu c}$  of Expr. 2, we introduce, according to [16], the function  $\text{Act}$  which permits to know the number of possible combinations of inputs and outputs on a reaction channel inside a given compartment.

**Definition 6** The activity  $\text{Act}$  of channel  $n$  inside compartment  $a$  in the system  $S$  is defined as

$$\text{Act}_{n@a}(S) = (\text{In}_{n@a}(S) \cdot \text{Out}_{n@a}(S)) - \text{Mix}_{n@a}(S)$$

where  $S$  is in standard form,  $\text{In}_{n@a}(S)$  and  $\text{Out}_{n@a}(S)$  are the number of unguarded inputs and outputs on channel  $n$  inside compartment  $a$ , and  $\text{Mix}_{n@a}(S)$  is the sum of  $\text{In}_{n@a}(\sum_i) \cdot \text{Out}_{n@a}(\sum_i)$  for each summation  $\sum_i$  in  $S$ .

The function  $\text{chan}$  allows to know all the active channels inside each compartment in a given system  $S$ .

**Definition 7** Given a  $S\pi@$  system  $S$  in standard form

$$S = (\nu \tilde{x})(P_1 \mid \cdots \mid P_j \mid !P_{j+1} \mid \cdots \mid !P_k)$$

the function  $\text{chan}$  is defined recursively as follows:

$$\begin{aligned} \text{chan}(S) &= \bigcup_{i=1}^k \text{chan}(P_i) \\ \text{chan}(\sum_{i \in I} \pi_i.P_i) &= \bigcup_{i \in I} \text{chan}(\pi_i) \\ \text{chan}(n@a:v(\tilde{x})) &= \text{chan}(\bar{n}@a:v(\tilde{x})) = \{n@a\} \end{aligned}$$

The volume  $V_c$  of Expr. 2 is calculated as the sum of the volumes occupied by each process ready to perform a reaction inside compartment  $c$ . It is worth noticing that in the case of nested compartments as in Fig. 1 (a), the volumes of the compartments inside  $c$  must not be summated to  $V_c$  in order

to respect Expr. 2, since the elements of  $c$  are not allowed to diffuse into the inner compartments and consequently their concentration does not depend on inner compartments volumes.

**Definition 8** Given a  $S\pi@$  system  $S$  in standard form

$$S = (\nu \tilde{x})(P_1 \mid \cdots \mid P_j \mid !P_{j+1} \mid \cdots \mid !P_k)$$

the volume  $\text{Vol}_a$  of the compartment  $a$  in the system  $S$  is calculated as follows:

$$\begin{aligned} \text{Vol}_a(S) &= \sum_{i=1}^k \text{Vol}_a(P_i) \\ \text{Vol}_a(\sum_{i \in I} \pi_i.P_i) &= \sum_{i \in I} \text{Vol}_a(\pi_i) \\ \text{Vol}_a(\bar{n}@a:v(\tilde{x})) &= \text{Vol}_a(n@a:v(\tilde{x})) \\ \text{Vol}_a(n@a:v(\tilde{x})) &= \begin{cases} v & a = b \\ 0 & \text{otherwise} \end{cases} \end{aligned}$$

If  $\text{Vol}_a(S) = 0$ , then  $a$  is given the default volume value 1.

When  $\text{Vol}_a(S) = 0$ , all the informations on the volumes  $v$  occupied by the reactants inside  $a$  have been omitted. This is the case of modelling in Spi, hence it is quite natural to redefine  $\text{Vol}_a(S)$  as 1 so that the reaction rates defined for Spi programs coincide with the reaction rates calculated in  $S\pi@$ .

The volume of a compartment depends on *all* the enclosed objects, not only on reacting ones. Consequently, the volume of non reacting elements must be taken in account and inserted into the program when such volume informations are relevant for the simulation.

Finally, the multi-compartment stochastic simulation algorithm (MSSA) is defined as extension of the SSA given in [9], according to the previous definitions and the MCGM:

**Definition 9** Given a  $S\pi@$  system  $S$  in standard form, the selection of the next reaction  $\text{Next}(S)$  and of the delay  $\text{Delay}(S)$  relative to the MSSA are described by the following algorithm:

1. For each channel  $c_i$  in  $\text{chan}(S)$ , with  $\text{chan}(S) = \{c_1, \dots, c_j\}$ , calculate

$$a_i = \text{Act}_{n@b}(S) * \text{rate}(n) / \text{Vol}_b(S)$$

where  $c_i = n@b$  for some  $n \in \mathcal{N}, b \in \mathcal{C}$ .

2. Calculate  $a_0 = \sum_{i=1}^j a_i$

3. Generate two random numbers  $k_1, k_2 \in [0, 1]$  and calculate  $\tau, \mu$  such that

$$\tau = (1/a_0) \ln(1/k_1)$$

$$\sum_{i=1}^{\mu-1} a_i < k_2 a_0 \leq \sum_{i=1}^{\mu} a_i$$

4.  $\text{Next}(S) = c_\mu$  and  $\text{Delay}(S) = \tau$ .

The value  $c_\mu = n@b$  for some  $n, b$  denotes the reaction channel  $n$  and the compartment  $c$  of the next reaction happening after  $\tau$  time. The two processes performing the synchronisation step on  $c_\mu$  are then randomly chosen as for SPiM.

As noted in [16] for the activities Act, the volumes Vol can be calculated by difference with the respective volumes of the previous execution step, so that the computational complexity order of the algorithm is unvaried.

## 4 Examples

In this section some examples of chemical and biological modelling are presented, in order to illustrate the simplicity of the language and the increased biochemical faithfulness.

### 4.1 Chemical Reactions

If  $m_1$  and  $m_2$  are two molecules floating inside the only, fixed-size compartment of the system, it is quite useless to specify the compartment they are in and the volume they occupy: this is the exact case of modelling in Spi. As discussed after Def. 1, the system composed of  $m_1$  and  $m_2$  could easily be written in  $S\pi@$  as well:

$$M_1 \equiv r_1.N \quad M_2 \equiv \bar{r}_1 \quad S \equiv M_1 \mid M_2$$

where  $m_1$  and  $m_2$  are supposed to undergo a reaction of type  $R_1$  (represented by channel  $r_1$ ) and produce a new chemical reactant  $n$ . As previously discussed, in absence of volume and compartment informations, the volume of the reactants is considered as 0, while that of the (only, unspecified) compartment is considered as 1, so that also the function  $\text{rate}(r_1)$  is consistent – and actually coincides, according to Expr. 2 – with the rate given for the simulation in Spi.

The situation in Fig. 1 (a), which presents multiple compartments could be encoded as

$$M_{11} \equiv r_1@c_1.N(c_1) \quad M_{21} \equiv \bar{r}_1@c_1$$

$$M_{12} \equiv r_1@c_2.N(c_2) \quad M_{22} \equiv \bar{r}_1@c_2$$

$$S \equiv M_{11} \mid M_{21} \mid M_{12} \mid M_{22}$$

if supposing  $c_1$  and  $c_2$  of the same, fixed-size. If their volumes  $v_1, v_2$  are constant, but different, it is possible to specify them by adding two further elements  $d_1, d_2$  which do not take part in the reactions but allow to insert the desired informations:

$$M_{11} \equiv r_1@c_1.N(c_1) \quad M_{21} \equiv \bar{r}_1@c_1$$

$$M_{12} \equiv r_1@c_2.N(c_2) \quad M_{22} \equiv \bar{r}_1@c_2$$

$$D_1 \equiv (\nu n)n@c_1 : v_1 \quad D_2 \equiv (\nu n)n@c_2 : v_2$$

$$S \equiv M_{11} \mid M_{21} \mid M_{12} \mid M_{22} \mid D_1 \mid D_2$$

Even if the processes  $D_1, D_2$  are unable to perform any reaction,  $v_1$  and  $v_2$  are taken into account when calculating  $c_1$  and  $c_2$  volumes. According to Def. 8, since the volume occupied by each of the other processes is 0,  $c_1$  and  $c_2$  volumes turn out to be  $v_1$  and  $v_2$  respectively, as required. In this way the volumes of  $K$  compartments (and consequently the rate of each reaction inside them) can be specified directly inside the program by adding not more than  $K$  non-reactive elements. Alternatively, it is possible to specify the volumes  $V_{m_1}, V_{m_2}$  occupied by each molecule and ensure that the volume of their chemical product  $n$  is equal to the sum of theirs (as it approximately happens in many real reactions). In general, it is possible to specify only the volumes of the desired elements while keeping consistent compartment volumes with reaction rates.

One of the major advantages of  $S\pi@$  approach is the possibility to specify once the behaviour of one element and create instances of it inside any compartment, which is very convenient when the same objects are present in several different places. The previous system could be rewritten as

$$T_1 \equiv !m_1(c).r_1@c.N(c) \quad T_2 \equiv !m_2(c).\bar{r}_1@c$$

$$S \equiv \bar{m}_1\langle c_1 \rangle \mid \bar{m}_1\langle c_2 \rangle \mid \bar{m}_1\langle c_1 \rangle \mid \bar{m}_1\langle c_2 \rangle \mid$$

$$T_1 \mid T_2 \mid D_1 \mid D_2$$

by exploiting the possibility of giving infinite rate to channels  $m_1, m_2$  so that the creation of the instances is immediate.

### 4.2 Biological Modellings

In this section we outline the way some simple biological entities and phenomena [11] can be intuitively represented in  $S\pi@$ . The interaction or exchanging of molecules between compartments constitute the easier and more interesting situations.

**Membrane Receptor.** Consider the simple scheme of a cell membrane receptor whose behaviour is to detect molecules of kind  $m_1$  outside the cell and signal their presence inside the cell by transforming molecules of kind  $m_2$  into  $m_3$ , as in Fig. 1 (b). If we use the membrane as compartment boundary, by modelling the space outside the cell as compartment  $e$  and  $c$  as the cell itself, the receptor – which is a transmembrane protein – lies partially in both compartments. This means that the process  $R$  representing the receptor needs to know the names of the two compartments and to use them for interacting both outside and inside the cell. The system could be for example encoded as

$$\begin{aligned} M_1 &\equiv \overline{bind\_m_1}@e \\ M_2 &\equiv \overline{conv\_m_2}@c \\ R &\equiv !bind\_m_1@e.\overline{conv\_m_2}@c.M_3(c) \end{aligned}$$

where the names  $bind\_m_1, conv\_m_2$  carry the interaction between the receptor  $R$  and the molecules  $m_1$  and  $m_2$ . The process  $M_3(c)$  represents the transformed molecule  $m_3$  inside compartment  $c$ .

**Pump.** The conveyance of molecules or ions across a membrane constitutes a simple example for showing how the exchange of elements between compartments is straightforward in  $S\pi@$ : the sodium/potassium pump represents a quite classical modelling candidate. The  $Na^+/K^+$  ATPase (Fig 1 (c)) is a ion pump which moves three  $Na^+$  ions out and two  $K^+$  ions into the cell for each consumed ATP molecule. The pump can be schematised by process  $P$

$$\begin{aligned} ATP(d) &\equiv \overline{atp}@d \\ ADP(d) &\equiv \overline{adp}@d \\ NA(d) &\equiv \overline{na}@d \\ K(d) &\equiv \overline{k}@d \\ P &\equiv !atp@c.na@c.na@c.na@c. \\ &\quad (ADP(c) \mid k@e.k@e. \\ &\quad (NA(e) \mid NA(e) \mid NA(e) \mid \\ &\quad K(c) \mid K(c))) \end{aligned}$$

which first recruit the ATP molecule, then binds the three sodium ions inside the cell, releases the ADP molecule deriving from the hydrolysed ATP, binds the two  $K^+$  ions outside the cell and finally releases the captured ions into the proper compartments.

**Osmosis.** The last considered example allows to observe the limits of the previous models and the

need to introduce volume information specified in Expr. 2, in order to manage correctly the relative rates of reactions in function not only of the number of elements for each reactant, but also of its concentration. The system in Fig. 1 (d) depicts a compartment  $c$  of variable size containing a water solution placed in a hypotonic environment  $e$ . The compartment is bounded by a semipermeable membrane, i.e. a filter which allows only the water to move across. Experience shows that this situation causes a net movement of water towards compartment  $c$ , due to the so called osmosis phenomenon. Biological occurrence of this circumstance is common both in animal and vegetal cells. A typical animal cell swells when placed in hypotonic and shrinks when in hypertonic solution even if the cell membrane is poorly permeable to water, thanks to the presence of water-channel proteins which allow only water to flow, in either directions. The process  $A$  may describe a simple abstraction of water-channel protein aquaporin:

$$\begin{aligned} HOH(d) &\equiv \overline{hoh}@d : v_{H_2O} & S(d) &\equiv \overline{s}@d : v_S \\ A(f, g) &\equiv !hoh@f.HOH(g) \mid !hoh@g.HOH(f) \end{aligned}$$

$S$  is the solute,  $HOH$  represents a water molecule able to interact with the aquaporin  $A$  and move through the membrane of the cell. In this very simple layout, the aquaporin is completely symmetrical. The system in Fig. 1 (d) may be written as

$$\begin{aligned} Sys &\equiv \underbrace{HOH(e) \mid \dots \mid HOH(e)}_{m_1} \mid \\ &\quad \underbrace{S(e) \mid \dots \mid S(e)}_{n_1} \mid A(c, e) \mid \\ &\quad \underbrace{HOH(c) \mid \dots \mid HOH(c)}_{m_2} \mid \\ &\quad \underbrace{S(c) \mid \dots \mid S(c)}_{n_2} \end{aligned}$$

where  $m_1$  and  $n_1$  represent the number of water molecules and salt ions outside,  $m_2$  and  $n_2$  the number of molecules and ions inside the cell membrane and only one aquaporin is present, for simplicity. If we discard the informations on the volumes  $V_e$  and  $V_c$  of the two compartments in Expr. 2, we are first forced to introduce asymmetry in the encoding of the aquaporin  $A$ , in order to differentiate the rate  $r_e$  of molecules entering from the rate  $r_c$  of molecules leaving the cell. The system can be considered in equilibrium when the the probability of a water molecule entering is equal to the probability of a water molecule leaving the cell, that is

when

$$h'_e r_e = h'_c r_c \quad (4)$$

where  $h'_e$  is the number of possible combinations of aquaporin–water molecule outside,  $h'_c$  the possible combinations inside the cell. Since only one aquaporin is present, we have that  $h_e = m'_1$  and  $h_c = m'_2$ , where  $m'_1$  and  $m'_2$  are the number of water molecules at equilibrium. Hence Expr. 4 becomes

$$m'_2/m'_1 = r_e/r_c$$

meaning that the equilibrium depends on the rate of the molecules initially conveyed, which is not true. In reality, under the hypothesis of uniform temperature and pressure, the equilibrium is reached when the concentration of the two solutions is the same, that is when

$$m'_2/n'_2 = m'_1/n'_1 \quad (5)$$

where  $n'_1$  and  $n'_2$  are the number of salt ions in the respective compartments. Conversely, this dependence is coherently modeled if we consider the right expressions for  $V_e$  and  $V_c$ . In fact, by Expr. 2, we have

$$V_e^{-1} h'_e r_e = V_c^{-1} h'_c r_c \quad (6)$$

Since the volumes are obtained as the sum of the average volume occupied by each element, we have

$$\begin{aligned} V_e &= \sum_e v_{H_2O} + \sum_e v_S = m'_1 v_{H_2O} + n'_1 v_S \\ V_c &= \sum_c v_{H_2O} + \sum_c v_S = m'_2 v_{H_2O} + n'_2 v_S \end{aligned}$$

which – under the initial symmetric assumption  $r_e = r_c$  – properly substituted in 6 leads to

$$\frac{1}{v_{H_2O} + \frac{n'_1}{m'_1} v_S} = \frac{1}{v_{H_2O} + \frac{n'_2}{m'_2} v_S}$$

that is clearly satisfied by equation 5.

Osmosis involves most of living cells, and expression (5) describes only one of the possible equilibrium conditions. For example, plant cells are surrounded by rigid walls which prevent them from increasing their volume. Consequently, if placed in hypotonic solution, these cells absorb water until the pressure on cell walls equals the osmotic pressure, which depends on the absolute temperature  $T$  and the difference of salt ions/molecules concentration. This equilibrium condition cannot be expressed in  $S\pi@$ , since any pressure evaluation lacks. Such kind of informations may be introduced in the model by defining reaction rates as functions

of the absolute temperature  $T$  and pressure  $p_c$  of the compartment, where  $p_c$  may be in turn calculated as function of the compartment  $c$ , the absolute temperature and the elements surrounded by  $c$ . This would allow to take into account temperature, pressure and some of the structural informations pertaining mechanical properties of compartment boundaries. Depending on the kind of function used for evaluating  $p_c$ , the computational complexity of the algorithm may grow significantly.

## 5 Related Work

Extensions of the SSA handling varying volumes were already considered in [12, 10]. The chosen approach consists in the expression of the propensity functions as known, time-dependent functions. While this approach provides faster simulations, it can be applied only in the case that the variation is deterministic and it is known in advance. Conversely, the MSSA allows to introduce the variation of volumes in the *discrete* and *stochastic* behaviour of the system, in perfect agreement with the intention of the original SSA.

The next subvolume method (NSM) [8] handles multiple volumes, in order to provide efficient simulation of chemical systems in presence of molecular diffusion. The NSM is thought for a high number of volumes with static structure, of fixed and equal size. Although the NSM is faster than the MSSA, it does not handle dynamical compartments with varying volumes and cannot be directly extended with this aim without losing its computational efficiency.

## 6 Conclusion

The osmosis example shows a simple (and biologically common) situation where the SSA happens not to be faithful if taken “as it is”, since the volume of each element (and not only compartment volumes) must be properly considered during the simulation in order to obtain the correct values for reaction rates in function of reactants concentration. The reason is that SSA already embodies the unique, fixed-size volume hypothesis, which is correct for fluid systems under the conditions stated in [9] and needs to be properly translated into another chemical or biological context. The presented adaptation of the SSA to multiple compartments allows to take in account the informations on the concentration of reactants in function of volumes without changing the underlying probabilistic/kinetic model, so that the MSSA results

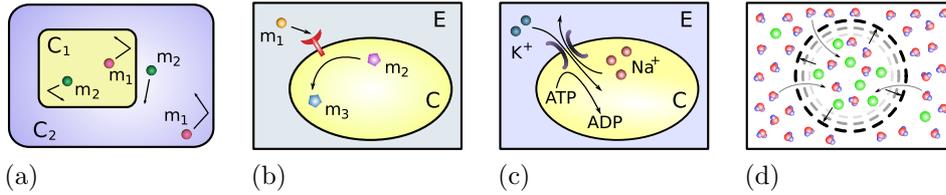


Figure 1: (a) Two equivalent molecule pairs floating in compartments of different volume. (b) Membrane receptor. (c) Ion Pump. (d) Osmosis by semipermeable membrane in hypotonic solution.

somehow closer to the biological scenario. Anyway,  $S\pi@$  modelling is still far from being complete w.r.t. biological reality, for a number of reasons. Here are some:

- temperature and pressure are considered constant and uniform, reasonable condition in most (but not all) of the biological situations, as discussed in the next section;
- many atomic-level phenomena (e.g. van der Waals or electrostatic effects, hydrogen bonds, ...) are disregarded in the kinetic model;
- all the particles are considered to be small, free floating and homogeneously distributed in their respective compartments while many biological elements are not small, neither homogeneously distributed nor allowed (or only partially) to move freely;
- all the physical and dimensional informations on biological elements and structures (e.g. properties like tridimensional position or shape, or surface area, fluidity, thickness of membranes, mechanical resistance, elasticity, ...) cannot be modeled.

In addition,  $S\pi@$  inherits two significant limits – strict seriality and not more than two reactants (but unlimited products) for each reaction, which are due to the point-to-point, interleaving synchronisation semantics of the  $\pi$ -Calculus.

On the other hand the presented approach has several advantages. The most important is the capability of modelling dynamical compartment structures. This can be achieved by preserving several compartment semantics, like for example atonality of BioAmbients [20] or bitonality of Brane Calculi [3]. In fact, both these calculi can be reproduced by adapting the encodings presented in [23] to  $S\pi@$ .

Despite of the increased expressiveness, the calculus is very conservative. The original kinetic/stochastic hypotheses of Gillespie’s model are

completely preserved and  $S\pi@$  semantics is very close to that of the stochastic  $\pi$ -Calculus. This allows to obtain a correct  $S\pi@$  machine by simple extension of SPiM and also to easily adapt the related graphical representation [17].

Beyond the fact that SPiM [16] programs can be directly executed on  $S\pi@$  by retaining the same semantics, the additional informations pertaining compartment and volume can be inserted only when necessary and only for the desired elements and do not increase the computational complexity of the original Gillespie algorithm. Fixed-size or variable volume, single or multiple compartments can be modeled but reaction rates have a unique representation, independent of (but coherent with the volume of) the compartment where the reaction happens. In the same way elements can be defined once and easily instantiated in any compartment, or moved across without readjusting their behaviour, typical feature only of calculi with explicit compartment semantics like BioAmbients or Brane.

The inclusion of further physical properties (like temperature and pressure) and the parallelisation of the algorithm in the way of [4] by Gillespie’s tau leaping [1] are left for future work.

## References

- [1] Y. Cao, D. T. Gillespie, and L. R. Petzold. Efficient step size selection for the tau-leaping simulation method. *Journal of Chemical Physics*, 124(4), 2006.
- [2] M. Carbone and S. Maffei. On the expressive power of polyadic synchronisation in pi-calculus. *Nord. J. Comput.*, 10(2):70–98, 2003.
- [3] L. Cardelli. Brane calculi. In Danos and Schächter [7], pages 257–278.
- [4] P. Cazzaniga, D. Pescini, D. Besozzi, and G. Mauri. Tau leaping stochastic simulation method in p systems. In H. J. Hoogeboom,

- G. Paun, G. Rozenberg, and A. Salomaa, editors, *Workshop on Membrane Computing*, volume 4361 of *Lecture Notes in Computer Science*, pages 298–313. Springer, 2006.
- [5] P. Cazzaniga, D. Pescini, F.-J. Romero-Campero, D. Besozzi, and G. Mauri. Stochastic approaches in P systems for simulating biological systems. In M. A. Gutiérrez-Naranjo, G. Paun, A. Riscos-Núñez, and F. J. Romero-Campero, editors, *Fourth Brainstorming Week on Membrane Computing, Sevilla, January 30 - February 3, 2006. Volume I*, pages 145–164. Fénix Editora, 2006.
- [6] R. Cleaveland, G. Lüttgen, and V. Natarajan. Priority in process algebra. In J. Bergstra, A. Ponse, and S. Smolka, editors, *Handbook of Process Algebra*, pages 711–765. Elsevier Science Publishers, February 2001.
- [7] V. Danos and V. Schächter, editors. *Computational Methods in Systems Biology, International Conference CMSB 2004, Paris, France, May 26-28, 2004, Revised Selected Papers*, volume 3082 of *Lecture Notes in Computer Science*. Springer, 2005.
- [8] J. Elf and M. Ehrenberg. Spontaneous separation of bi-stable biochemical systems into spatial domains of opposite phases. *Systems Biology*, 1(2):230–236, 2004.
- [9] D. T. Gillespie. Exact stochastic simulation of coupled chemical reactions. *J. Phys. Chem.*, 81(25):2340–2361, 1977.
- [10] P. Lecca. A time-dependent extension of gillespie algorithm for biochemical stochastic  $\pi$ -calculus. In *SAC '06: Proceedings of the 2006 ACM symposium on Applied computing*, pages 137–144, New York, NY, USA, 2006. ACM Press.
- [11] H. Lodish, A. Berk, P. Matsudaira, C. A. Kaiser, M. Krieger, M. P. Scott, L. Zipursky, and J. Darnell. *Molecular Cell Biology*. W. H. Freeman, 2004.
- [12] T. Lu, D. Volfson, L. Tsimring, and J. Hasty. Cellular growth and division in the gillespie algorithm. In *Systems Biology, IEE Proceedings*, pages 121–128, 2004.
- [13] R. Milner. *Communicating and mobile systems: the  $\pi$ -calculus*. Cambridge University Press, New York, NY, USA, 1999.
- [14] R. Milner, J. Parrow, and D. Walker. A calculus of mobile processes, i. *Inf. Comput.*, 100(1):1–40, 1992.
- [15] R. Milner, J. Parrow, and D. Walker. A calculus of mobile processes, ii. *Inf. Comput.*, 100(1):41–77, 1992.
- [16] A. Phillips and L. Cardelli. A correct abstract machine for the stochastic pi-calculus. In *Bioconcur'04*. ENTCS, August 2004.
- [17] A. Phillips, L. Cardelli, and G. Castagna. A graphical representation for biological processes in the stochastic pi-calculus. 4230:123–152, 2006.
- [18] C. Priami and P. Quaglia. Beta binders for biological interactions. In Danos and Schächter [7], pages 20–33.
- [19] C. Priami, A. Regev, E. Y. Shapiro, and W. Silverman. Application of a stochastic name-passing calculus to representation and simulation of molecular processes. *Inf. Process. Lett.*, 80(1):25–31, 2001.
- [20] A. Regev, E. M. Panina, W. Silverman, L. Cardelli, and E. Y. Shapiro. Bioambients: an abstraction for biological compartments. *Theor. Comput. Sci.*, 325(1):141–167, 2004.
- [21] A. Regev, W. Silverman, and E. Y. Shapiro. Representation and simulation of biochemical processes using the pi-calculus process algebra. In *Pacific Symposium on Biocomputing*, pages 459–470, 2001.
- [22] C. Versari. Compartments in the stochastic  $\pi$ -calculus with polyadic synchronisation. Available at <http://www.cs.unibo.it/~versari/files/cversari-spiat07.pdf>, 2007.
- [23] C. Versari. A core calculus for a comparative analysis of bio-inspired calculi. *European Symposium on Programming*, 2007. Available at <http://www.cs.unibo.it/~versari/files/cversari-esop07.pdf>.