

## RESEARCH OUTPUTS / RÉSULTATS DE RECHERCHE

### **The stochastic evolution of a protocell. The Gillespie algorithm in a dynamically varying volume.**

Carletti, Timoteo; Filisetti, Alessandro

*Published in:*

Computational and Mathematical Methods in Medicine

*Publication date:*

2012

*Document Version*

Early version, also known as pre-print

[Link to publication](#)

*Citation for published version (HARVARD):*

Carletti, T & Filisetti, A 2012, 'The stochastic evolution of a protocell. The Gillespie algorithm in a dynamically varying volume.', *Computational and Mathematical Methods in Medicine*, vol. 423627, pp. 1-13.

#### **General rights**

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal ?

#### **Take down policy**

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

# THE STOCHASTIC EVOLUTION OF A PROTOCELL. THE GILLESPIE ALGORITHM IN A DYNAMICALLY VARYING VOLUME.

T. CARLETTI AND A. FILISSETTI

ABSTRACT. In the present paper we propose an improvement of the Gillespie algorithm allowing us to study the time evolution of an ensemble of chemical reactions occurring in a varying volume, whose growth is directly related to the amount of some specific molecules, belonging to the reactions set.

This allows us to study the stochastic evolution of a protocell, whose volume increases because of the production of container molecules. Several protocells models are considered and compared with the deterministic models.

## 1. INTRODUCTION

All known life forms are composed of basic units called *cells*; this holds true from the single-cell prokaryote bacterium to the highly sophisticated eucaryotes, whose existence is the result of the coordination, in term of self-organization and emergence, of the behavior of each single basic unit.

While present day cells are endowed with highly sophisticated regulatory mechanisms, that represent the outcome of almost four billion-years of evolution, it is generally believed that the first life-forms were much simpler. Such primordial life-bricks, the *protocells*, were most probably exhibiting only few simplified functionalities, that required a primitive embodiment structure, a protometabolism and a rudimentary genetics, so to guarantee that offsprings were “similar” to their parents [1, 15, 17].

Intense research programs are being established aiming at obtaining protocells capable of growth and duplication, endowed with some limited form of genetics [12, 13, 14, 17]. Despite all efforts, artificial protocells have not yet been reproduced in laboratory and it is thus extremely important to develop reference models [3, 10, 14, 16] that capture the essence of the first protocells appeared on Earth and enable to monitor their subsequent evolution. Due to the uncertainties about the details, high-level abstract models are particularly relevant. Quoting Kaneko [7] it is necessary to consider “simplified models able to capture universal behaviors, without carefully adding complicating details”.

Most of the models present in the literature are based on deterministic differential equations governing the evolution of the concentrations of the involved reacting molecules. Even if the results are worth discussing and provide important insights, it should be stressed that the former assumptions are rarely satisfied in a cell [5]. Firstly, the number of involved molecules is small and should be counted by integer numbers, hence the use of the concentrations can be questioned; secondly, the presence of the thermal noise introduces in the system a degree of stochasticity

than cannot be trivially encoded by a differential equation, mostly because this makes the time evolution a stochastic process. One possible way to overcome such difficulties is to use the Chemical Master equation: given the present state of the system, namely the number of available molecules for each species, and the possible reactions among them, one can compute the transition probabilities to reach and leave the given state and thus get a partial differential equation describing the time evolution of the probability distribution of having a given number of molecules at any future times [5, 6].

Analytically solving the resulting equation is normally a very hard task, one should thus resort to use numerical methods. A particularly suitable one is the algorithm presented by Gillespie [5, 6], allowing to determine, as a function of the present state of the system, the most probable reaction and the most probable reaction time, i.e. the time at which such reaction will occur.

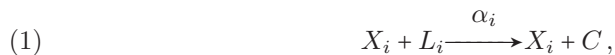
Let us however observe that in the setting we are hereby interested in, the chemical reactions occur in a varying volume, because of the protocell growth; we thus need to adapt the Gillespie method to account for this factor. To the best of our knowledge, there are in the literature very few papers dealing with the Gillespie algorithm in a varying volume [8, 9]. Moreover in all these papers, the volume variation can be considered as an exogenous factor, not being directly related to the number of lipids forming the protocell membrane. So our main contribution is to improve the Gillespie algorithm taking into account the protocell varying volume which is moreover consistent with the increase of the number of lipids constituting the protocell membrane.

The paper is organized as follow. In Section 2 we briefly recall the Surface Reaction Models of protocell, that would be used to compare our stochastic numerical scheme. Then in Sections 3 and 4 we will present our implementation of the Gillespie algorithm in a dynamically varying volume. Finally in Section 5 we will present some applications of our method.

## 2. SURFACE REACTION MODELS

Among the available models for protocells, a particularly interesting one is the Surface Reactions Model [3, 16], SRM for short, and its applicability to the synchronization problem. Such model is roughly inspired by the Los Alamos bug hypothesis [14, 15] but which, due to its abstraction level, the SRM can be applied to a wider set of protocell hypotheses.

The SRM is build on the assumption that a protocell should comprise at least one kind of “container” molecule (typically a lipid or amphiphile), hereby called  $C$  molecule, and one kind of replicator molecule - loosely speaking “genetic material”, hereafter called Genetic Memory Molecule, GMM for short, and named with the letter  $X$ . There are therefore two kinds of reactions which are crucial for the working of the protocell: those which synthesize the container molecules Eq. (1) and those which synthesize the GMM replicators Eq. (2)



In both cases  $L_i$  and  $P_j$  are the buffered precursors, respectively of container molecules and of the  $j$ -th GMM, while  $\alpha_i$  and  $M_{ij}$  are the reaction kinetic constants.

A second main assumption of the SRM, is that such reactions occur on the *surface* of the protocell, exposed to the external medium where precursors are free to move. Hence, as long as container molecules are produced, they are incorporated in the membrane that thus increases its size, until a critical point at which, due to physical instabilities, the membrane splits and two offsprings are obtained, each one getting half of the mother's GMMs and whose size is roughly half that of the mother just before the division.

Under the previous assumptions and in the deterministic setting, one can prove [3, 16] that the number of membrane molecules and the number of GMMs evolve in time according to:

$$(3) \quad \begin{cases} \dot{C} &= \left(\frac{C}{\rho}\right)^{\beta-1} \bar{\alpha} \cdot \bar{X} \\ \dot{\bar{X}} &= C^{\beta-1} M \cdot \bar{X}, \end{cases}$$

where  $\bar{X} = (X_1, \dots, X_N)$  represents the amount of each GMM,  $\bar{\alpha} = (\alpha_1, \dots, \alpha_N)$  is the vector of the reaction constants responsible for the production of  $C$  molecules from the  $X$  molecules plus some appropriate precursor.  $(M_{ij})$  denotes the reaction constant at which  $X_i$  is produced by  $X_j$  plus some precursor.  $\beta \in [2/3, 1]$  is a geometrical shape factor that relates the surface to the volume of the protocell and  $\rho$  is the lipid density (for more details the interested reader can consult [3, 16]). Let us observe that in this setting the precursors are assumed to be buffered and thus their amount to be constant, hence the latter can be incorporated into the constants  $\alpha$  and  $M$ .

So starting with a initial value of container molecules,  $C(t_0) = C_0$ , and of GMMs,  $\bar{X}(t_0) = \bar{X}_0$ , the protocell will grow until some time  $t_0 + \Delta T_1$  at which the amount of  $C$  molecules has doubled with respect to the initial value,  $C(t_0 + \Delta T_1) = 2C_0$  and thus the protocell undergoes a division. Each offspring will get half of the GMMs the mother protocell had just before the division,  $\bar{X}^{(1)} = \bar{X}(t_0 + \Delta T_1)/2$ . And the protocell cycle starts once again. One can prove [3, 16] that under suitable conditions  $\bar{X}^{(n)}$  tends to a constant value once  $n$  goes to infinity, implying thus the emergence of synchronization of growth and information production.

### 3. THE METHOD

Let us now improve the previous scheme by introducing a probabilistic setting *à la Gillespie*. We thus consider a protocell made by a lipidic vesicle and containing a well stirred mixture of  $N$  GMMs,  $X_1, \dots, X_N$ , that may react through  $m$  elementary reaction channels  $R_\mu$ ,  $\mu = 1, \dots, m$ , running within the volume  $V(t)$  of the protocell.

Let us observe that because of the protocell growth the volume is an increasing function of time. Actually one can relate the volume to the amount of container molecules via their density  $V = C/\rho$  where  $C$  denotes the integer number of molecules forming the lipidic membrane. We will hereby use the same symbol  $X_i$  to denote both the  $i$ -th GMM and the integer number of molecules of type  $X_i$  in the system.

For each reaction channel  $R_\mu$  assume that there exists a scalar rate  $c_\mu$  such that  $c_\mu dt + o(dt)$  is the probability that a random combination of molecules from channel  $R_\mu$  will react in the interval  $[t, t + dt)$  within the volume  $V(t)$ .

Let  $h_\mu(Y)$  be the total number of possible distinct combinations of molecules for a channel  $R_\mu$  when the system is in state  $Y = (X_1, \dots, X_N, C)$ , then we can define the *propensity* [9] of the reaction  $R_\mu$  to be  $a_\mu(Y) = h_\mu(Y)c_\mu$ .

One can prove [5] that for a binary reaction the rate  $c_\mu$  can be written in the form  $c_\mu = k_\mu/V$ , where  $k_\mu$  is a fixed constant. Similarly one can prove that for a reaction involving  $n$  different species, we get:  $c_\mu = k_\mu/V^{n-1}$ . And thus for a single molecule reaction, i.e. a decay, we get  $c_\mu = k_\mu$ , namely independently from the volume.

Let us now assume that among the  $m$  reactions,  $Q_1$  involve one single molecule,  $Q_2$  are binary reactions,  $Q_3$  are ternary reactions and so on. Of course  $Q_1 + Q_2 + \dots + Q_{N+1} = m$ . We recall that we have  $N$  GMMs and the container type molecule  $C$ , hence  $N + 1$  species. For short we will denote  $\mathcal{Q}_1$  the set of indices  $\mu$  for mono molecule reactions, and by  $\mathcal{Q}$  the remaining ones. Let us observe that in this way some coefficient  $a_\mu$ , will depend both on the system state  $Y$  and on the time via the volume  $V(t)$ :  $a_\mu(Y, t)$  for  $\mu \in \mathcal{Q}$ .

More precisely to study the time evolution of the system we need to determine the probability  $P_\mu(\tau|Y, t)d\tau$ , that given the system in the state  $Y = (X_1, \dots, X_n, C)$  at time  $t$ , then the next reaction will occur in the infinitesimal time interval  $(t + \tau, t + \tau + d\tau)$  and it will be the reaction  $R_\mu$ . This probability will be computed as

$$(4) \quad P_\mu(\tau|Y, t)d\tau = P_{not}(\tau|Y, t) \times a_\mu(Y, t + \tau)d\tau,$$

where  $P_{not}(\tau|Y, t)$  is the probability that no reaction occurs in  $(t, t + \tau)$  given the state  $Y$  at time  $t$ , whereas the rightmost term denotes the probability to have a reaction  $R_\mu$  in  $(t + \tau, t + \tau + d\tau)$  given the state  $Y$  at time  $t + \tau$ .

To compute the first term  $P_{not}$ , let us take  $s \in [t, t + \tau]$  and observe that:

$$P_{not}(s + ds|Y, t) = P_{not}(s|Y, t)P_{not}(ds|Y, t + s) = P_{not}(s|Y, t) \left( 1 - \sum_{\mu} a_\mu(Y, t + s)ds \right),$$

being  $1 - \sum_{\mu} a_\mu(Y, t + s)ds$  the probability that no reaction will occur in  $(t + s, t + s + ds)$  once we are in state  $Y$  at time  $t + s$ . Thus rewriting the previous difference equation as a differential equation, passing to the limit  $ds \rightarrow 0$ , and observing that  $P_{not}(0|Y, t) = 1$ , we get the solution:

$$(5) \quad P_{not}(\tau|Y, t) = \exp \left[ -A_{Q_1}(Y)\tau - \int_0^\tau A_Q(Y, s + t) ds \right],$$

where

$$A_{Q_1}(Y) = \sum_{\mu \in \mathcal{Q}_1} a_\mu(Y) \quad \text{and} \quad A_Q(Y, s + t) = \sum_{\mu \in \mathcal{Q}} a_\mu(Y, s + t).$$

The apparent asymmetry in the exponential term in (5) is easily recovered by observing that  $A_{Q_1}(Y)\tau = \int_0^\tau A_{Q_1}(Y)ds$ .

We can thus conclude that

$$(6) \quad P_\mu(\tau|Y, t)d\tau = \exp \left[ -A_{Q_1}(Y)\tau - \int_t^{t+\tau} A_Q(Y, s) ds \right] a_\mu(Y, t + \tau)d\tau.$$

Let us observe that the rightmost term is correctly  $a_\mu(Y, t + \tau)$ , namely the system is still in the state  $Y$  at time  $t + \tau$ , because no reaction has been produced in  $(t, t + \tau)$ .

Let us recall that the volume enters in the previous relation via the function  $A_Q$ , more explicitly one has

$$(7) \quad A_Q(Y, s) = \sum_{\mu \in \mathcal{Q}_2} \frac{h_\mu(Y)k_\mu}{V(s)} + \sum_{\mu \in \mathcal{Q}_3} \frac{h_\mu(Y)k_\mu}{(V(s))^2} + \dots + \sum_{\mu \in \mathcal{Q}_{N+1}} \frac{h_\mu(Y)k_\mu}{(V(s))^N},$$

that can be rewritten in terms of  $C$  molecules using the relation  $C = \rho V$ . So our method applies to a different problem with respect to the one considered in [9], in fact in our case the volume growth is not imposed a priori but dynamically evolves accordingly to the reaction scheme, if  $C$  is produced then  $V$  increases otherwise it will keep a constant value, while in [9] the volume growth is an exogenous variable.

#### 4. THE STOCHASTIC SIMULATION ALGORITHM IN A GROWING VOLUME

Once we have the probability function  $P_\mu(\tau|Y, t)$  we can build an algorithm that reproduces the time evolution given by the model defined above.

Given the system in some state  $Y$  at time  $t$ , we must determine the interval of time  $\tau$  and the reaction channel  $R_\mu$  according to the probability distribution function  $P_\mu(\tau|Y, t)$ , and finally update the state  $Y \rightarrow Y + \nu_\mu$ , where  $\nu_\mu$  is a stoichiometric vector representing the increase and decrease of molecular abundance due to the reaction  $R_\mu$ . This will be accomplished following the standard approach by Gillespie [5] but taking care of the time dependence of the propensities. We will thus need to compute the cumulative probability distribution function and then make use of the inversion method [6], to determine the channel  $\mu$  and the next reaction time  $\tau$ , distributed according to  $P_\mu(\tau|Y, t)$ .

From (6) we can compute the *cumulative distribution function*

$$(8) \quad F(\tau|Y, t) = \int_0^\tau \sum_\mu P_\mu(s|Y, t) ds,$$

providing the probability that any reaction will occur in  $(t, t + \tau)$  starting from the state  $Y$  at time  $t$ . The function  $F(\tau|Y, t)$  can be explicitly computed by

**Proposition 4.1.** *Under the above assumptions we have*

$$(9) \quad F(\tau|Y, t) = 1 - \exp \left[ -A_{Q_1}(Y)\tau - \int_t^{t+\tau} A_Q(Y, s) ds \right].$$

*Proof.* The first step is to use (6) and perform a sum over all the channels  $\mu$  to rewrite (8) as

$$F(\tau|Y, t) = \int_0^\tau (A_{Q_1}(Y) + A_Q(Y, t+s)) \exp \left[ -A_{Q_1}(Y)s - \int_t^{t+s} A_Q(Y, r) dr \right] ds.$$

Then we can observe that

$$\begin{aligned} & \frac{\partial}{\partial s} \left( \exp \left[ -A_{Q_1}(Y)s - \int_t^{t+s} A_Q(Y, r) dr \right] \right) = \\ & = - (A_{Q_1}(Y) + A_Q(Y, t+s)) \exp \left[ -A_{Q_1}(Y)s - \int_t^{t+s} A_Q(Y, r) dr \right], \end{aligned}$$

and thus

$$\begin{aligned} F(\tau|Y, t) &= - \int_0^\tau \frac{\partial}{\partial s} \left( \exp \left[ -A_{Q_1}(Y)s - \int_t^{t+s} A_Q(Y, r) dr \right] \right) ds \\ &= 1 - \exp \left[ -A_{Q_1}(Y)\tau - \int_t^{t+\tau} A_Q(Y, r) dr \right]. \end{aligned}$$

□

Once we have the cumulative distribution function we can obtain the value  $\tau$  by drawing a random number  $u_1$  from a uniform distribution in  $[0, 1]$  and then solve with respect to  $\tau$  the implicit equation:

$$(10) \quad u_1 = 1 - \exp \left[ -A_{Q_1}(Y)\tau - \int_t^{t+\tau} A_Q(Y, s) ds \right].$$

Let us stress once again that this is not as straightforward as for the original Gillespie [5] scheme, or the simplified one presented in [9], because of the time dependence of  $A_Q$  via the volume. One can nevertheless find suitable approximations for the integral, this will be the goal of the next sections.

**4.1. The adiabatic assumption.** Let us assume that  $\tau$  is very small, or which is equivalent, that the time scale of the chemical reactions involving the GMMs is much faster than the production of container molecules, hence the volume growth is very slow compared with the production of the chemicals  $X_i$ .

Under this hypothesis one can assume that in the interval  $(t, t + \tau)$  the volume doesn't vary and thus one can make the following approximation

$$(11) \quad \int_t^{t+\tau} A_Q(Y, s) ds \sim A_Q(Y, t)\tau.$$

One can thus explicitly solve equation (10) to get:

$$(12) \quad \tau_{Gill} = -\frac{1}{A_{Q_1}(Y) + A_Q(Y, t)} \log(1 - u_1),$$

that is the standard Gillespie result except now that  $A_Q(Y, t)$  depends on time and as long the volume increases, then the contribution arising from  $A_Q(Y, t)$  might become smaller because  $A_Q \sim 1/V$ .

**4.2. The next order correction.** One can obtain a somehow better estimate valid in the case of comparable time scales for the reactions involving GMM and the container growth. The idea is to compute the integral in Eq. (10) using the following approximation:

$$(13) \quad \begin{aligned} \int_t^{t+\tau} A_Q(Y, s) ds &= \int_0^\tau A_Q(Y, t+s) ds = \int_0^\tau \left( A_Q(Y, t) + \frac{\partial A_Q(Y, t)}{\partial t} s + \dots \right) ds \\ &= A_Q(Y, t)\tau + \frac{\partial A_Q(Y, t)}{\partial t} \frac{\tau^2}{2} + \mathcal{O}(\tau^3). \end{aligned}$$

where  $\partial A_Q(Y, t)/\partial t$  can be obtained using the definition (7) and expressing the volume in terms of  $C = V(t)\rho$ , namely:

$$\frac{\partial A_Q(Y, t)}{\partial t} = -\frac{\dot{C}}{C} \left( \sum_{\mu \in \mathcal{Q}_2} \frac{h_\mu(Y)k_\mu}{C(t)} + 2 \sum_{\mu \in \mathcal{Q}_3} \frac{h_\mu(Y)k_\mu}{(C(t))^2} + \dots + N \sum_{\mu \in \mathcal{Q}_{N+1}} \frac{h_\mu(Y)k_\mu}{(C(t))^N} \right).$$

To compute  $\dot{C}/C$  we make the assumption that in a very short time interval, as the one we are interested in, the deterministic growth of the container is a good approximation for the stochastic underlying mechanism; this implies that we can use (3)

$$\frac{\dot{C}}{C} = \left( \frac{C(t)}{\rho} \right)^{\beta-1} \frac{\tilde{\alpha} \cdot \tilde{X}(t)}{C(t)}.$$

Inserting the previous result into (13) and finally solving (10) with respect to  $\tau$ , we can compute the next reaction time up to correction of the order of  $\tau^3$ , as:

$$(14) \quad \tau_{Gill} = \frac{-(A_{Q_1}(Y) + A_Q(Y, t)) + \sqrt{(A_{Q_1}(Y) + A_Q(Y, t))^2 - 2 \log(1 - u_1) \dot{A}_Q(Y, t)}}{\dot{A}_Q(Y, t)},$$

where we wrote for short  $\dot{A}_Q(Y, t) = \partial A_Q(Y, t)/\partial t$  and we selected the positive square root in such a way in the limit  $\dot{A}_Q(Y, t) \rightarrow 0$  we recover the previous solution (12).

**Remark 4.2** (On the existence of  $\tau_{Gill}$ ). *In the case of variable volume a new phenomenon can arise: the volume growth can be so fast that no reaction can occur in the interval  $(t, t + \tau + d\tau)$  for any  $\tau$ . Mathematically this translates into a sign condition for the term under square root in (14), if:*

$$(15) \quad \log(1 - u_1) < (A_{Q_1}(Y) + A_Q(Y, t))^2 / (2\dot{A}_Q(Y, t)),$$

then equation (10) has no real solution.

This can be geometrically interpreted as follows. The relation (10) determines  $\tau_{Gill}$  as the intersection of the parabola  $-A_{Q_1}(Y) - A_Q(Y, t)\tau - \dot{A}_Q(Y, t)\tau^2/2$  with the horizontal line  $\log(1 - u_1)$ , which is negative because  $u_1 \in (0, 1)$ . Such parabola intersect the y-axis at  $\tau_1 = 0$  and  $\tau_2 = -2(A_{Q_1}(Y) + A_Q(Y, t))/\dot{A}_Q(Y, t) > 0$  and it is concave. Then its absolute (negative) minimum is reached at the vertex  $\tau_V = (\tau_1 + \tau_2)/2$  and its value is  $(A_{Q_1}(Y) + A_Q(Y, t))^2 / (2\dot{A}_Q(Y, t))$  and it is negative because  $\dot{A}_Q(Y, t)$  is negative. Hence if the horizontal line is below this value, i.e. condition (15) is verified, the parabola and the line do not have any real intersections (see Fig. 1).

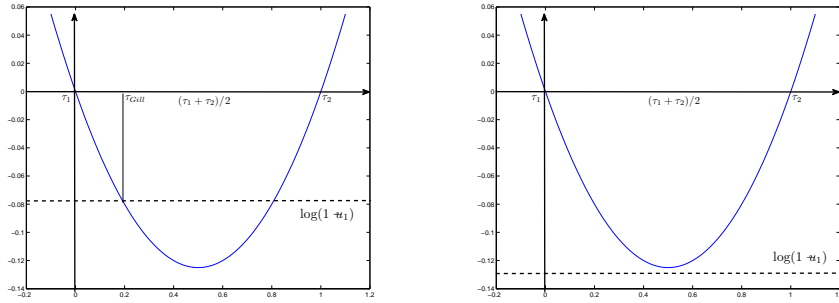


FIGURE 1. Geometrical interpretation of the existence of the next reaction time  $\tau_{Gill}$ . Left panel :  $\tau_{Gill}$  is the smallest intersection between the parabola and the horizontal line  $\log(1 - u_1)$ . Right panel  $\tau_{Gill}$  doesn't exist, the horizontal line is located below the minimum of the parabola.

Let us also observe that, whenever it exists,  $\tau_{Gill}$  is always positive as it should be. In the case of a protocell the non existence of such next reaction time could be translated into the death by dilution of the protocell.

**4.3. The next reaction channel.** Whenever the next reaction time does exist, the next reaction channel is determined using the classical Gillespie method, namely by drawing a second uniformly distributed random number  $u_2 \in [0, 1]$  and fix the channel  $\mu$  such that:

$$(16) \quad \sum_{\nu=1}^{\mu-1} a_{\nu}(Y, t + \tau) \leq u_2 a_0(Y, t + \tau) \leq \sum_{\nu=1}^{\mu} a_{\nu}(Y, t + \tau),$$

where  $a_0(Y, t + \tau) = A_{Q_1}(Y) + A_Q(Y, t + \tau) = \sum_{\nu=1}^m a_{\nu}(Y, t + \tau)$ .

**Remark 4.3.** *Let us observe that if all the reactions involve the same number of chemicals, then the determination of which reaction channel  $\mu$  will be activated in the next reaction, doesn't depend on the volume which factorizes out from (16). In fact assuming all the reactions to involve  $p$  chemical, we obtain by definition*

$$a_{\nu}(Y, t + \tau) = \frac{h_{\nu}(Y)k_{\nu}}{[V(t + \tau)]^p} \quad \forall \nu \in \{1, \dots, m\},$$

and thus (16) rewrites:

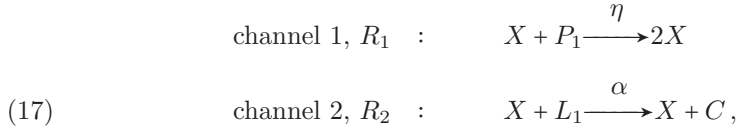
$$\sum_{\nu=1}^{\mu-1} \frac{h_{\nu}(Y)k_{\nu}}{[V(t + \tau)]^p} \leq u_2 \sum_{\nu=1}^m \frac{h_{\nu}(Y)k_{\nu}}{[V(t + \tau)]^p} \leq \sum_{\nu=1}^{\mu} \frac{h_{\nu}(Y)k_{\nu}}{[V(t + \tau)]^p},$$

which is clearly independent of the volume value  $V$ .

## 5. SOME APPLICATIONS

The aim of this section is to provide some applications of the previous algorithm to the study of the evolution of a protocell.

**5.1. One single Genetic Memory Molecule.** The simplest model is the one where only one GMM specie is present in the protocell [16] and thus only two chemical channels are active:



where  $P_1$  and  $L_1$  are, respectively, precursors of GMM, i.e. nucleotide, and precursors of amphiphiles.

One can thus compute the propensities in the state  $Y = (X, C)$  at time  $t$ :

$$(18) \quad a_1(X, C, t) = h_1(X, C) \frac{\eta}{V(t)} = \eta \frac{P_1 X}{V(t)} \quad \text{and} \quad a_2(X, C, t) = h_2(X, C) \frac{\alpha}{V(t)} = \alpha \frac{L_1 X}{V(t)};$$

let us observe that we assume that precursors are buffered and thus they are constant.

Because system (17) contains only bimolecular reactions, all the propensities are time dependent, hence  $A_{Q_1} = 0$  and  $A_Q = a_1(X, C, t) + a_2(X, C, t) = (P_1 \eta + L_1 \alpha) X / V(t)$ , thus (10) simplifies into

$$u_1 = 1 - \exp \left[ - \int_t^{t+\tau} A_Q(Y, s) ds \right],$$

whose second order solution (14) is given by

$$\tau_{Gill} = \frac{-A_Q(Y, t) + \sqrt{(A_Q(Y, t))^2 - 2 \log(1 - u_1) \dot{A}_Q(Y, t)}}{\dot{A}_Q(Y, t)},$$

and

$$\frac{\partial A_Q(X, C, t)}{\partial t} = -\frac{\dot{V}(t)}{V(t)} \left( \frac{P_1 \eta X}{V(t)} + \frac{L_1 \alpha X}{V(t)} \right) \Big|_{V(t)=C(t)/\rho} = -\left( \frac{C}{\rho} \right)^{\beta-1} \frac{\rho L_1 \alpha X^2}{C^2} (P_1 \eta + L_1 \alpha).$$

So we can finally obtain

$$\tau_{Gill} = \frac{C}{L_1 \alpha X} \left( \frac{\rho}{C} \right)^{\beta-1} - \sqrt{\left[ \frac{C}{L_1 \alpha X} \left( \frac{\rho}{C} \right)^{\beta-1} \right]^2 + 2 \frac{C^2}{L_1 \alpha \rho X^2 (P_1 \eta + L_1 \alpha)} \log(1 - u_1)},$$

provided

$$\log(1 - u_1) \geq -\frac{\rho}{2\alpha} \left( \frac{\rho}{c} \right)^{2(\beta-1)} (P_1 \eta + L_1 \alpha).$$

Which reaction channel  $\mu$  will active in the time interval  $[t, t+\tau]$  can be obtained according to :

$$\begin{aligned} & \text{if } u_2 \frac{(P_1 \eta + L_1 \alpha) X}{V} \leq \frac{P_1 \eta X}{V} \quad \text{namely } 0 \leq u_2 \leq \frac{P_1 \eta}{P_1 \eta + L_1 \alpha} \quad \text{then } \mu = 1 \\ \text{if } \frac{P_1 \eta X}{V} < u_2 \frac{(P_1 \eta + L_1 \alpha) X}{V} \leq \frac{(P_1 \eta + L_1 \alpha) X}{V} & \quad \text{namely } \frac{P_1 \eta}{P_1 \eta + L_1 \alpha} < u_2 \leq 1 \quad \text{then } \mu = 2. \end{aligned}$$

Let us observe that according to remark 4.3, the choice of  $\mu$  doesn't depend on the volume, because only binary reactions are present.

Let  $C_0$  be the initial amount of container molecules, then we assume that once  $C(t) = 2C_0$  the protocell splits into two offspring, almost halving the GMM amount. More precisely we assume that the first offspring will get a number of GMMs drawn according to a Binomial distribution with parameter  $p = 1/2$  and  $n = X(t)$ . From this step, for technical reason, only one randomly chosen offspring will be studied during each generation.

In Fig. 2 we report a comparison between the deterministic (3) and the stochastic dynamics, under the adiabatic assumption for  $\tau_{Gill}$ , corresponding to the continuous growth phase of the container between two successive divisions. As one should expect, a system composed by a large number of molecules exhibits small stochastic fluctuations whose average is not too far from the dynamics described by the deterministic model.

In Fig. 3 we report the amount of GMM,  $X^{(k)}$  (panel a), at the beginning of each protocell cycle and the duplication time (panel b), namely the interval of time needed to double the amount of  $C$  molecules, for both the stochastic and deterministic models. Once again one can clearly observe the small fluctuations of the stochastic system around the value obtained by the numerical integration of the deterministic description, Eq. (3). Let us observe that these fluctuations are due to the stochastic integrator scheme and also on the division mechanism.

We are now interested in studying the fluctuations dependence on the amount of molecules. We already know that for a sufficiently large number of molecules the stochastic dynamics follows closely the deterministic one and thus the fluctuations are small. On the other hand, one should expect that when the number of molecules

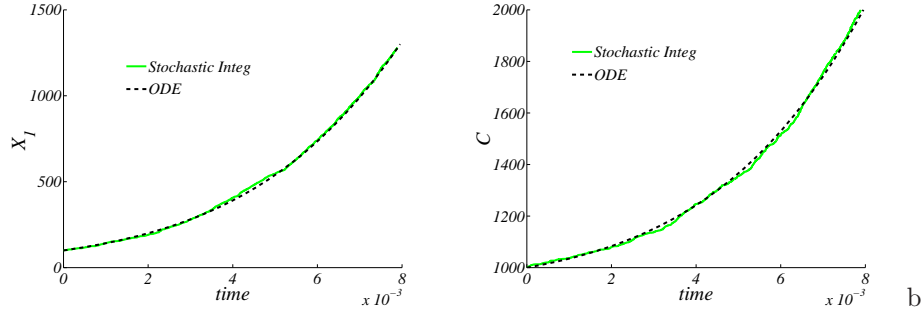


FIGURE 2. Stochastic vs ODE SRM protocell (3). Case of one GMM, left panel the time evolution of the amount of GMM, right panel the time evolution of the amount of C. Parameters are :  $\eta = 1$ ,  $\alpha = 1$ ,  $L1 = 500$ ,  $P1 = 600$ ,  $X_1(0) = 100$ ,  $C(0) = 1000$ ,  $\rho = 200$  and  $\beta = 2/3$ .

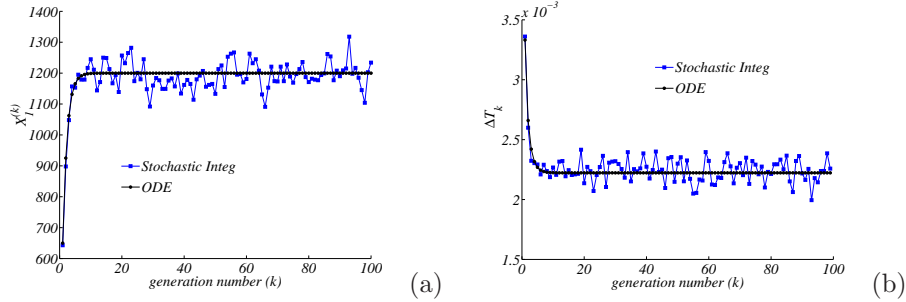


FIGURE 3. Stochastic vs ODE SRM protocell (3). Case of one GMM, left panel the amount of GMM at the beginning of each division cycle, right panel the division time as a function of the number of elapsed divisions. Parameters are :  $\eta = 1$ ,  $\alpha = 1$ ,  $L1 = 500$ ,  $P1 = 600$ ,  $X_1(0) = 100$ ,  $C(0) = 1000$ ,  $\rho = 200$  and  $\beta = 2/3$ .

decreases, then the fluctuation will rise and the system behavior could not be completely described by means of a deterministic approach. This is confirmed by Fig. 4 and Fig. 5, where we can observe that a model composed by a small number of initial molecules, 20 times lesser than in the model presented in Fig. 2 exhibits larger stochastic fluctuations.

In Fig. 6 we summarize the results of several protocell models each one with a different amount of initial molecules, in order to appreciate the influence of the latter on the stochastic fluctuations. To compare with, we also report the case of the deterministic model. Because the kinetic constants are kept constant, the analytical theory for the deterministic model ensures that the division time doesn't vary [3]. Nevertheless the fewer is the initial amount of  $X_0$  and  $C_0$ , the larger are the fluctuations present in the stochastic integration.

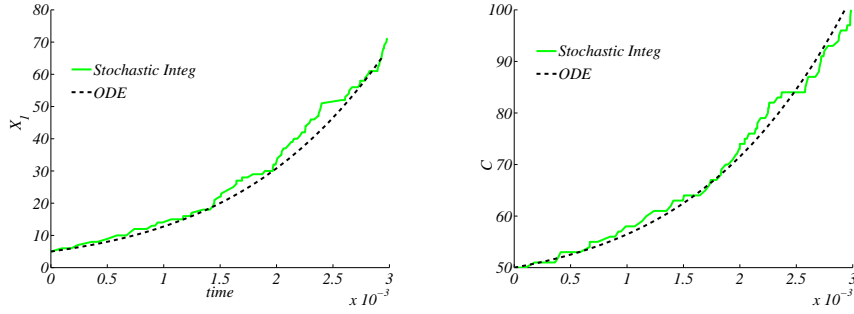


FIGURE 4. Stochastic vs ODE SRM protocell (3). Case of one GMM, left panel the time evolution of the amount of GMM, right panel the time evolution of the amount of  $C$ . Parameters are :  $\eta = 1$ ,  $\alpha = 1$ ,  $L1 = 500$ ,  $P1 = 600$ ,  $X_1(0) = 5$ ,  $C(0) = 50$ ,  $\rho = 200$  and  $\beta = 2/3$ .

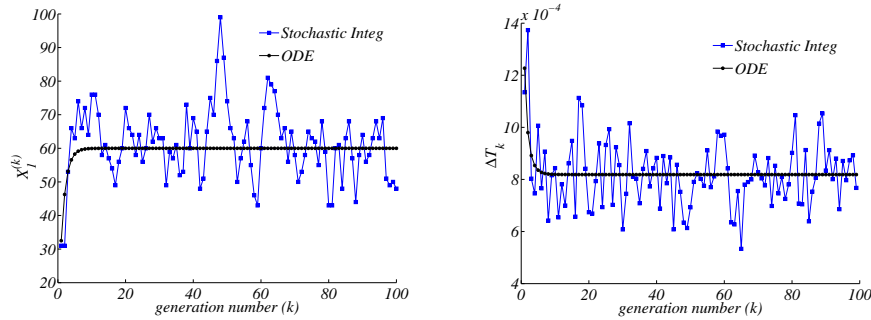


FIGURE 5. Stochastic vs ODE SRM protocell (3). Case of one GMM, left panel the amount of GMM at the beginning of each division cycle, right panel the division time as a function of the number of elapsed divisions. Parameters are :  $\eta = 1$ ,  $\alpha = 1$ ,  $L1 = 500$ ,  $P1 = 600$ ,  $X_1(0) = 5$ ,  $C(0) = 50$ ,  $\rho = 200$  and  $\beta = 2/3$ .

To get a more complete understanding of the fluctuations dependence, we decided to measure them using the standard deviation of the protocell division time (after a sufficiently long transient phase). In Fig. 7 we report the standard deviation of the division time  $\Delta T$  as a function of the initial amount of molecules. As expected the fluctuations strength decreases rapidly as soon as the number of molecules increases and the relation can be very well approximated by a power law distribution with exponent  $-0.54 \pm 0.03$  (linear best fit).

**5.2. Two non-interacting Genetic Memory Molecules.** A slightly more sophisticated model can be obtained by considering two linear non interacting GMMs. The system can be described by the following chemical reactions:

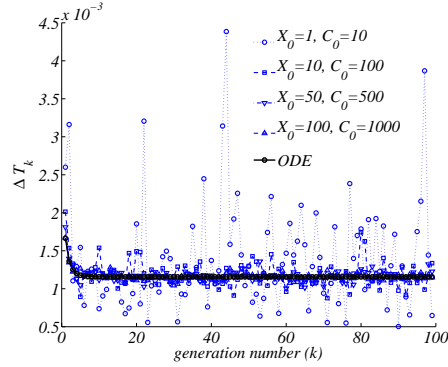


FIGURE 6. Fluctuation dependence on the initial conditions. We report the division times as a function of the number of elapsed divisions, for 5 different protocells models. Protocell  $\circ$ :  $X_1(0) = 5$ ,  $C(0) = 10$ , protocell  $\square$ :  $X_1(0) = 10$ ,  $C(0) = 100$ , protocell  $\nabla$ :  $X_1(0) = 50$ ,  $C(0) = 500$ , protocell  $\triangle$ :  $X_1(0) = 100$ ,  $C(0) = 1000$ . The black line denotes the deterministic protocell. All the remaining parameters have been fixed to:  $\eta = 1$ ,  $\alpha = 1$ ,  $L1 = 500$ ,  $P1 = 600$ ,  $\rho = 100$  and  $\beta = 1$ .

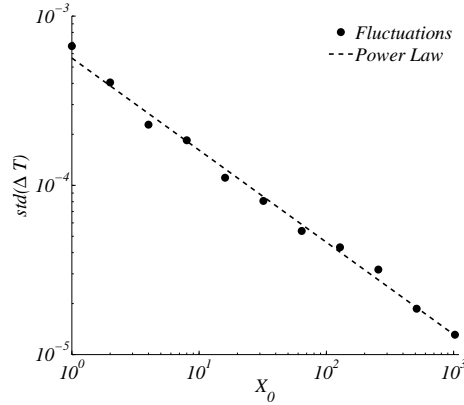
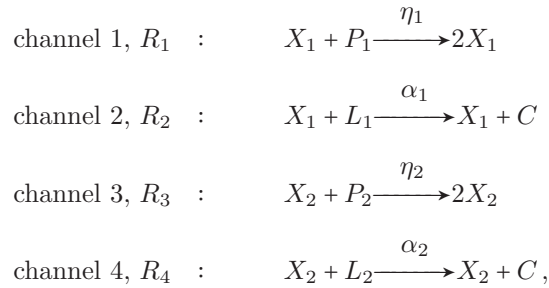


FIGURE 7. Fluctuation dependence on the initial conditions. We report the standard deviation of the protocell division time as a function of the initial amount of molecules  $X_0$  ( $\bullet$ ) and a linear best fit, whose slope is  $= -0.54 \pm 0.03$ . Parameters are:  $X(0) = 2^n$  with  $n = 0, \dots, 10$ ,  $C(0) = 10X(0)$ ,  $\eta = 1$ ,  $\alpha = 1$ ,  $L1 = 500$ ,  $P1 = 600$ ,  $\rho = 100$  and  $\beta = 1$ .



where  $P_i$  and  $L_i$  are, respectively, precursors of the  $i$ -th GMM, i.e. nucleotide, and precursors of amphiphiles used by the  $i$ -th GMM to build a  $C$  molecule.

As previously done, we compare the stochastic and the deterministic models. Results are reported in Figure 8 and one can still observe that in presence of a large number of molecules the deterministic dynamics well approximates the stochastic model. On the other hand, the protocell division time exhibits large fluctuations around the deterministic value even in presence of a quite large number of molecules (see right panel Fig. 9).

The parameters have been set in such a way only one GMM will survive according to the analytical theory for the deterministic model. One can observe that, despite the fluctuations, the same fate is obtained for the stochastic model (see right panel Fig. 9).

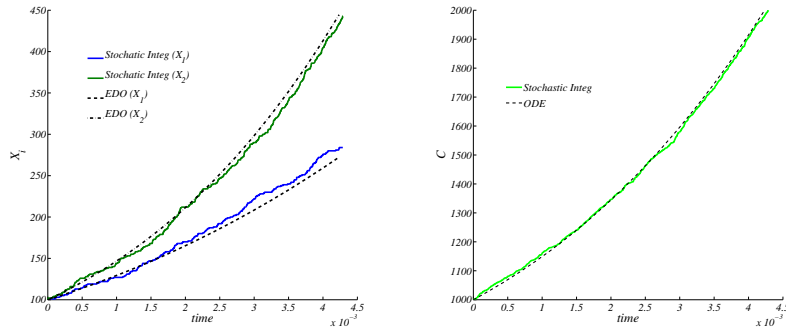


FIGURE 8. Stochastic vs ODE SRM protocell (3). Case of two GMMs, left panel the time evolution of the amount of GMM during a division cycle, right panel the time evolution of the amount of  $C$  molecules. Parameters are :  $\eta_1 = \eta_2 = 1$ ,  $\alpha_1 = \alpha_2 = 2$ ,  $L_1 = 500$ ,  $L_2 = 600$ ,  $P_1 = 600$ ,  $P_2 = 670$ ,  $X_1(0) = X_2(0) = 100$ ,  $C(0) = 1000$ ,  $\rho = 200$  and  $\beta = 2/3$ .

Once we reduce the number of involved molecules, the stochastic fluctuations dramatically increase (see Fig. 10 and Fig. 11).

As in the case of only one GMM, when two non interacting linear GMMs are present the size of the stochastic fluctuations as a function of the initial number of molecules follows a power law distribution with exponent  $-0.51 \pm 0.05$  (linear best fit), see Fig. 12: the fewer are the molecules in the system, the larger are the fluctuations around the deterministic dynamics.

A new phenomenon arises in the case of two GMMs modeled by a stochastic process. There can be a *breaking of the symmetry* emerging in systems composed of two identical GMMs (i.e equal kinetic constants, equal initial amounts and availability of precursors) present with a few initial amount of each one. Although adopting a deterministic approach the dynamics of the two replicators would be perfectly the

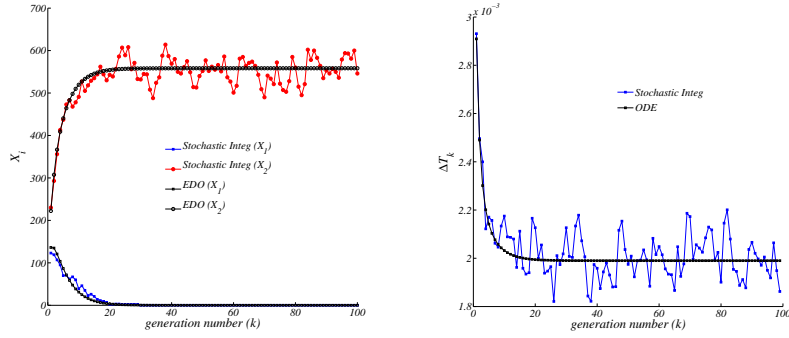


FIGURE 9. Stochastic vs ODE SRM protocell (3). Case of two GMMs, left panel the amount of GMM at the beginning of each division cycle, right panel the division time as a function of the number of elapsed divisions. Parameters are :  $\eta_1 = \eta_2 = 1$ ,  $\alpha_1 = \alpha_2 = 2$ ,  $L_1 = 500$ ,  $L_2 = 600$ ,  $P_1 = 600$ ,  $P_2 = 670$ ,  $X_1(0) = X_2(0) = 100$ ,  $C(0) = 1000$ ,  $\rho = 200$  and  $\beta = 2/3$ .

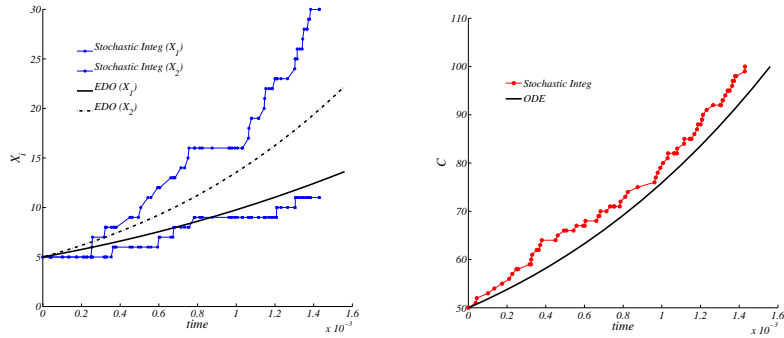


FIGURE 10. Stochastic vs ODE SRM protocell (3). Case of two GMMs, left panel the time evolution of the amount of GMM during a division cycle, right panel the time evolution of the amount of  $C$  molecules. Parameters are :  $\eta_1 = \eta_2 = 1$ ,  $\alpha_1 = \alpha_2 = 2$ ,  $L_1 = 500$ ,  $L_2 = 600$ ,  $P_1 = 450$ ,  $P_2 = 670$ ,  $X_1(0) = X_2(0) = 5$ ,  $C(0) = 50$ ,  $\rho = 200$  and  $\beta = 2/3$ .

same, a small fluctuation in the very first instants of the protocell evolution entails the dilution of one of the two replicators and thus a different fate for the protocell. Let us observe that the probability to have a large fluctuation is never zero, thus waiting for a sufficiently long time, a specie can always disappear from the system and thus giving rise to the the breaking of the symmetry phenomenon. See Fig. 13 where we report, as a function of the initial amount of molecules  $X_i(0)$ ,  $i = 1, 2$ , the proportion of simulations where the symmetry breaking has been observed repeating 50 times each simulation with the same set of parameters and initial conditions during 100 generations.

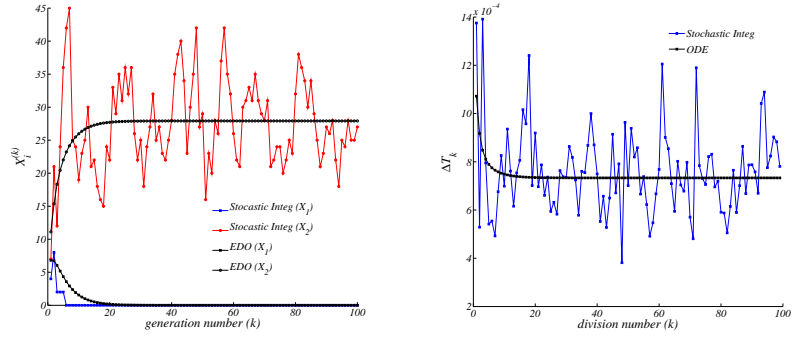


FIGURE 11. Stochastic vs ODE SRM protocell (3). Case of two GMMs, left panel the amount of GMM at the beginning of each division cycle, right panel the division time as a function of the number of elapsed divisions. Parameters are :  $\eta_1 = \eta_2 = 1$ ,  $\alpha_1 = \alpha_2 = 2$ ,  $L_1 = 500$ ,  $L_2 = 600$ ,  $P_1 = 450$ ,  $P_2 = 670$ ,  $X_1(0) = X_2(0) = 5$ ,  $C(0) = 50$ ,  $\rho = 200$  and  $\beta = 2/3$ .

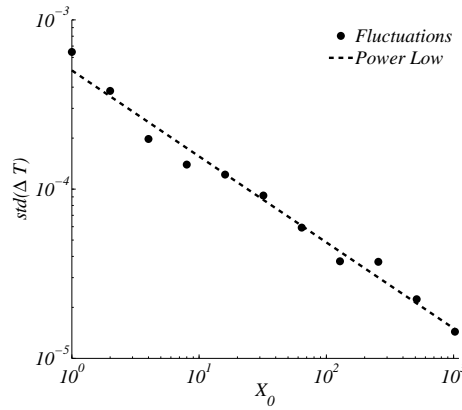


FIGURE 12. Fluctuation dependence on the initial conditions. We report the standard deviation of the protocell division time as a function of the initial amount of molecules  $X_i(0)$ ,  $i = 1, 2$ , ( $\bullet$ ) and a linear best fit, whose slope is  $= -0.51 \pm 0.05$ . Parameters are:  $X_1(0) = X_2(0) = 2^n$  with  $n = 0, \dots, 10$ ,  $C(0) = 10X_1(0)$ ,  $\eta_1 = \eta_2 = 1$ ,  $\alpha_1 = \alpha_2 = 2$ ,  $L_1 = 500$ ,  $L_2 = 500$ ,  $P_1 = 500$ ,  $P_2 = 600$ ,  $\rho = 100$  and  $\beta = 1$ .

## 6. CONCLUSION

In this paper we presented a new stochastic integration algorithm based on the one introduced by Gillespie. Our contribution is devoted to the explicit introduction of the volume variation in the algorithm, which moreover is directly related to the amount of contained molecules, and thus it evolves in a self-consistent way.

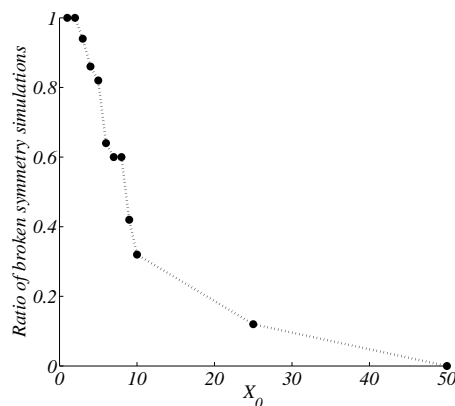


FIGURE 13. Symmetry breaking phenomenon. Each point denotes the fraction of runs exhibiting the symmetry breaking phenomenon, during 100 generations, over 50 identical replicas. Parameters are:  $X_1(0) = X_2(0) = [1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 25, 50]$ ,  $C(0) = 10X$ ,  $\eta_1 = \eta_2 = 1$ ,  $\alpha_1 = \alpha_2 = 2$ ,  $L1 = 500$ ,  $L2 = 500$ ,  $P1 = 600$ ,  $P_2 = 600$ ,  $\rho = 100$  and  $\beta = 1$ .

This algorithm straightforwardly adapts to the study of the evolution of a protocell, simplified form of cells, where an ensemble of chemical reactions occurs in a varying volume, the volume of the protocell, that in turn increases because of the production of container molecules.

We presented several protocell models and we compare them with the analogous deterministic protocell models, namely solved using the ODE. In this preliminary study, we emphasized the role of the fluctuations and their dependence on the initial amount of molecules. The dynamics is richer than the deterministic one and thus it is worth studying, in particular we deserve to future investigations the case where several molecules interact in a linear way but including cross catalysis, i.e. the interaction matrix is not diagonal, or they interact in a non-linear way. Also the study of the emergence of time-periodic patterns due to the fluctuations, will be analyzed. An analytical treatment of the latter case could be possible using some recent technics developed by [11, 4], see also [2] where the space is also taken into account.

**Acknowledgment** The work of TC has been partially supported by the FNRS grant “Mission Scientifique 2010-2011”. TC would like also to thank the European Center for Living Technology in Venice (Italy) for the warm hospitality during a short visiting stay that originates this collaboration.

#### REFERENCES

- [1] Alberts B. et al., *Molecular Biology of the Cell*, (Garland, New York) 2002.
- [2] Anna P. et al., *Spatial model of autocatalytic reactions*, PRE, **81**, (2010), pp. 056110
- [3] T. Carletti et al, *Sufficient conditions for emergent synchronization in protocell models*, J. Theor. Biol., **254**, (2008), pp. 741.
- [4] T. Dauxois et al, *Enhanced stochastic oscillations in autocatalytic reactions*, PRE, **79**, (2009), pp. 036112

- [5] D.T. Gillespie, *Exact stochastic simulation of coupled chemical reactions*, J. of Phys Chem, **81**, 25, (1977), pp. 2340.
- [6] D.T. Gillespie, *Markov Processes: An Introduction for Physical Scientists*, Academic Press, Boston, (1992).
- [7] Kaneko, K., *Life : An introduction to complex system biology*, Springer, The Netherlands, (2006)
- [8] A.M. Kierzek, *STOKS: STOChastic Kinetic Simulations of biochemical systems with Gillespie algorithm*, 18, (3), Bioinformatics, (2002), pp. 470-481.
- [9] T. Lu, D. Volfson, L. Tsimring and J. Hasty, *Cellular growth and division in the Gillespie algorithm*, Syst. Biol., **1**, 1, (2004), pp. 121.
- [10] Luisi P. L., Ferri F. and Stano P, *Naturwissenschaften*, **93**, (2006), pp. 1.
- [11] A.J. McKane and T.J. Newman, *Predator-Prey Cycles from Resonant Amplification of Demographic Stochasticity*, PRL, **94**, (2005), pp. 218102.
- [12] Mansy, S.S. et al., *Template-directed synthesis of a genetic polymer in a model protocell*, Nature Letters, (2008), doi:10.1038/nature07018
- [13] Oberholzer, T. et al., *Enzymatic RNA replication in selfreproducing vesicles: an approach to a minimal cell*, Biochemical and Biophysical Research Communications, **207**, (1995), pp. 250257.
- [14] Rasmussen S., Chen L., Stadler B. and Stadler P., *Proto-Organism kinetics: evolutionary dynamics of lipid aggregates with genes and metabolism*, Origins Life Evol. Biosphere, **34**, (2004), pp. 171.
- [15] Rasmussen S. et al., *Transitions from non living to living matter*, Science, **303**, (2004), pp. 963-965.
- [16] R. Serra et al., , A. Life, **13**, (2007), pp. 123.
- [17] Szostak D., Bartel P. B. and Luisi P. L., *Synthesizing life*, Nature, **409**, (2001), pp. 387-390.

(Timoteo Carletti) NAXYS, NAMUR CENTER FOR COMPLEX SYSTEMS AND UNIVERSITY OF NAMUR, NAMUR, BELGIUM 5000

(Alessandro Filisetti) CIRI, ENERGY AND ENVIRONMENT INTERDIPARTIMENTAL CENTER FOR INDUSTRIAL RESEARCH, UNIVERSITY OF BOLOGNA AND EUROPEAN CENTRE FOR LIVING TECHNOLOGY, UNIVERSITY CÁ FOSCARI OF VENICE