Spatial extension of stochastic π calculus

Anton Stefanek Maria G. Vigliotti Jeremy T. Bradley

Department of Computing, Imperial College London

{as1005|mgv98|jb}@doc.ic.ac.uk

August 18, 2009

Abstract

We introduce a spatial extension of stochastic π-calculus that provides a formalism to model systems of discrete, connected locations. We define the extended stochastic semantics and also give deterministic semantics in terms of a system of ordinary differential equations. We describe two simple examples, one based on a standard epidemic model and one modelling resistance in plant tissues.

1 Introduction

Stochastic process algebras are becoming increasingly important in Systems Biology. Several frameworks have been developed (such as Bio-PEPA [6]) that allow convenient description of the models and different ways of formal analysis. Moreover, SPAs are naturally suited for extensions – for example in [5] Bio-PEPA is extended with spatial descriptions, a very important feature for modelling of various biological systems.

Stochastic π-calculus is a SPA that has been successfully applied to modelling in Systems Biology [4, 3, 16, 10]. This gives a motivation for its further extensions. We introduce certain spatial features to stochastic π-calculus. This has been done to some extent in the Bio-Ambient calculus [15]. However, our aim is to provide an extension that would allow a definition of an alternative, deterministic and continuous, semantics (it is not clear to us how this could be done for the Bio-Ambient calculus). In this work we will first remind the reader of the stochastic π-calculus and provide a definition of the deterministic semantics, based on the continuous π-calculus [11] and PEPA [9]. Then we introduce a spatial extension $\mathcal{L}\pi$ that allows modelling of discrete, connected locations and show how it keeps both the stochastic and deterministic semantics. We illustrate our ideas on two examples, one a standard epidemic model and another from plant physiology.

2 Stochastic π calculus

We will use a formalism (shortened to $\mathcal{S}\pi$) based on the standard stochastic π-calculus, as described in [14].

The basic primitives of $\mathcal{S}\pi$ are *processes* that communicate over *channels* or evolve independently. Communication happens via *actions*. On a channel a, a process can perform an *output action* $!a$, possibly involving transmission of a message, $!a\langle\widetilde{\psi}\rangle$ (where $\widetilde{\psi}$ is a vector of variables and channel names). On the other hand, a process can perform an *input action* $?a$, possibly receiving a message, $?a(\widetilde{x})$ (where $\widetilde{x}$ is a vector of variables which become bound by the values in the received message). Each channel has an associated constant rate, which corresponds to the rate of communication over that channel. To evolve independently, a process can perform a *silent action* $\tau@r$ at a specified rate r.

To summarize, the actions are

$$\alpha = !a \setminus ?a \setminus !a\langle\widetilde{\psi}\rangle \setminus ?a(\widetilde{x}) \setminus \tau@r.$$

The processes are built inductively from actions and the basic *zero* process $\mathbf{0}$ not capable of any action. A *continuation* is a process of the form $\alpha.P$ (where P is a process), capable of performing an action α and thereby evolving into the process P. A *summation* $\sum_{i\in I}\alpha_i.P_i$ is a

process with multiple such capabilities. A process $P|Q$ is a *parallel composition* of two processes P, Q that can communicate together. The restriction operator 'new' ensures that the channels in the set φ are private to the *restriction* process $(\text{new}\,\varphi)P$, unless sent to another process via channel communication. Finally, for a more convenient modelling and, more importantly, to allow recursion, a *process identifier instance* can be used in place of a process it defines, possibly with some parameters, $A\langle\widetilde{\psi}\rangle$.

To summarize, the set of processes of $\mathcal{S}\pi$ (denoted by $\mathcal{P}$) contains

$$P, Q = \mathbf{0} \setminus \sum_{i \in I} \alpha_i.P_i \setminus P|Q \setminus (\text{new}\,\varphi)P \setminus A\langle\widetilde{\psi}\rangle.$$

The binding between identifiers and the processes they define is specified in an *environment*, a collection of defining equations of the form

$$A(\widetilde{\psi}) \stackrel{\text{def}}{=} P.$$

An environment E together with an initial process S form a *system of* $\mathcal{S}\pi$ – a complete description of the underlying model.

For an example, consider a simple epidemic model. Due to a disease, a population can be divided into three types of individuals – *susceptible*, *infected* and *recovered*. A susceptible individual can catch the disease when meeting an infected one, who can also recover from the disease after a period of time. We can take the environment E_{SIR} consisting of the equations

$$\begin{aligned} S &\stackrel{\text{def}}{=} ?i.I, \\ I &\stackrel{\text{def}}{=} !i.I + \tau@r_{rec}.R, \\ R &\stackrel{\text{def}}{=} \mathbf{0} \end{aligned}$$

representing the three different types of individuals and specifying their behaviour. The communication on the channel i between the processes I and S corresponds to the transmission of the disease from an infected to a susceptible individual and the silent transition from a process I to the process R corresponds to the recovery of an infected individual. The population can be represented by an initial process of the form $s \times S|i \times I|r \times R$ (where $n \times P$ is a shorthand for a parallel composition of p copies of P).

2.1 Stochastic semantics

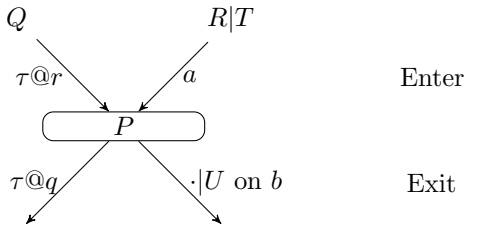
Traditionally, stochastic π-calculus has a discrete, stochastic semantics [14], given in the form of a continuous time Markov chain (CTMC). The states of the CTMC correspond to the processes and the transition rates are obtained from the rates associated to the channels and the rates of the silent actions according to rules defined on the structure of processes. Parallel compositions are capable of the same actions as their components. Those can can evolve independently or communicate on channels. If two processes capable of complementary actions (input and output) on the same channel are in a parallel composition, they evolve together with rate corresponding to the rate of the channel. If the process performing output action also sends a channel name coming from a restriction, the other process is put under the same restriction. Identifier instances are treated as the processes they define.

Unlike for some of the other SPAs, such as PEPA, the CTMCs arising from $\mathcal{S}\pi$ models can have an infinite state space (products of continuations can be parallel compositions) and so the only standard method of analysis is the stochastic simulation using a variant of the Gillespie algorithm [8]. The efficiency of this algorithm can be improved by aggregating identical processes within parallel compositions (using properties of the structural congruence for $\mathcal{S}\pi$) and hence making the complexity independent of individual process populations. The aggregation also shows that $\mathcal{S}\pi$ naturally obeys the law of mass action - it can be shown that the rate of communication between two processes depends on the product of their populations. See [17] for details.

2.2 Deterministic semantics

A recent trend in SPAs is to provide an alternative continuous semantics that serves as a deterministic approximation to the underlying CTMC, [9, 1, 11]. This has been done for an extension of the stochastic π-calculus, the *continuous π-calculus*, in [11]. We will employ a style more similar to the case of PEPA [9].

In the deterministic semantics, populations of processes are approximated by real valued functions over time that are mutually related via a system of ordinary differential equations (ODEs). For $\mathcal{S}\pi$, similarly to [11], we can define a notion of a *prime process* – a process that cannot be split into a parallel composition of non-zero processes – and show that each process (corresponding to a state in the CTMC) can be uniquely expressed as (is structurally congruent to) a parallel composition of prime processes. Then we can define for each prime process P a real valued function $[P]$ giving the population (its approximation) of P over time. To obtain the system of ODEs, we look at the possible ways the population of P can increase and decrease in a short period of time. It can increase as a result of a communication between two prime processes R, T or as a result of a silent transition of some prime process Q. On the other hand, it can decrease due to communication between P and another prime process Q or a silent transition of P.



We can keep track of these possibilities together with their multiplicities (for example $\tau@r.(P|P)$ increases the population of P by two) in form of Enter and Exit multisets and define the *system of ODEs of a system* (S, E) as consisting of the following, for each reachable prime process P:

$$\frac{\mathrm{d}[P]}{\mathrm{d}t} = \sum_{(r,Q)\in \mathrm{Enter}_{\tau,S,E}(P)} r \cdot [Q](t) + \sum_{(a,R,T)\in \mathrm{Enter}_{ch,S,E}(P)} r_a \cdot [R](t)[T](t)$$
$$\sum_{q\in \mathrm{Exit}_{\tau,S,E}(P)} q \cdot [P](t) - \sum_{(b,U)\in \mathrm{Exit}_{ch,S,E}(P)} r_b \cdot [P](t)[U](t).$$

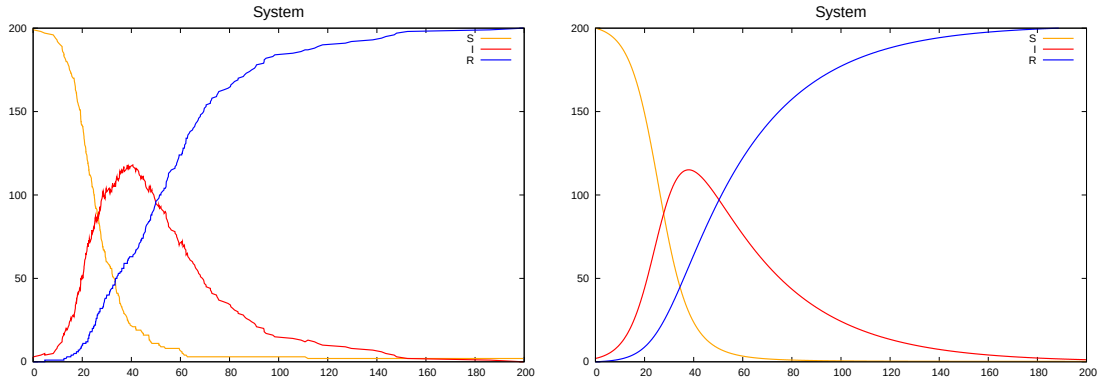
The initial values of the functions $[P]$ are given by the populations of P in the initial process S. This system of ODEs can then be numerically solved to give an approximation of process populations over time.

For the epidemic example, we get the following system of ODEs

$$\begin{aligned}
\mathrm{d}[S]/\mathrm{d}t &= -r_i \cdot [S](t)[I](t),\\
\mathrm{d}[I]/\mathrm{d}t &= r_i \cdot [S](t)[I](t) - r_{rec} \cdot [I](t),\\
\mathrm{d}[R]/\mathrm{d}t &= r_{rec} \cdot [I](t).
\end{aligned}$$

with initial values given by s, i and r in $s \times S|i \times I|r \times R$. See Figure 1 for a numerical solution of this system and a comparison with a sample stochastic simulation.

Unfortunately, the set of ODEs from the deterministic semantics of a $\mathcal{S}\pi$ system does not necessarily have to be finite. The interplay between the recursion and restriction can potentially result in an arbitrary number of newly created channels that are "connected" in arbitrary many different ways and thus form an arbitrary number of prime processes and the corresponding real valued functions. This is one of the distinctive features of stochastic π-calculus and has been used to model similar structures in nature, such as polymerization of actin filaments in [3]. However, most of the stochastic π-calculus models from the literature do not require such features (see [12] for a collection of some of them) – it makes sense to provide a characterization that will ensure a finite set of ODEs. A very crude solution is to restrict the stochastic π-calculus to the *Chemical*



(a) Stochastic simulation (b) ODE solution

Figure 1: A simple epidemic example. Comparison of the stochastic simulation of the CTMC from the discrete semantics and a numerical solution to the ODEs from the continuous semantics.

Ground Form (CGF) – a subset excluding the restriction operator and action parameters. We can show that this ensures finiteness of the resulting set of ODEs. This agrees with [2], where the author provides a translation of models in CGF to ODEs via chemical equations.

Finally, most of the models in [12] are in CGF or can be translated into "equivalent" ones in CGF (see [17] for some examples). This allows comparison of the two semantics and therefore further investigation into their relationship, with the possibility of formal limit results such as for PEPA [7].

3 Spatial extension – $\mathcal{L}\pi$

We can now define a simple spatial extension of $\mathcal{S}\pi$ which we will call $\mathcal{L}\pi$. We aim for a minimal extension that can provide a basis for further improvements. Considering applications in biology, we focus on *discrete* compartments between which the processes can move. The Bio-Ambient calculus [15] models this to a certain extent, but additionally allows dynamics of the compartmental structure. Albeit suitable for some applications (e.g. membrane modelling), we believe that it would not be straightforward to maintain the deterministic semantics of such extension – we therefore consider compartments with fixed structure. This can be justified by the fact that the current knowledge about biological systems is limited and the existing models mainly describe fixed compartments (as is argued in [5] where a similar extension is provided for Bio-PEPA).

3.1 Location graphs

We can argue that in case the compartments are non-overlapping (but possibly nested), their structure can be represented by a graph – the vertices represent the compartments and the edges represent how the processes (molecules, proteins, etc.) can move between them. This will form the basis of our extension. We define *location graphs* that give the structure of the *locations* (by which we mean generalized "compartments", not necessarily having physical boundaries). Inside each location, there are $\mathcal{S}\pi$ processes that can independently evolve, with rates affected by the locations volume. Moreover, these processes are allowed to move between the locations, with rate given by the location graph. We also assume that communication can happen only between processes inside the same location.

Formally, location graphs are of the form

$$[l_1 \colon P_1, \ldots, l_n \colon P_n]_{v,m}$$

where $L = \{l_1, \ldots, l_n\}$ is the set of location names, P_i are $\mathcal{S}\pi$ processes, $v \colon L \to \mathbb{R}$ is a *volume function* assigning a fixed volume to each location and $m \colon \mathcal{P} \times L \times L \to \mathbb{R}$ is a *movement function*

giving the rate of movement of processes between pairs of locations. A *system of* $\mathcal{L}\pi$ consists of an environment and a location graph.

Returning to the epidemic example, we can be interested in a system with a quarantine where the infected individuals get placed after a period of time and are kept until they recover. We can model this with a location graph with two locations, one corresponding to the original "world" (say a) and the other to the quarantine (say b), with processes coming from the original environment. There are two possible movements in this model – of I processes from a to b and of R processes from b to a. The rate of the first, say $r_{diagnose}$ can correspond to how fast an infected individual gets diagnosed and the rate of the second, say $r_{recover}$ to how long it takes to verify a recovery. Therefore we take $m(I, a, b) = r_{diagnose}$ and $m(R, b, a) = r_{recover}$ and $m(P, l_1, l_2) = 0$ for all the other processes P and locations l_1, l_2. We can ignore the location volume and set $v(a) = v(b) = 1$. As the initial location graph we can take (starting with an empty quarantine)

$$G = [a\colon s \times S | i \times I | r \times R, b\colon \mathbf{0}]_{v,m}$$

and get a system of $\mathcal{L}\pi$ (E_{SIR}, G).

3.2 Stochastic semantics

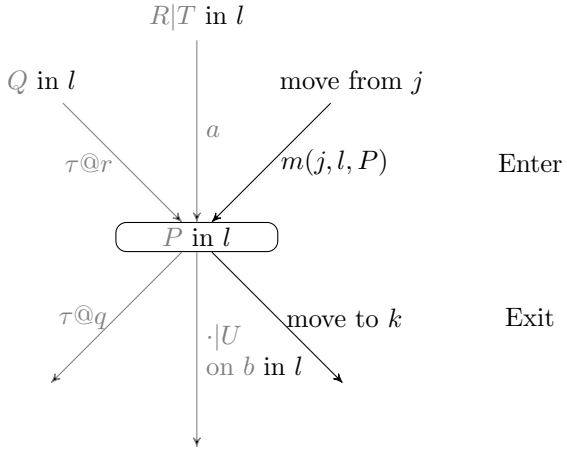
The stochastic semantics of $\mathcal{L}\pi$ is an extension of the stochastic semantics of $\mathcal{S}\pi$. Each state of the CTMC is a location graph. The transitions can be either internal to a single location or correspond to a movement between locations. In an internal transition, the processes inside a location l can communicate or evolve independently in the same way as in $\mathcal{S}\pi$, resulting in a transition to a location graph with only the contents of l changed accordingly. The rate is potentially affected by the volume of l – mimicking the chemistry, the rate of communication should be inversely proportional to the volume of l. Therefore the original $\mathcal{S}\pi$ rate gets divided by $v(l)$ in case it corresponds to a communication transition.

In the movement transition, a process P "moves" between two locations. If the process in l_1 contains P in a parallel composition and the movement function allows the movement of P between l_1 and another location l_2, i.e. $m(P, l_1, l_2) \neq 0$, then there is a possible transition to a graph with only locations l_1 and l_2 changed, where P is removed from the parallel composition in l_1 and added to the parallel composition in l_2. Ideally, we would only allow the movement of prime processes. However, the movement of restrictions would require more complicated semantics in order to respect the structural congruence. Therefore only summations and identifier instances defining summations can have a non-zero movement function. This can be further extended if suitable models that need movement of restrictions (complexes) are described.

The aggregation results and hence the efficient Gillespie algorithm from $\mathcal{S}\pi$ can be re-formulated for $\mathcal{L}\pi$ in an obvious way.

3.3 Deterministic semantics

The deterministic semantics of $\mathcal{L}\pi$ is an obvious extension. By the properties of $\mathcal{S}\pi$, the process in each location can be uniquely expressed as a parallel composition of prime processes. Therefore we can take a real valued function $[P]_l$ for each process P and location l. Population of a process P in each location l can change due to an internal transition – the corresponding terms in the resulting ODE are the same – and due to movement of P to and from l.



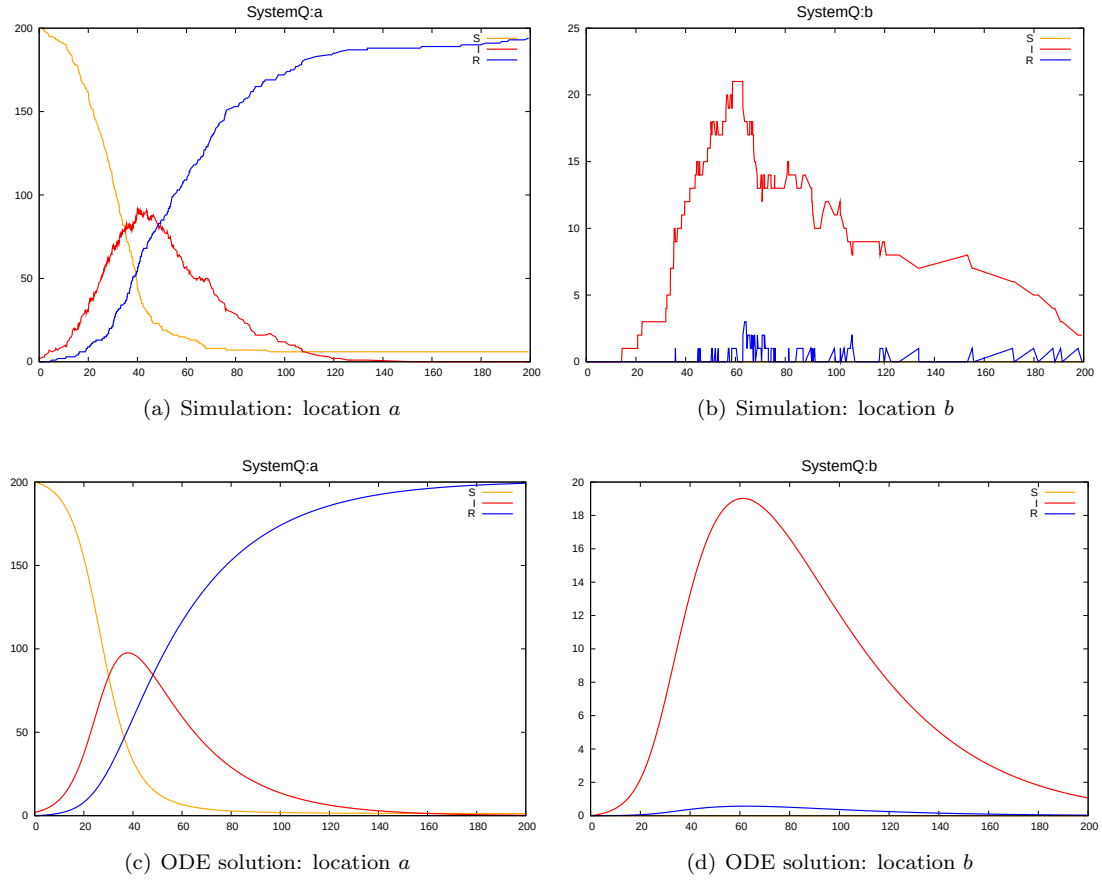
The Enter and Exit multisets can be constructed in a similar way to $\mathcal{S}\pi$, with the difference that we need to consider the movement in the set of all reachable prime processes. This then leads to an ODE for each real valued function $[P]_l$:

$$\begin{aligned}\frac{\mathrm{d}[P]_l}{\mathrm{d}t} = &\sum_{(r,Q)\in \mathrm{Enter}_{\tau,S,E}(P)} r[Q]_l(t) + \sum_{(a,R,T)\in \mathrm{Enter}_{ch,S,E}(P)} r_a[R]_l(t)[T]_l(t)/v(l) \\ &- \sum_{q\in \mathrm{Exit}_{\tau,S,E}(P)} q[P]_l(t) - \sum_{(a,U)\in \mathrm{Exit}_{ch,S,E}(P)} r_b[P]_l(t)[U]_l(t)/v(l) \\ &+ \sum_{m(j,l,P)\neq 0} m(j,l,P)[P]_j(t) - \sum_{m(l,k,P)\neq 0} m(l,k,P)[P]_l(t)\end{aligned}$$

For the epidemic example we get real valued functions $[S]_a, [I]_a, [R]_a, [I]_b, [R]_b$ (we ignore $[S]_b$ as, due to the definition of m, it clearly stays constant zero all the time). See Figure 2 for a numerical solution to the extended set of ODEs and a comparison with a sample simulation.

We will look at a simplified example from plant physiology [18]. Consider a hypothetical plant tissue consisting of cells arranged in a two dimensional grid. A cell can be attacked by a virus. A hypothesis is that in this case, it sends out a signal to the neighbouring cells, which in turn become more resistant to the virus and thus eventually prevent its spreading to the whole tissue. We will show how $\mathcal{L}\pi$ can be used to model this situation and so to carry out experiments in-silico to confirm the hypothesis.

Our location graph will represent the structure of the tissue – we take a grid with only adjacent nodes connected. Each location will correspond to a cell – initially, it will contain a *Cell* process. The *Virus* process will be able to attack a cell when in the same location. In that case, the cell releases warnings to the neighbouring cells – it will create several *Warning* processes, that are allowed to move to the neighbouring locations – and starts fighting the virus. The life of the cell will be represented by the process *Life* and the resistance against the virus by the *Resistance* processes. The resistance processes will be able to attack the virus (output action on the channel *defeat*), while the virus attacks the life of the cell (output action on the channel *fight*) – the likelihood of the cell surviving therefore depends on the number of resistance processes it releases. When the virus wins, the cell gets defeated and the virus multiplies, otherwise a resistance process destroys the virus and notifies the cell (output action on the channel *defeated*) which then switches back to the normal state (but with resistance to the virus). When a cell gets warned (communicates with a warning process), it switches to the resistant state (the process *RCell*), which is identical to the *Cell* with the difference that it releases more *Resistance* processes.



(a) Simulation: location a

(b) Simulation: location b

(c) ODE solution: location a

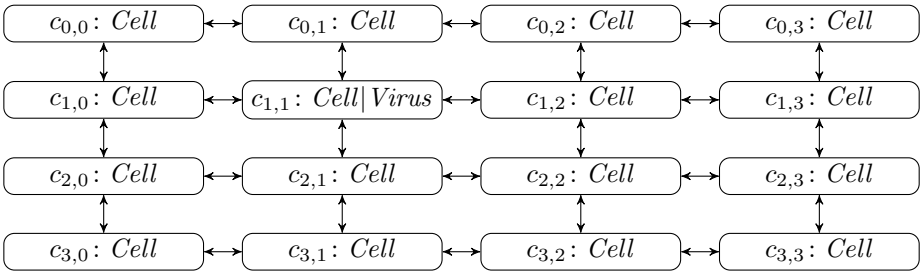
(d) ODE solution: location b

Figure 2:

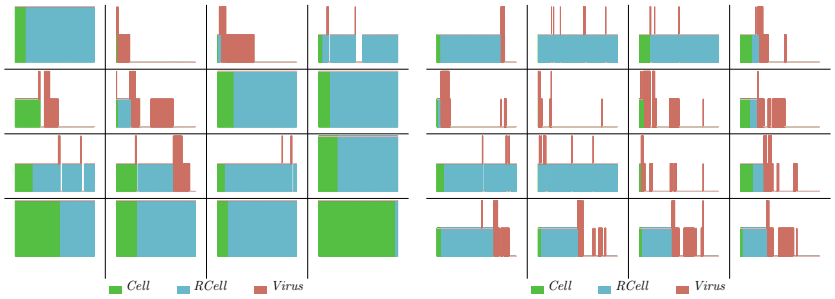
To model this behaviour we take an environment consisting of the defining equations

$$\begin{aligned} Cell &= ?attack.(Life|6 \times Reistance|4 \times Warning) + ?warn.RCell, \\ RCell &= ?attack.(Life|20 \times Reistance|4 \times Warning) + ?warn.RCell, \\ Resistance &= !defeat.!defeated + delay@expire, \\ Life &= ?fight + ?defeated.RCell, \\ Virus &= !attack.(!fight.(2 \times Virus) + ?defeat). \end{aligned}$$

To model the tissue structure, we take a location graph which can be drawn as the following:



The volume of all locations is a constant 1 and the *Warning* and *Virus* processes are allowed to move on the edges. See Figure 3 for a sample simulation of this $\mathcal{L}\pi$ system.



(a) Virus contained (b) Virus spreading

Figure 3: Sample simulation of the plant patogen model. Each cell in the grid represents time evolution of the corresponding compartment. The system starts with the virus in the location $c_{1,1}$. Figure (a) shows an example of a simulation where the virus is contained after attacking the neighbouring cells of $c_{1,1}$. Figure (b) shows a simulation where the virus spreads to the neighbouring cells and survives.

4 Conclusion and Further work

We introduced $\mathcal{L}\pi$, an extension of stochastic π-calculus that provides basic features for modelling systems of discrete, connected locations. To support this formalism, we developed a tool *JSPiM* (to be released in due time, see [17] for details of the implementation) that allows simulation and ODE generation and numerical solution of $\mathcal{S}\pi$ and $\mathcal{L}\pi$ models. Written in Java, it also serves as an alternative to the existing stochastic π-calculus tool *SPiM* [13]. We hope that with the help of this tool, more realistic examples from biology can be proposed, thus verifying the suitability of $\mathcal{L}\pi$ and giving direction for possible extensions. These could include the already mentioned movement of restrictions (usually representing chemical complexes). A syntactical extension of $\mathcal{L}\pi$ providing constructs for active movement could also prove useful – for example in the plant tissue model, the warning cells released would be each directed to a different neighbouring location instead of just allowed to randomly move.

A more ambitious task would be to relate $\mathcal{L}\pi$ to reaction-diffusion systems and work towards a formalism that would model the space continuously.

References

[1] L. Bortolussi and A. Policriti. Stochastic concurrent constraint programming and differential equations. *Electr. Notes Theor. Comput. Sci.*, 190(3):27–42, 2007.

[2] L. Cardelli. From processes to odes by chemistry. In *IFIP TCS*, 2008.

[3] L. Cardelli, E. Caron, P. Gardner, O. Kahramanoğulları, and A. Phillips. A process model of actin polymerisation. In N. Cannata, E. Merelli, and I. Ulidowski, editors, *Proceedings of the Workshop "From Biology To Concurrency and back (FBTC 2008)", July 2008*, volume 229 of *Electronic Notes in Theoretical Computer Science*, pages 127–144, Reykjavik, Iceland, 2008. Elsevier.

[4] L. Cardelli, P. Gardner, and O. Kahramanoğulları. A process model of rho gtp-binding proteins in the context of phagocytosis. *Electron. Notes Theor. Comput. Sci.*, 194(3):87–102, 2008.

[5] F. Ciocchetta and M. L. Guerriero. Modelling biological compartments in bio-pepa. *Electron. Notes Theor. Comput. Sci.*, 227:77–95, 2009.

[6] F. Ciocchetta and J. Hillston. Bio-pepa: a framework for the modelling and analysis of biological systems, 2008. Theoretical Computer Science.

[7] N. Geisweiller, J. Hillston, and M. Stenico. Relating continuous and discrete PEPA models of signalling pathways. *Theor. Comput. Sci.*, 404(1-2):97–111, 2008.

[8] D. Gillespie. Exact stochastic simulation of coupled chemical reactions. *The Journal of Physical Chemistry*, 81(25):2340–2361, 1977.

[9] J. Hillston. Fluid flow approximation of PEPA models. In *Proceedings of the Second International Conference on the Quantitative Evaluation of Systems*, pages 33–43, Torino, Italy, Sept. 2005. IEEE Computer Society Press.

[10] C. Kuttler and J. Niehren. Gene regulation in the pi calculus: simulating cooperativity at the lambda switch. *Transactions on Computational Systems Biology VII*, 4230:24–55, 2006.

[11] M. Kwiatkowski and I. Stark. The continuous π-calculus: A process algebra for biochemical modelling. In *Computational Methods in Systems Biology: Process of the Sixth International Conference CMSB 2008*, number 5307 in Lecture Notes in Computer Science, pages 103–122. Springer-Verlag, 2008.

[12] A. Phillips. Examples in spim. `http://research.microsoft.com/en-us/projects/spim/examples.pdf`.

[13] A. Phillips and L. Cardelli. Efficient, correct simulation of biological processes in stochastic π-calculus. *Proceedings of Computational Methods in Systems Biology*, pages 184–199, 2007.

[14] C. Priami. Stochastic π-calculus. *The Computer Journal*, 38(7), 1995.

[15] A. Regev, E. Panina, W. Silverman, L. Cardelli, and E. Shapiro. Bioambients: An abstraction for biological compartments. *Theoretical Computer Science, Special Issue on Computational Methods in Systems Biology*, 325(1):141–167, 2004.

[16] N. Segata and E. Blanzieri. Stochastic pi-calculus modelling of multisite phosphorylation based signalling: The pho pathway in sccharomyces cerevisiae. *Lecture Notes in Computer Science*, 2008.

[17] A. Stefanek. Continuous and spatial extension of stochastic π-calculus. Master's thesis, Imperial College of Science, Technology and Medicine, 2009.

[18] L. Taiz and E. Zeiger. *Plant Physiology*. Sinauer Associates, 2006.