Stochastic Bigraphs

Jean Krivine, Robin Milner, Angelo Troina

Stochastic Bigraphs[1]

Jean Krivine

Harvard Medical School, Systems Biology Department,
200 Longwood Avenue 02115 Boston MA, United States

Robin Milner

The Computer Laboratory, University of Cambridge,
JJ Thomson Avenue, Cambridge CB3 0FD, United Kingdom

Angelo Troina

Dipartimento di Informatica, Università di Torino,
Corso Svizzera 185, 10149 Torino, Italy

Abstract

In this paper we present a stochastic semantics for Bigraphical Reactive Systems. A reduction and a labelled stochastic semantics for bigraphs are defined. As a sanity check, we prove that the two semantics are consistent with each other. We illustrate the expressiveness of the framework with an example of membrane budding in a biological system.

1 Introduction

Bigraphical reactive systems (BRSs) [13,20] are conceived as a unifying framework for designing models of concurrent and mobile systems. These reactive systems are construed as a set of rewriting rules together with an initial bigraph on which the rules operate. Bigraphs are algebraic terms that may be represented as a particular kind of graphs allowing the representation of communication among agents as well as their spatial configuration (nodes may be nested within each other). This particular structure of bigraphs allows one to embed a wide class of calculi which focus on communications, such as CCS or the π-calculus [19], or on localities such as mobile ambients [5]. Although expressive power of BRSs is rather clear, they still lack means of "playing" with bigraphs dynamics in a quantitative way.

[1] This work has been mainly carried out during the three authors' stay at Laboratoire d'Informatique de l'École Polytechnique (LIX). The first two authors were supported by the State and Region Ile-de-France and the international *Blaise Pascal* research chair (assigned to Prof. Robin Milner for the academic year 2006/2007). Angelo Troina was supported by the INRIA/ARC project ProNoBiS.

This paper is electronically published in
Electronic Notes in Theoretical Computer Science
URL: `www.elsevier.nl/locate/entcs`

BRSs may be used to define independent dynamical systems that can be directly studied and simulated using an appropriate theory. The achievement of the paper is twofold. First, we obtain a uniform stochastic interpretation of behaviour in bigraphs that can be specialised to the entire range of BRSs. Second, using a simple but non trivial example of a biological system that requires dynamic compartmentalisation, we show that bigraphs are a good candidate for a formal algebraic and graphical representation of complex bio-molecular reactions.

Defining stochastic BRSs requires a correct enumeration of all the possible and *distinct* ways to apply each rule. Although enumerating such occurrences presents no difficulty in a more rigid calculus such as the π-calculus, a single redex may have several, though isomorphic, algebraic matches within a bigraph. In this paper, we shall develop mainly on that enumeration issue and assume some familiarity with traditional stochastic interpretation of algebraic processes [12,10,21] using Gillespie's method [11].

The present approach is an extension of the work on stochastic κ-calculus [8,7,6] in which great care was taken to define a stochastic semantics based on enumeration of rule instances. Here, we show that this approach can be generalized to a more powerful language where rules may have parameters, which entails that their occurrences in a given bigraph vary according to the choice of the outer *and* inner contexts. Indeed, in arbitrary bigraphs, choosing the outer context in which a rule left-hand side is to be placed does not guarantee the unicity of the inner context. We characterize a large fragment of bigraphs called *solid* where this property is always satisfied. We argue that this should turn out to be an important property from an efficient implementation perspective.

It is not the aim of the present paper to introduce a fully structured language for modelling biology nor to compare formally the present approach with existing algebraic formalisations of molecular compartments [4,23,10,18,2].

2 Bigraphical framework

In this section we define bigraphs informally, with enough detail to support both the biological example and the calculation of rates to be presented later.

Signature and interface

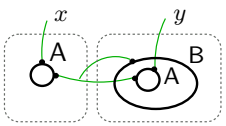
Bigraphs are the arrows in a category in which the objects are *(inter)faces*. We shall explain them by examples. First, each node of a bigraph is assigned a *control*, which determines how many *ports* it has. These controls are given by a *signature*, such as $\mathcal{K} = \{\mathsf{A}:2, \mathsf{B}:1, \mathsf{C}:0\}$.

Nodes may be nested arbitrarily, and their ports may be linked independently of the nesting.

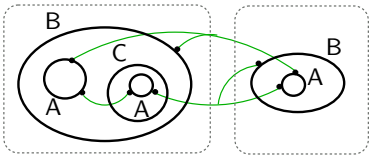
Each bigraph has an *inner* face and an *outer* face; these interfaces take the form $I = \langle m, X\rangle$ where m is a finite ordinal called a *width*, and X is a finite set of names such as $x, y, \ldots$. We write an interface $\langle m, \emptyset\rangle$ as m, and the trivial interface $\langle 0, \emptyset\rangle$ as ϵ. A bigraph $G : \langle m, X\rangle \to \langle n, Y\rangle$ has m *sites* and n *regions*; it also has *inner names* X and *outer names* Y. The *identity* arrow at interface I is written $\mathsf{id}_I : I \to I$. A bigraph with inner face ϵ is called *ground*.

Composition

Here are two ground bigraphs, each with two regions, conforming to the above signature:



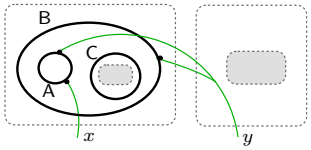
$F : \epsilon \to \langle 2, \{xy\} \rangle$



$G : \epsilon \to \langle 2, \emptyset \rangle$

F has three links; one *closed*, and the other two *open*—i.e. named by an outer name. Dashed rectangles represent regions. All the links of G are closed. Informally you may see that F occurs in G. This is confirmed by the following non-ground bigraph, which has two sites represented by shaded rectangles:

$H : \langle 2, \{xy\} \rangle \to \langle 2, \emptyset \rangle$



The outer face of F equals the inner face of H, so that we can form the categorical *composition* $H \circ F : \epsilon \to \langle 2, \emptyset \rangle$. In general, to form a composition $F' \circ F$ we place the regions of F in the corresponding sites of F', then delete those rectangles; we also join each open link of F, named x say, to the link in F' that contains the inner name x, then delete those names. These two operations are completely independent. If you form $H \circ F$ in this way, you find that $G = H \circ F$. Henceforth we shall omit the composition symbol, writing HF instead of $H \circ F$.

Product

Two interfaces $I = \langle m, X \rangle$ and $J = \langle n, Y \rangle$ are *disjoint* if $X \cap Y = \emptyset$; then they have a *tensor product* $I \otimes J \stackrel{\text{def}}{=} \langle m + n, X \uplus Y \rangle$. If the pairs (I_0, I_1) and (J_0, J_1) are disjoint, then the pair $G_i : I_i \to J_i$ has a *tensor product*

$$G_0 \otimes G_1 : I_0 \otimes I_1 \to J_0 \otimes J_1$$

in which G_0 and G_1 are placed side-by-side, in that order. Note that the tensor product is partial, unlike in a standard monoidal category, since it requires the name-sets X and Y to be disjoint. This is because names are drawn from a single global alphabet, thus allowing for a much easier derivation of other operations, including the parallel and prime products (which are central to process calculi), while making little difference to the mathematical treatment.

Other forms of product can be derived, in which the outer faces $J_0 = \langle m, X \rangle$ and $J_1 = \langle n, Y \rangle$ need not be disjoint; their *parallel product* is $J_0 \parallel J_1 \stackrel{\text{def}}{=} \langle m + n, X \cup Y \rangle$ and their *prime product* is $J_0 \mid J_1 \stackrel{\text{def}}{=} \langle 1, X \cup Y \rangle$. This yields the *parallel* and *prime products*

$$G_0 \parallel G_1 : I_0 \otimes I_1 \to J_0 \parallel J_1 \quad \text{and}$$

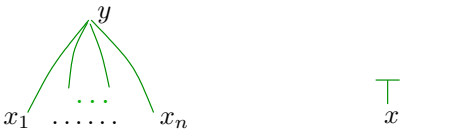
$$G_0 \mid G_1 : I_0 \otimes I_1 \to J_0 \mid J_1 \ .$$

In both cases the two bigraphs may be linked via shared outer names. In the second case their regions are also merged into one. Prime product is essentially the 'parallel composition' of the π-calculus. All these products are useful and have nice algebraic properties.

Linking

An interface $\langle 0, X\rangle$ with zero width will be written as X. Also, in interfaces, a set $X = \{x, y, \ldots\}$ will be written $xy\ldots$. A *linking* [2] $\lambda : X \to Y$ is a node-free bigraph whose faces have zero width. All linkings are generated by composition and tensor product from elementary *substitutions* $y_{/x_1 \cdots x_n}$ and *closures* $/x$:

elementary linkings:



$y_{/x_1 \cdots x_n} : x_1 \cdots x_n \to y$ $\qquad$ $/x : x \to \epsilon$

The substitution $y_{/y}$ is just the identity id_y.

To illustrate parallel product and closure, we may factorise our previously defined F as $F = (\mathsf{id}_I \otimes /z)(F_0 \parallel F_1)$ where $I = \langle 2, xy\rangle$:

$F : \epsilon \to \langle 2, xy\rangle$ $\qquad$ $F_0 : \epsilon \to \langle 1, xz\rangle$ $\qquad$ $F_1 : \epsilon \to \langle 1, zy\rangle$

This concludes our brief survey of the structure of bigraphs. For any signature $\mathcal{K}$, they form essentially a *strict symmetric monoidal category*, which we shall call here an *ssm category*.

We now move on to dynamics. We need a little terminology. Note that the sites, nodes and regions in a bigraph form a forest—i.e. an ordered set of trees whose roots are the regions. Thus 'root' and 'region' are synonyms.

Terminology

Every site or node in a bigraph has a *parent* which is either a node or a root. A root is *idle* if it has no children; an outer name is *idle* if its link has no ports or inner names. A site is *guarding* if its parent is a node. The *siblings* of a site or node are those with the same parent; the *siblings* of an inner name or port are those in the same link. A bigraph is *discrete* if every link is open and contains exactly one port or one inner name.

From now on we shall use lower case letters $a, \ldots, g, \ldots$ for ground bigraphs, and upper case letters $A, \ldots, G, \ldots$ for arbitrary bigraphs. But we reserve I, J, K for interfaces. We shall call ground bigraphs *agents*, since we ascribe dynamic behaviour to them.

[2] There is an analogous definition of a *placing*, a node-free bigraph whose faces have no names; we need not to detail it here.

Definition 2.1 [solid bigraph] A bigraph is *solid* if these conditions hold:

(a) no roots or outer names are idle;

(b) no two sites or inner names are siblings;

(c) every site is guarding;

(d) no outer name is linked to an inner name. □

Definition 2.2 [reaction rule] [3] A *(linear, parametric) reaction rule* is a pair

$$\mathsf{R} = (R : m \to J,\ R' : m \to J)$$

where R is the *parametric redex* and R' the *parametric reactum*, and R is solid. The rule generates all *ground reaction rules* (r, r'), where $r = (\mathsf{id}_Y \otimes R)d$ and $r' = (\mathsf{id}_Y \otimes R')d$ for some discrete ground *parameter* $d : \epsilon \to \langle m, Y \rangle$. The *reaction relation* $\longrightarrow\!\!\triangleright_{\mathsf{R}}$ over ground bigraphs is defined by

$$g \longrightarrow\!\!\triangleright_{\mathsf{R}} g' \text{ iff } g = Dr \text{ and } g' = Dr'$$

for some bigraph D and some ground reaction rule (r, r') generated by R. □

Four examples of reaction rules are shown in Figure 2. Parameters are shown as shaded rectangles. The fourth rule has $m = 1$ and $J = \langle 2, \{x\} \rangle$.

The constraint that d is discrete does not affect the reaction relation (since if d is non-discrete it can be replaced by a discrete parameter by adjusting the context D). This constraint simplifies the analysis in the computation of rates.

Now the central definition of the bigraphical framework follows.

Definition 2.3 [bigraphical reactive system] A *bigraphical reactive system (BRS)*, written $\textsc{Big}(\mathcal{K},\mathcal{R})$ consists of the bigraphical category $\textsc{Big}(\mathcal{K})$ determined by a signature $\mathcal{K}$, together with a set $\mathcal{R}$ of parametric reaction rules over $\mathcal{K}$. It has a reaction relation

$$\longrightarrow\!\!\triangleright_{\mathcal{R}} \ \stackrel{\text{def}}{=} \bigcup_{\mathsf{R} \in \mathcal{R}} \longrightarrow\!\!\triangleright_{\mathsf{R}}$$

which will be written as $\longrightarrow\!\!\triangleright$ when $\mathcal{R}$ is understood. □

These are the systems we wish to understand. We also need a refinement of them, for two reasons. First, much work in bigraphs has been done to derive *(labelled) transition systems* from the reaction rules, and this required a rigorous method to distinguish *occurrences* of redexes within agents. Second, it turns out that the very same method provides a means to count the number of ways (i.e. redex-occurrences) that can give rise to a reaction or to a transition, thus determining its rate. In what follows we shall derive rates for both reactions and transitions; the relation between the two will provide us with a good consistency check.

[3] This definition simplifies the standard one, but is sufficient for present purposes. There are two simplifications. First, we here restrict a rule to be *linear*, while in general a non-linear rule may discard or replicate some factors of the parameter d. Second, the standard definition requires the context D to be *active*, recognising that some controls prevent reaction within them.

We shall call a BRS *abstract* if it is defined as in Definition 2.3. Thus an abstract BRS is based upon an ssm category. In contrast, a *concrete* BRS is based upon a refinement of such a category, which we now define.

Definition 2.4 [s-category] An *s-category* is just as an ssm category, but each arrow A is equipped with a finite set $|A|$, its *support*. Furthermore, composition and tensor product are defined only for two arrows F, G with disjoint supports, and the support of the result is then $|F| \uplus |G|$. When defined, composition and product obey the usual properties of an ssm category.

In providing a way to identify occurrences, this refinement acts very like a labelling discipline in the λ-calculus.

Two arrows A and B are *support-equivalent*, written $A \simeq B$, if they differ only by a bijection between their supports which respects structure [4]. Denote the $\simeq$-equivalence class of A by $[\![A]\!]$. □

In bigraphs, the structure that must be respected by a support bijection consists of its placing and linking. If, e.g., $u, u' \mapsto v, v'$ under a support bijection from A to B, then node u is the parent of node u' in A iff v is the parent of v' in B.

Assume two disjoint infinite sets: $\mathcal{V}$ for *node-identifiers*, and $\mathcal{E}$ for *edge-identifiers*. (An *edge* is a closed link—one that has no outer name.)

We shall now define concrete bigraphs, as an s-category. Thereafter we shall often deal with concrete and abstract bigraphs simultaneously, so we need a convention to indicate which is which. It will turn out that concrete bigraphs get mentioned more often than abstract ones, so we shall adopt the following convention: Abstract bigraphs will always be denoted using a tilde, as in $\tilde{A}, \tilde{g}$, and concrete bigraphs without a tilde.

Definition 2.5 [concrete BRS] If $\tilde{G}$ is an abstract bigraph, then a *concrete* bigraph G, called a *concretion* of $\tilde{G}$, is obtained by assigning to each node a unique identifier $v \in \mathcal{V}$, and to each edge a unique identifier $e \in \mathcal{E}$. The *support* of G is given by $|G| \stackrel{\text{def}}{=} V \uplus E$, where $V \subset \mathcal{V}$ and $E \subset \mathcal{E}$ are the identifiers used.

Let $\textsc{Big}(\mathcal{K}, \mathcal{R})$ be an abstract BRS. Then the corresponding *concrete BRS*, $\grave{}\textsc{Big}(\mathcal{K}, \mathcal{R})$, is defined thus: Form the bigraphical s-category $\grave{}\textsc{Big}(\mathcal{K})$ over the given signature, then equip it with the reaction rules $\mathcal{R}$. The concrete reaction relation is defined as follows: $g \longrightarrow\!\!\!\!\triangleright_{\mathcal{R}} g'$ in $\grave{}\textsc{Big}(\mathcal{K}, \mathcal{R})$ whenever (g, g') is a concretion of $(\tilde{g}, \tilde{g}')$ and $\tilde{g} \longrightarrow\!\!\!\!\triangleright_{\mathcal{R}} \tilde{g}'$ in $\textsc{Big}(\mathcal{K}, \mathcal{R})$. □

From now on, when we do not need to specify $\mathcal{K}$ or $\mathcal{R}$ we shall write just Big or $\grave{}$Big for an abstract or concrete BRS.

Definition 2.6 If $\mathsf{R} = (\tilde{R}, \tilde{R}')$ is a rule in Big, then any concretion (R, R') of this pair is called a *concrete* parametric rule of $\grave{}$Big. Similarly, any concretion (r, r') of an abstract ground rule $(\tilde{r}, \tilde{r}')$ is called a *concrete* ground rule of $\grave{}$Big. □

[4] At the general level of s-categories, structure-respecting is defined by a few axioms which can be found in Definition 2.3 of [20]. They are omitted here.

Concrete bigraphs have structural properties that abstract bigraphs lack. For example, recall that an *epi(morphism)* A is an arrow such that $BA = B'A$ implies $B = B'$; similarly, A is a *mono(morphism)* if $AC = AC'$ implies $C = C'$. Epis and monos rarely exist in abstract bigraphs; in contrast we have:

Proposition 2.7 (concrete epis and monos) *A concrete bigraph is*

(a) epi if no roots or outer names are idle;

(b) mono if no two sites or inner names are siblings.

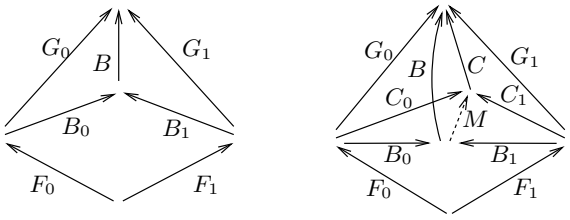
Now recall the notion 'solid', as defined in Definition 2.1; it applies equally to abstract and concrete bigraphs. But it implies the properties in Proposition 2.7, so by requiring redexes to be solid we know that in `BIG they are both epi and mono. In fact, they have an even stronger property:

Proposition 2.8 *If a concrete bigraph A is solid then $BAC = B'AC'$ implies $B = B'$ and $C = C'$.*

Proposition 2.8 states that once an occurrence of a concrete solid bigraph A is found within $D = BAC$, then the occurrence uniquely determines its 'environment' in D. Both these propositions will help in computing rates [5] .

We now come to another property found in concrete bigraphs: the notions of relative pushout and idem pushout.

Terminology (RPO and IPO):
A pair (F_0, F_1) of arrows is a *span* if they have the same inner face, and a *cospan* if they have the same outer face. A cospan (G_0, G_1) *bounds* a span (F_0, F_1) if $G_0F_0 = G_1F_1$. In this case, if (B_0, B_1) also bounds (F_0, F_1) and $BB_i = G_i$ $(i = 0, 1)$ then the triple (B_0, B_1, B) is a *bound for* (F_0, F_1) *relative to* (G_0, G_1). See the left-hand diagram below.



Definition 2.9 [relative pushout] Let (G_0, G_1) be a bound for (F_0, F_1). A *relative pushout (RPO) for* (F_0, F_1) *relative to* (G_0, G_1) is a relative bound (B_0, B_1, B) such that for any relative bound (C_0, C_1, C) there is a unique arrow M for which $MB_i = C_i$ $(i = 0, 1)$ and $CM = B$. (See the right-hand diagram.)

We say that an (s-)category *has RPOs* if, whenever (F_0, F_1) has a bound, it also has an RPO relative to that bound. □

[5] The result of Proposition 2.8 will mainly be applied with A a redex of a reaction rule. In such a case, condition (d) of Definition 2.1 is given for free since redexes have no inner names (see Definition 2.2).

In a sense, this makes (B_0, B_1) a *minimal* bound for (F_0, F_1); this is a weaker notion than a *minimum* bound, which is a pushout. We can express this minimality using the following special case of an RPO:

Definition 2.10 [idem pushout] (B_0, B_1) is an *idem pushout (IPO)* for (F_0, F_1) if (B_0, B_1, id) is an RPO for (F_0, F_1) relative to (B_0, B_1). □

Then the minimality of RPOs is expressed by the property (b) in the next proposition. RPOs and IPOs behave nicely, but for our present purpose we only need the following properties:

Proposition 2.11 (RPO properties) *It holds that:*

(a) Any RPO is unique up to isomorphism.

(b) The lower pair in any RPO is an IPO.

(c) `BIG *has RPOs, but* BIG *does not.*

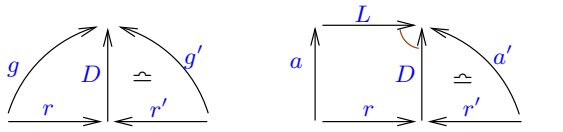
(d) Any IPO for a span with no idle roots or outer names is unique up to iso.

Property (a) of the proposition says that an RPO is unique except for any variant obtained by applying an isomorphism (iso) at its mediating interface. Similarly, when we compute rates by counting the number of *distinct* arrow structures (of certain kinds), variants by an iso at a mediating interface are not counted. An iso at an interface $I = \langle m, X \rangle$ consists of a permutation of its regions m and a bijective renaming of its names X.

The main use of RPOs has been to derive *(labelled) transitions* $a \xrightarrow{L} a'$, from a given set $\mathcal{R}$ of reaction rules. The *label* L is a bigraph which, when composed with a, completes an occurrence of a ground redex. The transition is *minimal* if L is no larger than needed to complete this occurrence, but to make this precise requires IPOs to exist. Since an abstract BRS BIG$(\mathcal{K},\mathcal{R})$ does not possess IPOs, we work in the concrete BRS `BIG$(\mathcal{K},\grave{}\mathcal{R})$, taking $\grave{}\mathcal{R}$ to consist of all concretions of the rules $\mathcal{R}$. Then:

Definition 2.12 [minimal transitions] Given a reaction rule R, a *minimal transition* $a \xrightarrow{L}_{\mathsf{R}} a'$ is a concrete triple[6] (a, L, a') such that, for some concrete ground rule (r, r') generated by R and some bigraph D, the cospan (L, D) is an IPO for the span (a, r) and $a' \eqsim Dr'$. □

Thus we may compare concrete reactions and transitions by the following diagram, in which an IPO is indicated by a small quarter-circle:



reaction $g \longrightarrow_{\mathsf{R}} g'$ minimal transition $a \xrightarrow{L}_{\mathsf{R}} a'$

[6] In previous work, transitions had a fourth component needed to ensure that the bisimilarity generated by minimal transitions is a congruence. This component makes no difference to the calculation of rates, so we ignore it here.

This completes the bigraphical background.

3 Modelling and simulating molecular systems

In this section we illustrate the expressivity of bigraphs in the context of a molecular system using compartments. The idea of using a formal language to model biology was pioneered by Regev and Shapiro [24] who modelled bio molecular interactions using the π-calculus [19]. Following the intuition, Danos and Laneve have proposed a flexible rule-based language called κ-calculus [9], that allows to model molecular complexation and post transcriptional modifications (activation or inhibition of binding domains of proteins) in a fairly direct way.

We show here how to represent a simple model of membrane budding in order to motivate the need of a stochastic semantics for bigraphs.

Membrane budding is a mechanism by which molecules living inside (or on the surface of) a cell may be transported to other cells *via* inter-cellular carriers [14]. We will first give a schematic description of the biological mechanisms at stake and then proceed with the bigraphical reactive system that models it qualitatively.

3.1 Membrane budding

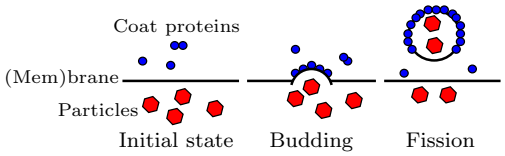


Fig. 1. Membrane budding

Our simple model of membrane budding proceeds in three steps. In the initial state, particles (the polygons in Fig.1) are freely floating inside the cell. Concurrently some coat proteins (the small circles in Fig.1) may assemble on the surface of the cell, modifying the lipid bi-layer of the membrane which in turns produces a bud [14]. If enough coat proteins are assembled, the bud may part completely from the cell. After fission, the bud carries away a certain amount of particles. Importantly, the size of the bud is constant in reality, thus implying that more coat proteins will lead to more buds and not to larger ones [15].

Node controls

Before presenting the reactive system itself, we proceed with the description of the different node controls that we will use in our example. Controls coat, gate and particle denote the atomic nodes of the system, meaning that no node may be nested inside them. The coat and gate nodes have arity 1 and particles have no ports. The non atomic controls (the compartments) are bud, with arity 1, and (mem)brane, with arity 0.

Rules

We may now use the nodes described above as lego parts to construct reaction rules. We will give each time the algebraic notation for the rules and depict the graphical

descriptions in Fig. 2. We will use $\Box$ to denote parameters and $\mathsf{K}_{x_1,\ldots,x_n}$ to denote a node of control K whose i^{th} port is connected to the outer name x_i. We shall also use the *nesting* operator, which is a simple but useful extension of the composition operation defined as:

$$G_0.G_1 \stackrel{def}{=} (id_X \parallel G_0) \circ G_1$$

where id_X is the identity on the outer names of G_1.

The first rule describes how a bud may form upon the collision of a coat protein and a membrane:

$$(/x\ \mathsf{coat}_x) \mid (\mathsf{brane}.\Box) \rightarrow /xy\ (\mathsf{coat}_x \mid \mathsf{bud}_x.\mathsf{gate}_y \mid \mathsf{brane}.(\mathsf{gate}_y \mid \Box)) \tag{1}$$

A bud is formed whenever a coat protein binds to the membrane. Note the creation of a channel between the bud and the membrane whose entries are denoted by gate nodes. The closure $/x$ on the left hand side ensures that the coat protein is not already bound to another bud in the system.

The second rule allows coat proteins to assemble around a bud.

$$\mathsf{bud}_x \mid (/y\ \mathsf{coat}_y) \rightarrow \mathsf{bud}_x \mid (x/y\ \mathsf{coat}_y) \tag{2}$$

A free coat protein may join a coat complex that is forming around a bud.

The next rule describes how particles may use a channel to pass from one compartment to another (here the bud and the membrane). Note that this rule is using explicitly two regions so we have to use here the parallel composition in addition to the prime product:

$$(\mathsf{gate}_x \mid \mathsf{particle}) \parallel \mathsf{gate}_x \rightarrow \mathsf{gate}_x \parallel (\mathsf{gate}_x \mid \mathsf{particle}) \tag{3}$$

A channel consists of two gate nodes on different regions connected to each other. A particle may pass from one region to the other whenever it collides with a gate. Note that rule (3) is symmetrical, therefore it may be applied either to pump particles into the bud or to push them back into the cell [7].

Our last rule produces the fission of the bud from the membrane as a consequence of the breaking of the channel that was linking them. As said, in this particular case of membrane budding, the size of a bud is given by the curvature of the coat proteins, hence we will suppose that a fission may only occur after a critical number (say n) of coat proteins have assembled.

$$/y\ \Big((\prod_n \mathsf{coat}_x \mid \mathsf{bud}_x.(\mathsf{gate}_y \mid \Box)) \parallel \mathsf{gate}_y\Big) \rightarrow (\prod_n \mathsf{coat}_x \mid \mathsf{bud}_x.(\Box)) \parallel 0 \tag{4}$$

When a sufficient amount of coat proteins have assembled, the channel attaching a bud to its membrane may break entailing the formation of a separated vesicle. The probability of fission increases with further coat binding after the n initial ones (see Example 1 of Section 4). In the appendix A we briefly show how one can study further this mechanism using the PRISM modele checker [17]. Note that the second region will always find its place inside a brane node.

[7] Although the rule is symmetrical, its stochastic version could be made asymmetric to express the differences of volumes between the cell and the bud.

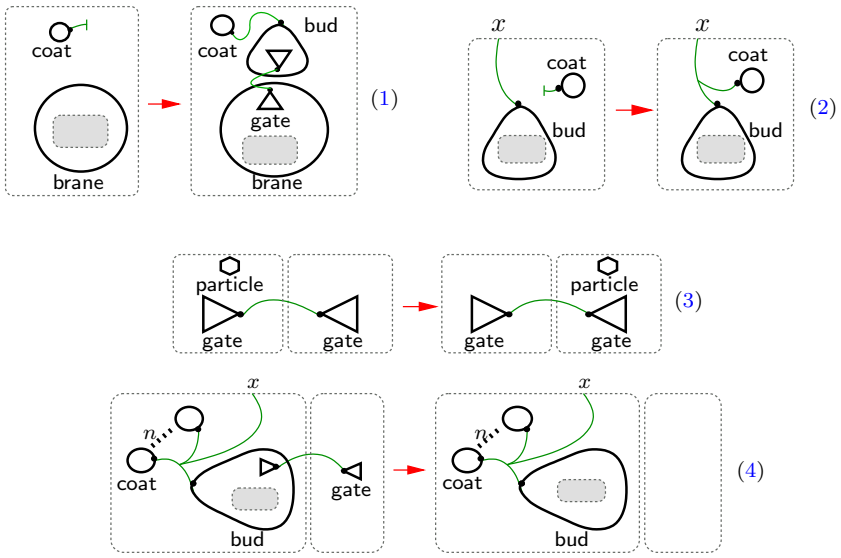


Fig. 2. Reactive System.

From a biological point of view, the reactive system we defined above only makes sense on particular initial graphs. To simplify, we can consider initial graphs where no buds have started to form and where coat nodes are siblings of brane nodes that contain particles. This kind of well formedness condition can be expressed, for example, using a sorting policy (see [13]).

The question is the following: given such an initial graph and the rules given in Fig. 2, can we make any quantitative description of the evolution of the system? In the next section we introduce the theoretical concepts which are necessary to answer this question.

4 Stochastic Bigraphs

In this section we associate a rate constant $\rho > 0$ with each abstract reaction rule. Based upon these *rated* rules, which take the form $\mathsf{R} = (\tilde{R}, \tilde{R}', \rho)$, we then define a rate for every reaction and every transition where redexes are assumed to be solid.

Since rates depend on counting occurrences, it is simplest to define rates first for *concrete* reactions and transitions. We begin with a general notion of occurrence. We talk of an abstract bigraph $\tilde{G}$ occurring within a concrete one H; different occurrences are determined by different concretions of $\tilde{G}$. With the help of this, we define occurrences of rules, and their results.

Definition 4.1 [inner occurrence] If $\tilde{G}$ is an abstract bigraph, an *inner occurrence* of $\tilde{G}$ in H is a pair (C, G) where G is a concretion of $\tilde{G}$ and $H = CG$. □

Definition 4.2 [occurrence] An *occurrence of* a rule R in a concrete agent g is an inner occurrence (C, r) in g, where (r, r') is a concrete ground rule generated by R. Two such occurrences are regarded as the same if they differ only by an iso on the interface between C and r; otherwise they are *distinct*. If also $g' \eqsim Cr'$ then we say the occurrence *results in* g'. □

We shall need the following, which can be proved from the fact that every parametric redex $\tilde{R}$ is solid and every parameter $\tilde{d}$ is discrete:

Proposition 4.3 (determining reaction) *Given a rule* $\mathsf{R} = (\tilde{R}, \tilde{R}', \rho)$*, let* (C, r) *be an occurrence of* R *in* g*, where* (r, r') *is a concrete ground rule generated by* R*. Then:*

(a) r *is an epimorphism.*

(b) C *is determined uniquely by* r*.*

(c) r' *is determined uniquely by* r *up to support equivalence.*

4.1 Stochastic Reactions

We now associate a rate with each reaction $g \longrightarrow\!\!\triangleright_{\mathsf{R}} g'$. Such a rate is the parameter of an exponential distribution that characterizes the stochastic behaviour of that reaction. The rate is obtained as the product of the rate constant ρ of the reaction rule and the number of distinct occurrences of the rule.

Definition 4.4 [rate of concrete reaction] Given g, g' concrete, and an abstract reaction rule $\mathsf{R} = (\tilde{R}, \tilde{R}', \rho)$, define $\mu_{\mathsf{R}}[g, g']$ to be the number of distinct occurrences (C, r) of R in g resulting in g'.

Each such occurrence is also called a *contribution* of R to the rate of $g \longrightarrow\!\!\triangleright_{\mathsf{R}} g'$.

The *reaction rate* of (g, g') for the rule R is defined formally by

$$\mathit{rate}_{\mathsf{R}}[g, g'] \stackrel{\text{def}}{=} \rho \cdot \mu_{\mathsf{R}}[g, g'] \ .$$

Finally, the reaction rate for a set $\mathcal{R}$ of rules is given by

$$\mathit{rate}_{\mathcal{R}}[g, g'] \stackrel{\text{def}}{=} \sum_{\mathsf{R} \in \mathcal{R}} \mathit{rate}_{\mathsf{R}}[g, g'] \ . \ \square$$

To compute the reaction rate of an abstract pair $(\tilde{g}, \tilde{g}')$, we can just compute the rate for an arbitrary concretion (g, g'), because the rate is independent of the chosen concretions, i.e.:

Proposition 4.5 *If* $f \backsimeq g$ *and* $f' \backsimeq g'$*, all concrete, then*

$$\mathit{rate}_{\mathcal{R}}[f, f'] = \mathit{rate}_{\mathcal{R}}[g, g'].$$

This justifies the following definition of the abstract reaction rate:

Definition 4.6 [rate of abstract reaction] Given a set $\mathcal{R}$ of rated reaction rules, the rate of an abstract reaction $\tilde{g} \longrightarrow\!\!\triangleright_{\mathcal{R}} \tilde{g}'$ is defined by

$$\mathit{rate}_{\mathcal{R}}[\tilde{g}, \tilde{g}'] \stackrel{\text{def}}{=} \mathit{rate}_{\mathcal{R}}[g, g']$$

where (g, g') is an arbitrary concretion of $(\tilde{g}, \tilde{g}')$. $\square$

Our rate calculation is consistent with the standard reaction relation, in the following sense:

Proposition 4.7 (consistency) *For any set $\mathcal{R}$ of stochastic reaction rules,*

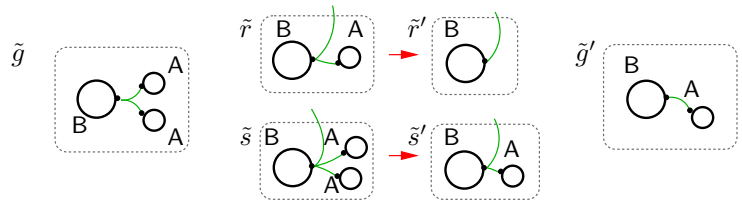
$$\mathit{rate}_{\mathcal{R}}[g, g'] > 0 \text{ iff } g \longrightarrow\!\!\!\!\triangleright_{\mathcal{R}} g' \ , \quad \text{and} \quad \mathit{rate}_{\mathcal{R}}[\tilde{g}, \tilde{g}'] > 0 \text{ iff } \tilde{g} \longrightarrow\!\!\!\!\triangleright_{\mathcal{R}} \tilde{g}' \ .$$

The following lemma states, intuitively, that to count the occurrences (C, r) of a rule we can just count the occurrences of its redex r (up to iso). It follows from Proposition 4.3.

Lemma 4.8 *Given a rule* R*, the count* $\mu_{\mathsf{R}}[g, g']$ *is the number of distinct concrete ground rules* (r, r') *generated by* R *such that, for some* C*,* $Cr = g$ *and* $Cr' \eqsim g'$*.*

Rates are especially simple when the parametric rules $\mathsf{R} = (\tilde{r}, \tilde{r}', \rho)$ are ground – i.e. the inner face of $\tilde{r}$ and $\tilde{r}'$ is ϵ. For then a generated concrete ground rule is simply a concretion (r, r') of $(\tilde{r}, \tilde{r}')$.

Example 4.9 [computing rates] Using ground rules, we now give a simple illustration of computing rates. The diagram shows an abstract agent $\tilde{g}$, two ground rules $\mathsf{R} = (\tilde{r}, \tilde{r}', \rho)$ and $\mathsf{S} = (\tilde{s}, \tilde{s}', \sigma)$, and the result $\tilde{g}'$ of applying each rule to $\tilde{g}$. (Each rule can occur in $\tilde{g}$ in two ways, but the abstract result $\tilde{g}'$ is the same in all cases.)



For each computation we fix a concretion g of $\tilde{g}$, assigning identifiers—say u_0 and u_1—to its two A-nodes, and identifier v to its B-node. Clearly this yields two occurrences of R in g, assigning u_0 and u_1 respectively to the A-node of $\tilde{r}$ and v to its B-node. Hence

$$\mathit{rate}_{\mathsf{R}}[g, g'] = 2\rho \ .$$

On the other hand there is only one occurrence of S in g; this is because any concretion of $\tilde{s}$ has a *support automorphism*, meaning that however we assign the identifiers to the two A-nodes of $\tilde{s}$ we get the same bigraph. Hence

$$\mathit{rate}_{\mathsf{S}}[g, g'] = \sigma \ .$$

The situation is more interesting if we keep R and S fixed but give $\tilde{g}$ any number $n \geq 2$ of A-nodes. The reader can easily check that the rates of $(\tilde{g}, \tilde{g}')$ for R and S will be respectively $\rho \cdot n$ and $\sigma \cdot n(n-1)/2$; the latter case involves the number of *unordered* pairs of A-nodes in $\tilde{g}$. Note that when there are more than two A-nodes, the open link in s can be linked to any other A-node in $\tilde{g}$. When there are just two A-nodes in $\tilde{g}$, as it is the case in the figure, the open link will be closed by the context. □

4.2 Stochastic Labelled Transitions

We now show that, just as minimal transitions $a \xrightarrow{L} a'$ can be derived from reaction rules, so can rates for these transitions be deduced from the rates for the rules. In computing these rates, we recall first that a certain transition may arise from more than one rule. So we begin by computing the rate of the concrete transition $a \xrightarrow{L}_{\mathsf{R}} a'$ based upon a given rule R. Such a computation is done by counting occurrences of R in La, as follows.

Definition 4.10 [rate of concrete transition] Let $\mathsf{R} = (\tilde{R}, \tilde{R}', \rho)$ be an abstract reaction rule, and let (a, L, a') be a triple of concrete bigraphs where the composition La exists. Define $\mu_{\mathsf{R}}[a, L, a']$ to be the number of distinct occurrences (D, r) of R in La such that (L, D) is an IPO for (a, r) and $a' \doteq Dr'$, where (r, r') is a concrete ground rule generated by R.

Then the *transition rate* of the triple for the rule R is defined as follows:

$$rate_{\mathsf{R}}[a, L, a'] \stackrel{\text{def}}{=} \rho \cdot \mu_{\mathsf{R}}[a, L, a'] \ .$$

The rate for a set $\mathcal{R}$ of rules is given by

$$rate_{\mathcal{R}}[a, L, a'] \stackrel{\text{def}}{=} \sum_{\mathsf{R} \in \mathcal{R}} rate_{\mathsf{R}}[a, L, a'] \ . \ \square$$

Now, just as for reactions, the rate of a concrete transition is independent of the chosen concretions:

Proposition 4.11 *In* `BIG, *let* $a \doteq b$, $L \doteq M$ *and* $a' \doteq b'$, *with* $|a| \cap |L| = |b| \cap |M| = \emptyset$, *where the compositions* La *and* Mb *exist. Then*

$$rate_{\mathcal{R}}[a, L, a'] = rate_{\mathcal{R}}[b, M, b'] \ .$$

This justifies the definition of the rate of an abstract transition:

Definition 4.12 [rate of abstract transition] Given a set $\mathcal{R}$ of rated reaction rules, the rate of an abstract transition $\tilde{a} \xrightarrow{\tilde{L}}_{\mathcal{R}} \tilde{a}'$ is given by

$$rate_{\mathcal{R}}[\tilde{a}, \tilde{L}, \tilde{a}'] \stackrel{\text{def}}{=} rate_{\mathcal{R}}[a, L, a']$$

where (a, L, a') is an arbitrary concretion of $(\tilde{a}, \tilde{L}, \tilde{a}')$ with $|a| \cap |L| = \emptyset$. $\square$

Again, our rate calculation is consistent with the standard transition relation.

Proposition 4.13 (consistency) *For any set* $\mathcal{R}$ *of stochastic reaction rules,*

$$rate_{\mathcal{R}}[\tilde{a}, \tilde{L}, \tilde{a}'] > 0 \ \textit{iff} \ \tilde{a} \xrightarrow{\tilde{L}}_{\mathcal{R}} \tilde{a}' \ .$$

4.3 A consistency check

In order to justify our definitions of rates for reactions and transitions we consider a consistency check, showing that they are related in a way we expect. The intuition

is as follows: Given an agent g and another agent a lying 'inside' g, the reactions possible for g should correspond exactly to the transitions that a can make using only labels that can be found within g. We now make this statement precise for a concrete BRS.

First we recall a useful property of IPOs containing epis.

Proposition 4.14 (IPO) *If F_0 and F_1 are epi then any IPO for (F_0, F_1) is unique up to iso.*

We also need to specify when we regard two transitions as distinct. Since an IPO can be varied by an iso ι, we find that whenever $a \xrightarrow{L}\!\triangleright\, a'$ then also $a \xrightarrow{\iota L}\!\triangleright\, \iota a'$, and we regard these as essentially the same transition. Also recall that the result a' of a transition can be varied by support equivalence. This leads to the following:

Definition 4.15 [distinct transitions] Two pairs (L_0, a'_0) and (L_1, a'_1), or two transitions $a \xrightarrow{L_0}\!\triangleright\, a'_0$ and $a \xrightarrow{L_1}\!\triangleright\, a'_1$, are *equivalent* if $L_1 = \iota L_0$ and $a'_1 \simeq \iota a'_0$, for some iso ι. Otherwise they are called *distinct*. □

Since this property is expressed in terms of occurrences, we prove it only for concrete reactions and transitions. A corresponding property for the abstract case follows as a corollary.

Theorem 4.16 *Let $\mathcal{R}$ be a set of abstract reaction rules. In* `BIG, *let g and a be fixed, with $Fa = g$. Assume that a has no idle names or roots. Then the rate of a concrete reaction from g equals the sum of the rates of all distinct transitions from a that are possible within g. Namely, for all g':*

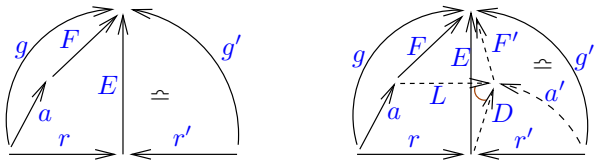
$$rate_{\mathcal{R}}[g, g'] = \sum\{rate_{\mathcal{R}}[a, L, a'] \mid \textit{distinct pairs } (L, a') \textit{ compatible with } g'\}$$

where (L, a') is compatible with g' *if $F = F'L$ and $F'a' \simeq g'$ for some F'.*

Proof. For simplicity we consider only ground rules with rate 1; the generalisation is straightforward. In particular, without loss of generality, we assume $\mathcal{R}$ to contain just one ground rule, $\mathsf{R} = (\tilde{r}, \tilde{r}', 1)$.

Let us refer to the left- and right-hand sides of the required equation as the LHS and RHS respectively.

Case $\leq$:
Since the rule is ground, a contribution to the LHS consists of a pair (E, r) with $g = Er$ and $g' \simeq Er'$, where (r, r') is a concretion of $(\tilde{r}, \tilde{r}')$. This is shown in the left-hand diagram.



reaction $g \longrightarrow\!\triangleright_{\mathsf{R}}\, g'$

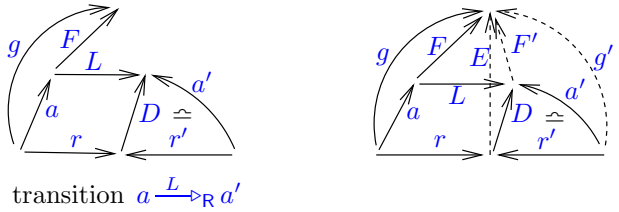
Now take an RPO (L, D, F') for (a, r) relative to (F, E) as shown on the right-hand diagram. Also define $a' = Dr'$.

Since $La = Dr$ is an IPO, (D, r) contributes to $rate_{\mathsf{R}}[a, L, a']$ in the RHS, the consistency condition is satisfied. Furthermore, by Proposition 4.3(b), each distinct contribution (E, r) to the LHS has a distinct component r, implying a distinct contribution (D, r) to $rate_{\mathsf{R}}[a, L, a']$. Thus the case $\leq$ is proven.

Case $\geq$:

We need to show that each distinct contribution to the RHS determines a distinct contribution to the LHS. Now a contribution to the RHS is determined in two stages; first a contribution to the transition rate for a with some pair (L, a'), and second the compatibility of this pair with g'.

(1) A contribution to $rate_{\mathsf{R}}[a, L, a']$: This consists of a pair (D, r) for some concretion (r, r') of the rule, such that (L, D) is an IPO for (a, r) and $a' \simeq Dr'$. This is shown in the left-hand diagram.



Since a and r are both epi, by Proposition 4.14 this IPO is unique up to iso. Hence, L, D and a' are determined by $(\tilde{r}, \tilde{r}')$ up to an iso at their outer face and up to support translation of a'. Thus every contribution (D, r) to the transition rate $rate_{\mathsf{R}}[a, L, a']$ is determined by a distinct concretion r of $\tilde{r}$. Furthermore this contribution is single; it cannot contribute to the transition rate for any other pair distinct from (L, a').

(2) Compatibility of (L, a') with g': Compatibility ensures that $F = F'L$ and $F'a' \simeq g'$, for some F'. Then, defining $E \stackrel{\text{def}}{=} F'D$, we easily check that $Er = g$ and $Er' \simeq g'$. Hence (E, r) is an occurrence of $\tilde{r}$ in g leading to the result g', so it contributes to the LHS.

Taking the two stages together, every contribution to the RHS determines a distinct contribution to the LHS. This concludes the proof of the case $\geq$. $\square$

Notice that the consistency check holds for any choice of a. As a varies, F also varies; by the condition $F = F'L$ this in turn affects the labels available in g for possible transitions of a.

Having established reasonable rates for transitions, it is natural to consider behavioural equivalences and pre-orders that take account of rates. We conjecture that bisimilarity, in which each transition of one agent is matched by one for the other agent with the same rate, is a congruence. We also expect a well-behaved pre-order in which the matching rate of the second agent must be at least as fast as that of the first. We leave these questions for future work.

5 Conclusions and Future Work

In this paper we have proposed a theoretical framework to deal with stochastic bigraphs. Here, rates are associated with rewrite rules from which it is possible to derive the rule activities on a given bigraph. We have pointed at biology to give an application that, on the one hand, shows the expressiveness of bigraphs, which are able to model dynamic compartmentalisation, on the other hand, convincingly underlines the necessity to extend the model with rates. We have the feeling that the method proposed here could be simplified to provide a more general collision-based stochastic semantics which could be applied in other graph transformations frameworks.

The stochastic engine behind PEPA [12] and the Stochastic π–calculus [22,21] is constructed on the intuition of cooperating agents under different bandwidth limits. If two agents are interacting, the time spent for a communication is given by the slowest of the agents involved. Our stochastic semantics is defined in terms of the collision–based paradigm introduced by Gillespie. A similar approach is taken in BioSPi ([22]) and the quantitative variants of the κ–calculus([7,16]), and the Calculus of Looping Sequences ([1]). Inspired by the law of mass action, here we need to count the number of the reactants present in a system in order to compute the exact rate of a reaction.

It remains to see how a stochastic machine could be implemented to run this kind of models. One might start by extending the highly scalable implementation of the stochastic κ-calculus [7] to bigraphs. The algebraic work on axiomatisation and implementation of bigraph matching [3] should also be considered.

The present paper is not limited to biology, it actually defines a model that can be used for a stochastic analysis of distributed systems in general. We plan, however, to study in more depth the possibility to model biological systems using a BRS. This certainly requires to narrow down the relevant subset of BRSs that is suitable for representing biological systems and to compare more precisely bigraphs expressiveness with existing algebraic formalisms for modelling membranes.

Acknowledgements

The authors wish to thank Sylvain Pradalier (LIX, École Polytechnique) for the valuable discussions during the initial phase of this work.

References

[1] R. Barbuti, A. Maggiolo-Schettini, P. Milazzo, P. Tiberi, and A. Troina. Stochastic calculus of looping sequences for the modelling and simulation of cellular pathways. *Transactions on Computational Systems Biology*, 2008. To appear.

[2] R. Barbuti, A. Maggiolo-Schettini, P. Milazzo, and A. Troina. A calculus of looping sequences for modelling microbiological systems. *Fundamenta Informaticæ*, 72(1–3):21–35, 2006.

[3] L. Birkedal, T. C. Damgaard, A. J. Glenstrup, and R. Milner. Matching of bigraphs. In *Proceedings of GT–VC'06*, ENTCS, 2006.

[4] L. Cardelli. Brane calculi - interactions of biological membranes. In *Computational Methods in Systems Biology*, pages 257–278. Springer, 2004.

[5] L. Cardelli and A. Gordon. Mobile ambients. *Theoretical Computer Science*, 24:177–213, 2000.

[6] V. Danos, J. Feret, W. Fontana, R. Harmer, and J. Krivine. Rule-based modelling, symmetries, refinements. In *Proceedings of FMSB 2008*, Lecture Notes in Bio Informatics, 2008. To appear.

[7] V. Danos, J. Feret, W. Fontana, and J. Krivine. Scalable modelling of biological pathways. In Z. Shao, editor, *Proceedings of APLAS 2007*, volume 4807, pages 139–157, 2007.

[8] V. Danos, J. Féret, W. Fontanta, R. Harmer, and J. Krivine. Rule based modeling of biological signaling. In *Proceedings of CONCUR 2007*, LNCS, 2007. Tutorial paper, to appear.

[9] V. Danos and C. Laneve. Formal molecular biology. *Theoretical Computer Science*, 325, 2004.

[10] P. Degano, D. Prandi, C. Priami, and P. Quaglia. Beta-binders for biological quantitative experiments. In *Proceedings of QAPL 2006*, volume 164 of *ENTCS*, pages 101–117, 2006.

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

[12] J. Hillston. *A Compositional Approach to Performance Modelling*. Cambridge University Press, 1996.

[13] O. H. Jensen and R. Milner. Bigraphs and mobile processes (revised). Technical Report UCAM-CL-TR-580, University of Cambridge Computer Laboratory, 2004. Available at `http://www.cl.cam.ac.uk/users/rm135`.

[14] T. Kirchausen. Three ways to make a vesicle. *Nature reviews. Molecular cell biology*, 1:187–198, 2000.

[15] M. M. Kozlov and L. Chernomordik. The protein coat in membrane fusion: Lessons from fission. *Traffic*, 3(4):256–267, 2002.

[16] J. Krivine. An exact and scalable stochastic simulation algorithm. In *Proceedings of the International Conference on Computational Methods in Science and Engineering 2007 (ICCMSE 2007)*, volume 963 of *American Institute of Physics conference proceedings.*, pages 638–641, 2007.

[17] M. Kwiatkowska, G. Norman, and D. Parker. Probabilistic symbolic model checking with prism: a hybrid approach. *Int. Journal on Software Tools for Technology Transfer*, 6(2):128–142, 2004.

[18] C. Laneve and F. Tarissan. A simple calculus for proteins and cells. In *Proceedings of the Workshop MeCBIC'06.*, 2006.

[19] R. Milner. *Communicating and mobile systems: the π-calculus*. Cambridge University Press, Cambridge, 1999.

[20] R. Milner. Pure bigraphs: structure and dynamics. *Information and computation*, 204:60–122, 2006.

[21] A. Phillips and L. Cardelli. Efficient, correct simulation of biological processes in the stochastic pi-calculus. In *CMSB 2007*, 2007. to appear.

[22] C. Priami, A. Regev, W. Silverman, and E. Shapiro. Application of stochastic name-passing calculus to representation and simulation of molecular processes. *Infomation Processing Letters*, 80(1):25–31, 2001.

[23] A. Regev, E. M. Panina, W. Silverman, L. Cardelli, and E. Shapiro. Bioambients: An abstraction for biological compartments. *Theoretical Computer Science*, 325:141–167, 2004.

[24] A. Regev and E. Shapiro. Cells as computation. *Nature*, 419, September 2002.

A Simulating Stochastic Bigraphs

In Section 3 we introduced a process of membrane budding as a motivating example for developing a stochastic semantics for bigraphs. In this appendix we would like to study the budding example when rates are associated to the rules. Since we still lack a simulator for stochastic bigraphs, it is hard to show more than the expressive power of bigraphs to advocate their use in the design of a language for biological cells. However, under some assumptions, the budding example could be made sufficiently simple to move the analysis on the Continuous Time Markov Chain that can be extracted from the transition system resulting from our quantitative semantics for bigraphs. Thus we can perform some simulations using existing tools like, e.g., the PRISM model checker [17].

A.1 Stochastic Bigraphs as CTMCs

In this section we show how the transition system deriving from our stochastic reactions can be transposed into a Continuous Time Markov Chain.

Definition A.1 A *Continuous Time Markov Chain* (CTMC) is a triple $\langle \mathsf{S}, \mathsf{T}, \pi \rangle$, where:

- S is the set of states,
- $\mathsf{T} : \mathsf{S} \times \mathsf{S} \mapsto \mathbb{R}^{\geq 0}$ is the *transition function*,
- $\pi : \mathsf{S} \mapsto [0, 1]$ is the *starting distribution*. □

A state $s \in \mathsf{S}$ denotes a possible configuration of the described system. The system is assumed to pass from a state s to a state s' by consuming an exponentially distributed quantity of time, in which the parameter of the exponential distribution is $\mathsf{T}(s, s')$. The summation $\sum_{s' \in \mathsf{S}} \mathsf{T}(s, s')$ is called the *exit rate* of state s. Finally, the system is assumed to start from a state $s \in \mathsf{S}$ with probability $\pi(s)$, and $\sum_{s \in \mathsf{S}} \pi(s) = 1$.

If the set of states of the CTMC is finite ($\mathsf{S} = \{s_1, \ldots, s_n\}$), then the transition function T can be represented as a square matrix of size n in which the element at position (i, j) is equal to $\mathsf{T}(s_i, s_j)$.

Many analysis techniques are available from mathematics and computer science for CTMCs. For example, if the set of states of the CTMC is finite, one can verify properties of the described system by using a probabilistic model checker such as PRISM.

The semantics of a concrete bigraph g can be naturally transformed into a CTMC by considering bigraphs as states, by setting $\pi(g) = 1$ and by defining $\mathsf{T}(g, g')$ as the rate of all reactions in $\mathcal{R}$ from g to g', namely:

$$\mathsf{T}(g, g') \;=\; \mathit{rate}_{\mathcal{R}}[g, g']$$

The set of states of the CTMC obtained by the semantics of a bigraph g can be restricted to the set of bigraphs which are reachable from g. Obviously, if such a set of bigraphs is finite, we obtain a finite state CTMC.

A.2 Simulating CTMCs

Given the CTMC obtained from the transition system resulting from our stochastic semantics, we can follow a standard simulation procedure. Roughly speaking, the algorithm starts from the initial state of the CTMC and performs a sequence of steps by moving from state to state. At each step a global clock variable (initially set to zero) is incremented by a random quantity which is exponentially distributed with the exit rate of the current state s as parameter, and the next state s' is randomly chosen with a probability proportional to $\mathsf{T}(s, s')$.

The same approach can be applied to Stochastic Bigraphs, where a state of the simulation is a pair (g, t) where g is the current concrete bigraph and $t \in \mathbb{R}^{\geq 0}$ is the global clock. Assuming a finite set of rewrite rules $\mathcal{R}$ and an initial graph g_0, the initial state of the simulation is the pair $(g_0, 0)$.

Given a simulation state (g, t), from the stochastic semantics, we have a finite set of transitions starting from g, namely the set of transitions $\{g \longrightarrow\!\!\triangleright\, g_j\}$, with $j \in [1, n]$, where n is the number of possible reactions starting from g. Now, a simulation step transforms the state (g, t) into $(g_j, t + \tau)$ where τ is exponentially distributed with parameter $p = \sum_{j=1}^{n} rate_{\mathcal{R}}[g, g_j]$ and the target bigraph g_j is chosen randomly with probability $rate_{\mathcal{R}}[g, g_j]/p$.

The *race condition* described above implements the fact that, on the lines of Gillespie's algorithm [11], when different reactions are competing with different rates, the ones which are not chosen should restart the competition at the following step.

A.3 Adding Rates to the Membrane Budding Example

We are now able to extend the example of Section 3 by adding rates to the rules. For simplicity we restrict ourselves to the case of a single membrane and a single bud formation (note, however, that the exposed example allows, in general, the creation of several buds in parallel).

Since we consider the creation of a single bud we may discard the first rule of the example (describing the interaction of a coat protein with the membrane starting the creation of a bud). Thus, we just consider the last three rules presented in Section 3, for *coating* (a coat protein joins the bud in formation), *diffuse_in_out_bud* (a particle may either enter or exit a bud in formation) and *fission* (the bud leaves the membrane).

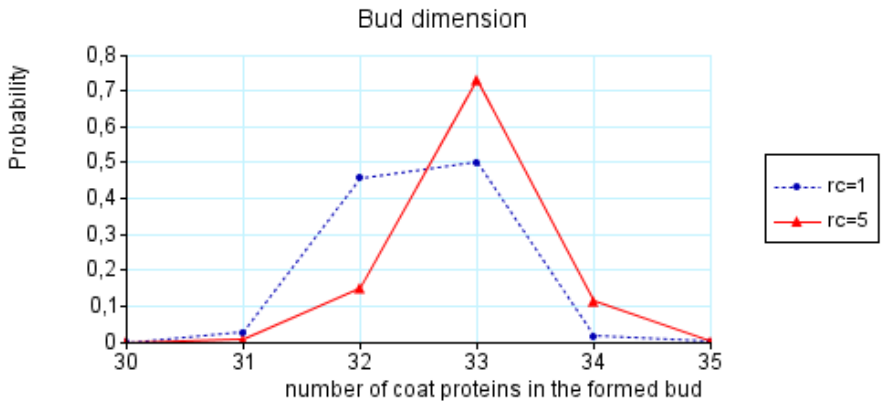
Note that removing the rule for the initial bud formation strongly simplifies the CTMC obtained from the stochastic transition system, which can now be extracted by hand. In fact, the restricted set of rules, together with the methodology for counting the occurrences in bigraphs and the procedure given in the previous subsection, can be used to get a quite simple CTMC which we are then able to simulate using PRISM. Actually, we obtained the following CTMC defined in the PRISM input language.

```
stochastic

const int NC=50;  //initial value of free coat proteins
const int NP=50;  //initial number of particles within the cell
const double rc;  //rate of coating
const double rd;  //rate of particles diffusion (in or out the bud)
const double rf;  //rate of fission

module budding

//State variables:

   c   : [0..NC] init 0; //number of coat proteins in the bud
   p   : [0..NP] init 0; //number of particles within the bud
   bud : [0..1]  init 0; //after fission bud=1 (termination state)
```

```
//Transitions ([label] "guard" -> "rate" : "state update"):

   [coating] (bud=0)&(NC-c>0) -> rc*(NC-c) : (c'=c+1);
     //NC-c=occurrences of free coat proteins

   [diffuse_in_bud] (bud=0)&(NP-p>0) -> rd*(NP-p): (p'=p+1);
     //NP-p=occurrences of particles within the membrane

   [diffuse_out_bud] (bud=0)&(p>0) -> rd*(p): (p'=p-1);
     //p=occurrences of particles in the forming bud

   [fission] (bud=0)&(c>30) -> rf*func(pow,c,c-31): (bud'=1);
     //since PRISM does not yet implement the binomial function
     //we used "func(pow,c,c-31)" to approximate it

   [self_loop] (bud=1) -> 1: (bud'=1);
     //self loop transition needed by PRISM
     //for computing the steady states space

endmodule
```

We considered a fixed initial configuration where 50 coat proteins are present in the environment surrounding the membrane and 50 particles are contained within the compartment, we set as $n = 31$ the minimum number of coat proteins needed for a bud to part from the membrane (fission rule). Note that the values used here have no relevant biological meaning, they are just used to analytically study the proposed application.

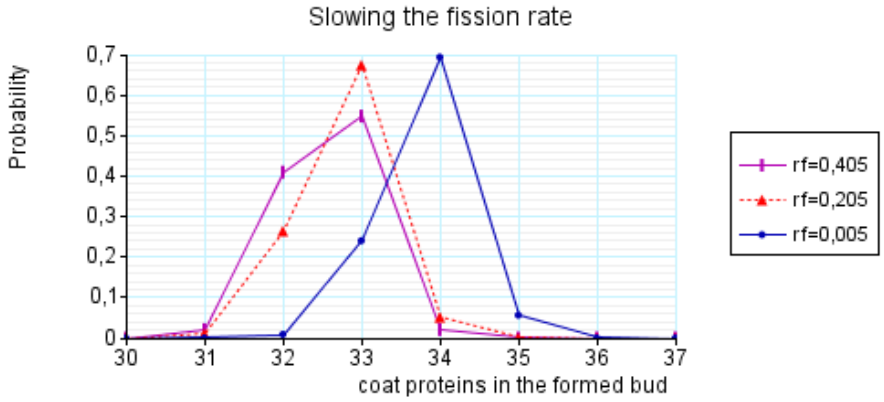
We start by computing the dimension of the bud after fission (i.e. the number of coat proteins which aggregated to form the bud). The next graph reports the probability of getting a bud whose dimension (i.e. the number of coat proteins assembled around the bud) is given by the value in the axis of abscissa. The study is performed for two different values of the rate at which coat proteins assemble, given by constant `rc`. The rate `rf` of the fission rule used for this experiment is 0.5, while `rd`, the rate of diffusion of particles in and out of the bud, was set to 1.



One sees that the number of coat proteins around a bud does not increase

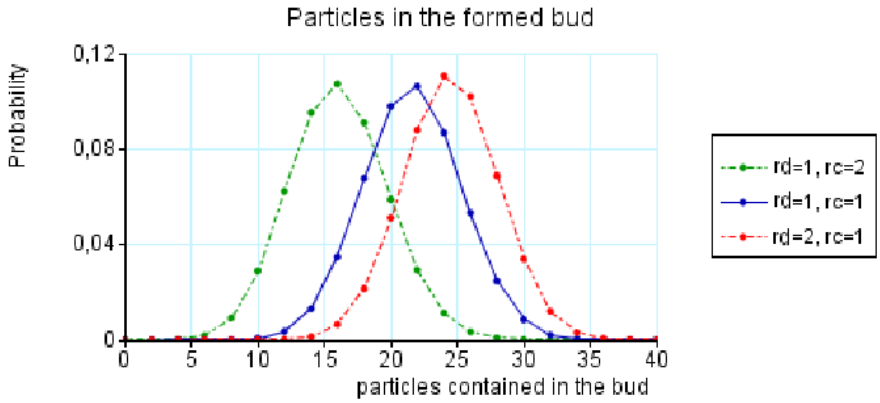
significantly on increasing the rate of the coating rule. Indeed, when the limit of 31 coat proteins is reached, the application rate of the fission rule grows exponentially.

In the next graph we see that a similar result arises when slowing the rate `rf` of the fission rule. Here we used `rc`=`rd`=1.



Hence, one may verify that the number of coat proteins that are carried out by a bud after fission is pretty much constant which coincide with the idea that the volume of a bud is also constant [15].

Another interesting analysis consists in the study of the number of particles present in the bud after fission. The following graph shows how such a value depends on the rates `rc` and, in particular, `rd` which models the movement of the particles between the two compartments. Again, we set `rf`=0.5.



We see that the final number of particles within the bud after fission increases when either increasing the rate `rd` (when particles move faster within the cell) or decreasing the rate `rc` (when coat proteins bind slower). Also note that the number of particles always follows a quite wide Gaussian distribution (and always with the same width).