

Hybrid Semantics of Stochastic Programs with Dynamic Reconfiguration

Luca Bortolussi

Dept. of Mathematics and Informatics,
University of Trieste, Italy.

luca@dimi.units.it

Alberto Policriti

Dept. of Mathematics and Informatics,
University of Udine, Italy.

Istituto di Genomica Applicata, Udine, Italy.

policriti@dimi.uniud.it

We begin by reviewing a technique to approximate the dynamics of stochastic programs—written in a stochastic process algebra—by a hybrid system, suitable to capture a mixed discrete/continuous evolution. In a nutshell, the discrete dynamics is kept stochastic while the continuous evolution is given in terms of ODEs, and the overall technique, therefore, naturally associates a Piecewise Deterministic Markov Process with a stochastic program.

The specific contribution in this work consists in an increase of the flexibility of the translation scheme, obtained by allowing a *dynamic* reconfiguration of the degree of discreteness/continuity of the semantics.

We also discuss the relationships of this approach with other hybrid simulation strategies for biochemical systems.

1 Introduction

Models in systems biology tend to cluster around two families of mathematical tools: differential equations and stochastic processes. Even though, physically speaking, stochastic models have firmer grounds [14, 15], their computational analysis is much more costly than that of their differential counterpart. In any case, ODE-based descriptions of biological systems are often valuable and provide deep insights. Indeed, it is known that, limiting to mass action models, ODE's are an approximation of the average of stochastic models, and the differences between the two vanish in the thermodynamic limit [13] (i.e. when populations and system's size go to infinity). Recently, there have been many attempts to mix these two techniques, at least as far as simulation of biological systems is concerned, resulting in several hybrid simulation algorithms [32, 25]. Hybrid dynamical systems have also been a hot topic in the last two decades, with much research work spanning across the boundary between computer science and engineering control. The best known model among hybrid dynamical systems are *hybrid automata* [21]. Stochastic extensions of such concept are also receiving recently much attention [10], although stochastic hybrid systems have a somewhat longer tradition [12]. In both cases, most of the interest is in the development of automated reasoning tools rather than in simulation.

It is widely recognized that Computational Systems Biology can highly benefit from modeling approaches embodying *some* stochastic ingredient. A very popular line along which such incorporation is realized, is based on the use of stochastic process algebras [27, 11], which are proposed as front-end languages to (automatically) generate mathematical models, usually Continuous Time Markov Chains (CTMC), see [32]. Recently, such process algebra based languages have also been endowed with semantics based on ODE [22], which increase the flexibility of such tools.

Many proposals of hybrid simulation algorithms for systems of biochemical reactions have been put forward [19, 24, 20, 28, 23, 16, 2]. Their salient feature is a description of one part of the system as

continuous, while keeping the other discrete and stochastic. The basic idea is to find the best trade off between accuracy and computational efficiency (stochastic simulations are much more expensive than ODE simulation).

In this paper we continue a programme which aims to increase even more the flexibility of stochastic process algebras by providing them with a very general semantics based on (stochastic) hybrid systems, encompassing CTMC and ODE as special cases. Such an approach is motivated not only by the gain in flexibility, but also by the goal of exploiting, in a systematic manner, automated reasoning tools to provide as much information as possible from a given model. Our stochastic process algebra of choice is stochastic Concurrent Constraint Programming (**sCCP**) [6], an extension of CCP [29] in the stochastic setting. In addition to the standard CTMC-based semantics, we have also provided sCCP with an ODE-based semantics [4] and with an hybrid automata based semantics. Moreover, hybrid semantics has been proposed both with a fixed or user-defined amount of continuously approximated components (see [8, 9]).

In this paper we extend our work by introducing a semantics based on *Stochastic Hybrid Automata*, thereby guaranteeing the possibility of parameterizing the degree of continuity introduced in the model. The approach allows also a *dynamic* reconfiguration of such degree, in accordance to properties of the current state of the system. This allows the description in a formal setting of different hybrid simulation strategies, opening the way for their use in the context of process algebra modelling.

We will start our presentation by introducing, in Section 2, a high level description of the target stochastic hybrid systems, suitable to be easily mapped to the well-established formalism of Piecewise Deterministic Markov Processes (see supplementary material [1]). The formalism introduced in Section 2, called Transition-Driven Stochastic Hybrid Automata (TDSHA), will act as the intermediate layer in the definition of the stochastic hybrid semantics of **sCCP**. Section 3 briefly introduces the **sCCP** language, while Section 4 presents the mapping from **sCCP** to TDSHA. Collections of TDSHAs can be organized in a lattice, whose definition and basic properties are presented in Section 4.1. Finally, in Section 5 we introduce the dynamic reconfiguration mechanism, briefly discussing also how to render, in such reconfigurations, partition strategies developed for hybrid simulation algorithms.

2 Transition-driven Stochastic Hybrid Automata

We define here a stochastic variant of *Transition-Driven Hybrid Automata*, introduced in [9] as an intermediate layer to map **sCCP** into hybrid automata. The emphasis is on *transitions* which, as always in hybrid automata, can be either discrete (corresponding to jumps) or continuous (representing flows acting on system's variables). The stochastic variant defined below contains two kind of discrete transitions: instantaneous—as in [9]—and stochastic, which happen with an hazard given by a rate function.

Definition 2.1. A Transition-Driven Stochastic Hybrid Automaton (TDSHA) is a tuple $\mathcal{T} = (Q, \mathbf{X}, \mathfrak{T}\mathcal{C}, \mathfrak{T}\mathcal{D}, \mathfrak{T}\mathcal{S}, \text{init})$, where:

- Q is a finite set of *control modes*.
- $\mathbf{X} = \{X_1, \dots, X_n\}$ is a set of real valued *system's variables*¹.
- $\mathfrak{T}\mathcal{C}$ is the set of *continuous transitions or flows*, whose elements τ are triples $(q, \text{stoich}, \text{rate})$, where: $q \in Q$ is a mode, *stoich* is a vector of size $|\mathbf{X}|$, and *rate* : $\mathbb{R}^n \rightarrow \mathbb{R}$ is a (sufficiently smooth) function. The elements of a triple τ are indicated by **cmode** $[\tau]$, **stoich** $[\tau]$, and **rate** $[\tau]$, respectively.

¹Notation: the time derivative of X_j is denoted by \dot{X}_j , while the value of X_j after a change of mode is indicated by X_j'

- $\mathfrak{T}\mathfrak{D}$ is the set of *instantaneous transitions*, whose elements δ are tuples of the form $(q_1, q_2, \text{priority}, \text{guard}, \text{reset})$, where: q_1 is the *exit-mode*, q_2 is the *enter-mode*, $\text{priority} : \mathbb{R}^n \rightarrow \mathbb{R}^+$ is a weight function used to resolve non-determinism between two or more active transitions. Moreover, guard is a quantifier-free first-order formula with free variables in \mathbf{X} , representing the *closed set* $G_\delta = \{\mathbf{x} \in \mathbb{R}^n \mid \text{guard}[\mathbf{x}]\}$ in which the transition is active, and reset is a deterministic update of the form $\mathbf{X}' = f(\mathbf{X})$.² The elements of a tuple δ are indicated by $\mathbf{e}_1[\delta]$, $\mathbf{e}_2[\delta]$, $\text{priority}[\delta]$, $\text{guard}[\delta]$, and $\text{reset}[\delta]$, respectively.
- $\mathfrak{T}\mathfrak{S}$ is the set of *stochastic transitions*, whose elements η are tuples of the form $\eta = (q_1, q_2, \text{guard}, \text{reset}, \text{rate})$, where q_1 , q_2 , guard , and reset are as for transitions in $\mathfrak{T}\mathfrak{D}$, while $\text{rate} : \mathbb{R}^n \rightarrow \mathbb{R}^+$ is the rate function giving the hazard of taking transition η . Such function is referred to by $\text{rate}[\eta]$.
- *init* is a point giving the initial state of the system.

A TDSHA has three types of transitions. Continuous transitions represent flows and, for each $\tau \in \mathfrak{T}\mathfrak{C}$, $\text{stoich}[\tau]$ and $\text{rate}[\tau]$ give the *magnitude* and the *form* of the flow of τ on each variable $X \in \mathbf{X}$, respectively (see below). Instantaneous transitions represent actions happening immediately when their guard becomes true. Finally, stochastic transitions happen at a specific rate. Both instantaneous and stochastic transitions can change system variables according to a specific reset function, depending on the variables' value at the point in time at which the jump occurs.

Remark 2.1. Both priority and rates introduced in Definition 2.1 make TDSHA stochastic. Priorities define, at each point, a discrete distribution of a random variable choosing among enabled instantaneous transitions. Rates, on the other hand, define a random race in continuous time, giving the delay for the next spontaneous jump.

Product of TDSHA. Given two TDSHA $\mathcal{T}_1 = (Q_1, \mathbf{X}_1, \mathfrak{T}\mathfrak{C}_1, \mathfrak{T}\mathfrak{D}_1, \mathfrak{T}\mathfrak{S}_1, \text{init}_1)$ and $\mathcal{T}_2 = (Q_2, \mathbf{X}_2, \mathfrak{T}\mathfrak{C}_2, \mathfrak{T}\mathfrak{D}_2, \mathfrak{T}\mathfrak{S}_2, \text{init}_2)$, the product TDSHA $\mathcal{T} = \mathcal{T}_1 \otimes \mathcal{T}_2$ can be defined in a simple way, along the path outlined in [9]. Essentially, the discrete states' space of the product automaton is $Q_1 \times Q_2$, while transitions from state (q_1, q_2) are all those issuing from q_1 or q_2 . Instantaneous or stochastic transitions of \mathcal{T}_1 going from state q_1 to state q'_1 , will go from a state (q_1, q_2) to (q'_1, q_2) for each $q_2 \in Q_2$. Symmetrically for transitions of \mathcal{T}_2 .

Dynamics of TDSHA. In order to formally define the dynamical evolution of TDSHA, we can map them into a well-studied model of Stochastic Hybrid Automata, namely Piecewise Deterministic Markov Processes [12]. In this sense, TDSHA are related to communicating PDMP [31], as they can also be seen as a compositional formalism to model PDMP. Due to space constraints, we just sketch here an informal description of PDMP. The interested reader can find a more formal treatment of PDMP and of their relation with TDSHA in the supplementary material [1].

Basically, PDMP are stochastic processes whose *state space* is given by a finite collection of *discrete modes* and by a set of *real-valued variables*. Within each mode, the continuous variables evolve following the solution of a set of *mode-specific ODE's*. While in a mode, variables must stay within the *allowed region*. If they touch the boundary of the allowed region, a *forced discrete transition* is taken, and the system may change mode and/or reset the value of the variables. Moreover, the system is subject to the happening *discrete stochastic events*, governed by an hazard rate that is function of the discrete

²Even though there is no real additional difficulty in considering stochastic resets—i.e. in assuming *reset* to be a transition measure—we decided to avoid such move for the sake of simplicity.

mode and of continuous variables. Also stochastic transitions trigger a reset of the state of the system. The main points of the mapping from TDSHA to PDMP are the following.

- Within each discrete mode $q \in \mathcal{Q}$, the system follows the solution of a system of ODE, constructed combining the effects of the continuous transitions τ acting on mode q . Essentially, the ODE for variable X_i is obtained by adding up the rate of all such τ times the i -th component of the vector **stoich** $[\tau]$:

$$\dot{\mathbf{X}} = \sum_{\tau \mid \mathbf{cmode}[\tau]=q} \mathbf{stoich}[\tau] \mathbf{rate}[\tau] \quad \text{in mode } q \in \mathcal{Q}.$$

- Two kinds of discrete jumps are possible: stochastic transitions are fired according to their rate, while instantaneous transitions are fired as soon as their guard becomes true. In both cases, the state of the system is reset according to the policy specified by **reset**. Choice among several active stochastic or instantaneous transitions is resolved probabilistically according to their rate or priority, see Remark 2.1.
- A trace of the system is therefore a sequence of (random) jumps interleaved by periods of continuous evolution.

3 Stochastic Concurrent Constraint Programming

In this section we briefly present (a simplified version of) stochastic Concurrent Constraint Programming (**sCCP** [3], a stochastic extension of CCP [30]), as it seems to be sufficiently expressive, compact, and especially easy to manipulate for our purposes³. In the following we just sketch the basic notions and the concepts needed in the rest of the paper. More details on the language can be found in [3, 6].

Definition 3.1. A **sCCP** program is a tuple $\mathcal{A} = (A, \mathcal{D}, \mathbf{X}, \mathit{init}(\mathbf{X}))$, where

1. The *initial network of agents* A and the *set of definitions* \mathcal{D} are given by the following grammar:

$$\begin{aligned} \mathcal{D} &= \emptyset \mid \mathcal{D} \cup \mathcal{D} \mid \{C \stackrel{\text{def}}{=} M\} \\ \pi &= [g(\mathbf{X}) \rightarrow u(\mathbf{X}, \mathbf{X}')]_{\lambda(\mathbf{X})} \quad M = \pi.C \mid M + M \quad A = M \mid A \parallel A \end{aligned}$$

2. \mathbf{X} is the set of variables of the store (with global scope);
3. $\mathit{init}(\mathbf{X})$ is a predicate on \mathbf{X} of the form $\mathbf{X} = \mathbf{x}_0$, assigning an *initial value* to store variables.

In the previous definition, basic actions are *guarded updates* of (some of the) variables: $g(\mathbf{X})$ is a quantifier-free first order formula whose atoms are inequality predicates on variables \mathbf{X} and $u(\mathbf{X}, \mathbf{X}')$ is a predicate on \mathbf{X}, \mathbf{X}' of the form $\mathbf{X}' = f(\mathbf{X})$ (\mathbf{X}' denotes variables of \mathbf{X} after the update), for some function $f : \mathbb{R}^n \rightarrow \mathbb{R}^n$. Each such action has a *stochastic duration*, specified by associating an exponentially distributed random variable to actions, whose rate depends on the state of the system through a function $\lambda : \mathbf{X} \rightarrow \mathbb{R}^+$.

Example 3.1. We will illustrate the notions introduced in the paper by means of an example coming from biological systems. Specifically, we consider a simple model of a (prokaryotic) genetic regulatory network with a single gene, expressing a protein acting, after dimerization, as a repressor of its own

³There are other probabilistic extensions of CCP studied in literature, like [26, 18, 17]. [26] provides CCP with a semantics based on discrete time Markov Chains, while in [18, 17] the stochastic ingredient is introduced by extending the store with random variables and adding a primitive for sampling. These approaches, however, are not suited for our purposes, as we need a model in which events happen probabilistically in continuous-time, as customary in biochemical modeling.

production. We assume a cooperative repression: two dimers are required to bind to the promoter region of the gene. The **sCCP** model is given by $\mathcal{A} = (A, \mathcal{D}, \mathbf{X}, \text{init})$, where the variables are $\mathbf{X} = \{X_p, X_{p2}\}$, storing the quantity of the protein p and of its dimer $p2$ and the components in \mathcal{D} are (* stands for true):

$$\begin{aligned}
 \text{gene}_0 &\stackrel{\text{def}}{=} [* \rightarrow X'_p = X_p + 1]_{k_{p1}} \cdot \text{gene}_0 + [X_{p2} > 0 \rightarrow *]_{k_{p1}X_{p2}} \cdot \text{gene}_1 \\
 \text{gene}_1 &\stackrel{\text{def}}{=} [* \rightarrow X'_p = X_p + 1]_{k_{p2}} \cdot \text{gene}_1 + [X_{p2} > 0 \rightarrow *]_{k_{p2}X_{p2}} \cdot \text{gene}_2 + \\
 &\quad [* \rightarrow *]_{k_{u1}} \cdot \text{gene}_0 \\
 \text{gene}_2 &\stackrel{\text{def}}{=} [* \rightarrow *]_{k_{u2}} \cdot \text{gene}_1 \\
 \text{deg} &\stackrel{\text{def}}{=} [* \rightarrow X'_p = X_p - 1]_{k_d X_p} \cdot \text{deg} \\
 \text{dimer} &\stackrel{\text{def}}{=} [* \rightarrow X'_p = X_p - 2 \wedge X'_{p2} = X_{p2} + 1]_{k_x X_p (X_p - 1) / 2} \cdot \text{dimer} + \\
 &\quad [* \rightarrow X'_p = X_p + 2 \wedge X'_{p2} = X_{p2} - 1]_{k_{-x} X_{p2}} \cdot \text{dimer}
 \end{aligned}$$

The initial network A is $\text{gene}_0 \parallel \text{deg} \parallel \text{dimer}$ with initial values of the store variables are given by

$$\text{init}(X_p, X_{p2}) = (X_p = 0) \wedge (X_{p2} = 0).$$

Notice: there is no need to introduce agents for proteins or dimers, as the quantity of these objects needs only to be measured by stream variables. The repression mechanism is represented by a gene unable of expressing a protein whenever in state gene_2 . We did not decrement X_{p2} before entering states gene_1 and gene_2 as we assume repression mechanism not requiring a binding of the dimer (inhibition by bumping).

Remark 3.1. The pros and cons of using **sCCP** as a modeling language for biological systems are discussed in detail in [6]. Basically, **sCCP** combines on one side the logical simplicity of process algebras and on the other side the computational power of constraints. As a matter of fact, the constraint store can be more general than that used in this paper, whereby more complex information (like spatiality) can be managed just by a simple programming activity. Further work is needed, however, to export the techniques developed here to a more general version of the store.

All agents definable in **sCCP**, i.e. all agents $C \stackrel{\text{def}}{=} M \in \mathcal{D}$,⁴ are *sequential*, i.e. they do not contain any occurrence of the parallel operator, whose usage is restricted at the upper level of the network. **sCCP** sequential agents can be seen as automata synchronizing on store variables and they can be conveniently represented as labeled graphs, called *Reduced Transition Systems* (RTS) (see [5]).

The steps to obtain an object suitable to our subsequent treatment are the following:

1. Define the collection of all possible states—the *derivative set* $\text{Der}(C)$ —and actions— $\text{action}(C)$ —of any sequential agent appearing in a **sCCP** program.
2. Restrict to **sCCP** *simple programs*, i.e. programs without multiple copies of the same agent running in parallel at the same time. Formally, it is required that the derivative sets of any two agents in parallel in the initial network are disjoint. This is only an apparent restriction, cf. [9] for a more detailed discussion.
3. Introduce the following multi-graph⁵ $\text{RTS}(C) = (S(C), E(C), \ell)$:
 - $S(C) = \text{Der}(C)$,
 - $E(C) = \{(\text{exit}(\pi), \text{enter}(\pi)) \mid \pi \in \text{action}(C)\}$,

⁴In the following, with a slight abuse of notation, we sometimes write $C \in \mathcal{D}$ for $C \stackrel{\text{def}}{=} M \in \mathcal{D}$.

⁵ $\text{exit}(\pi), \text{enter}(\pi), \text{guard}(\pi), \text{update}(\pi), \text{rate}(\pi)$ give the executing agent, the target agent, the guard, the update and the rate of an action π , respectively.

- $\ell(e) = (\text{guard}(\pi), \text{update}(\pi), \text{rate}(\pi))$, where π is the action defining $e \in E(C)$.

In Figure 1, we show the RTS for the agent gene_0 , defined in Example 3.1.

4. Introduce the notion of *extended sCCP* program

$$\mathcal{A}^+ = (A^+, \mathcal{D}^+, \mathbf{XU}\{P_C \mid C \in \mathcal{D}\}, \text{init}^+(\mathbf{XU}\{P_C \mid C \in \mathcal{D}\})),$$

in which a variable P_C for run-time recording the number of parallel copies of each agent $C \in \mathcal{D}$ is available, and prove \mathcal{A}^+ is isomorphic to \mathcal{A} (see [9] for further details).

Essentially, the last step is a technical trick that simplifies the overall treatment. The variable P_C counts the number of copies of C present in parallel within the system at a given point in time. To take into account the effects of transitions on agents, we modify updates and rate functions, by increasing/decreasing counter P_C relative to actions adding/removing a copy of C . The reason for introducing state variables will be apparent in next section. They are required to control a cluster of discrete states (continuously approximated) and the *real* value of a state variable will indicate the “tendency” of the system to be in that particular state.

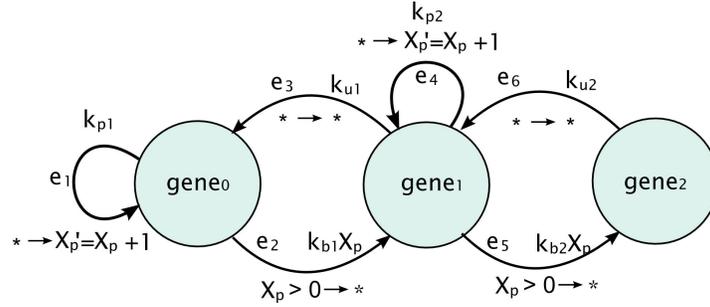


Figure 1: Reduced Transition Systems for the agent gene_0 defined in Example 3.1. Each edge is labeled by its rate function and by the guard and the update (depicted in the *guard* \rightarrow *update* notation).

4 From sCCP to TDSHA

In this section we define a semantics for **sCCP** in terms of TDSHA. The basic idea is to partition all possible transitions executable by an **sCCP** agent into two classes: those remaining discrete-stochastic and those to be approximated as continuous. Different partitions’ schemata correspond to different TDSHA. By parameterizing upon such schemata, we will obtain a *lattice* of different TDSHA’s.

Note that this approach is different from [9], as *we do not remove stochasticity*. Indeed, Stochastic Hybrid Automata can be seen as an intermediate layer between stochastic programs and (non-stochastic) hybrid systems. The reader is referred to [8] for further discussions.

The mapping proceeds in two steps. First we convert into TDSHA’s each sequential component of a **sCCP** program, then all these TDSHA’s are combined using the product construction.

Given a $\mathcal{A}^+ = (A^+, \mathcal{D}^+, \mathbf{Y}, \text{init}^+(\mathbf{Y}))$, let $C \in \mathcal{D}^+$ be one of the components of the initial network A , and let $\text{RTS}(C) = (S(C), E(C), \ell)$ be its RTS.

A specific continuous/discrete scheme of approximation is formalized by the choice of a boolean vector $\kappa \in \{0, 1\}^m$, $m = |E(C)|$, indexed by edges in $E(C)$: for $e \in E(C)$, $\kappa[e] = 1$ stands for a continuous

approximation of the transition, while $\kappa[e] = 0$ implies that the transition will remain discrete. Let $E(\kappa, C) = \{e \in E(C) \mid \kappa[e] = 1\}$ and $E(\neg\kappa, C) = \{e \in E(C) \mid \kappa[e] = 0\}$.

In order to guarantee that the vector field constructed from continuous transitions is sufficiently regular, we identify as *continuously approximable* only those actions π such that $rate(\pi)$ is differentiable and $rate(\pi)[\mathbf{X}] = 0$ whenever $guard(\pi)[\mathbf{X}]$ is false.⁶ We call *consistent* a vector κ such that $\kappa[e] = 1$ only for edges e that are continuously approximable. In the following, we suppose to work only with consistent κ .

At this point we are ready to introduce the basic components of our target TDSHA.

Discrete Modes. The modes of the TDSHA will be essentially the states $S(C)$ of the $RTS(C)$. However, as continuous transitions cannot change mode, we need to consider as equivalent those states that can be reached by a path of continuous edges. Let us denote by \sim_κ the equivalence relation among states of $S(C)$ relating two states if and only if they are connected by a path of continuous edges (i.e. edges in $E(\kappa, C)$ of the non-oriented version of $RTS(C)$). Let $S_\kappa(C) = S(C) / \sim_\kappa$. For each edge $e \in E(\kappa, C)$, we define the *stoichiometric vector* $\mathbf{v}_{\mathbf{Y},e}$ as an $|\mathbf{Y}|$ -vector, $\mathbf{Y} = \mathbf{X} \cup \{P_C \mid C \in \mathcal{D}\}$, such that $\mathbf{v}_{\mathbf{Y},e}[X] = h$ if and only if variable X is updated by transition e according to the formula $X' = X + h$.

Example. Consider the gene component of Example 3.1. Its RTS, shown in Figure 1, has three states, corresponding to the three components $gene_0$, $gene_1$, and $gene_2$, with state variables denoted by P_0 , P_1 , and P_2 , respectively. The RTS has also 6 transitions, indexed by e_1, \dots, e_6 . Consider the κ vector equal to $(1, 0, 0, 1, 1, 1)$: edges e_1, e_4, e_5, e_6 will be approximated as continuous, while the other three remain discrete. The relation \sim_κ has a quotient state space containing two classes: $S_1 = \{gene_0\}$ and $S_2 = \{gene_1, gene_2\}$. Such a partitioning of the gene's states can be seen as a way to render a slower dynamics for the binding/unbinding mechanism of the first repressor, to be compared to a faster one relative the second copy of the repressor.

Continuous flow. The continuous evolution for TDSHA is given by the following set of continuous transitions:

$$\mathfrak{T}\mathcal{C} = \{([exit(e)], \mathbf{v}_{\mathbf{Y},e}, rate(e) \mid e \in E(\kappa, C)\}.$$

Stochastic transitions. Stochastic transitions are defined in a very simple way, as guards and rates are basically copied from the **sCCP** edge. The only technicality is the definition of the reset.

Consider the state counting variables $\mathbf{P} = \{P_C \mid C \in \mathcal{D}\}$. They can assume values less than or equal to one, as the initial program is *simple*. Moreover, they range in the whole real-valued interval $[0, 1]$ whenever we are in a clustered state $[s]$ collapsing s_1, \dots, s_k of $RTS(C)$. In this case, the state variables P_{s_1}, \dots, P_{s_k} must sum exactly to 1, their value representing the likelihood of state s_1, \dots, s_k of the cluster $[s]$, respectively. In order to deal with state clusters correctly, we have to ensure that when a state $[s]$ is left, all its state variables are set to zero. Moreover, if a discrete transition looping in $[s]$ takes place, then the variable of its target state s_i must be set to 1, while all other variables of $[s]$ are to be reset to 0. To enforce this, consider an **sCCP** edge connecting states s_1 and s_2 , with

$$update(e) \stackrel{\text{def}}{=} \mathbf{X}' = f(\mathbf{X}) \wedge P'_{s_1} = P_{s_1} - 1 \wedge P'_{s_2} = P_{s_2} + 1,$$

⁶Guards of continuously approximable π are, in fact, redundant.

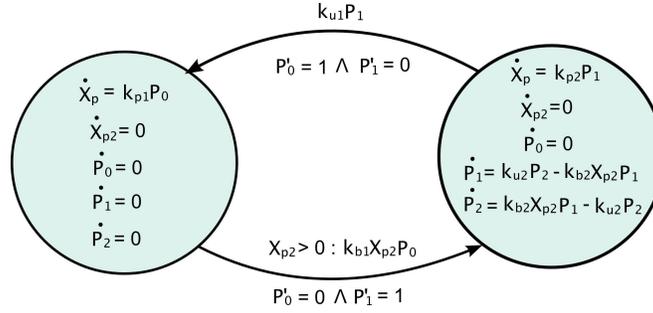


Figure 2: PDMP associated with the gene component of Example 3.1. Transitions approximated continuously determine a set of ODE, while discrete transitions are stochastic and are depicted here as edges of a graph (showing rates, guards and resets). The ODEs are obtained from continuous transitions according to the recipe of Section 2. We chose to display TDSHA in this format as it is similar to the classic representation of hybrid automata.

and define the function f_P on \mathbf{P} which is 1 on the component corresponding to P_{s_2} and zero elsewhere.

In this way, $\mathbf{P}' = f_P(\mathbf{P})$ implements the correct updating policy. Let now \bar{f} combine f and f_P : $\bar{f} \begin{pmatrix} \mathbf{X} \\ \mathbf{P} \end{pmatrix} = \begin{pmatrix} f(\mathbf{X}) \\ f_P(\mathbf{P}) \end{pmatrix}$.

Putting everything together, we have that the discrete transition associated with $e \in E(\neg\kappa, C)$ with $e = (s_1, s_2)$ is

$$([s_1], [s_2], \text{guard}(e), \mathbf{Y}' = \bar{f}(\mathbf{Y}), \text{rate}(e)) \in \mathfrak{TS}.$$

Instantaneous transitions. At this stage, there is no need to define instantaneous transitions. They will be used in Section 5 to deal with dynamic partitioning.

We can now collect all our considerations into the following definition.

Definition 4.1. Let $\mathcal{A} = (A, \mathcal{D}, \mathbf{X}, \text{init}_0)$ be a simple sCCP program and $\mathcal{A}^+ = (A^+, \mathcal{D}^+, \mathbf{Y}, \text{init}_0^+)$ be its extended version. Let C be a sequential component in parallel in A^+ , with $RTS(C) = (S(C), E(C), \ell)$. Fix a boolean vector $\kappa \in \{0, 1\}^m$, $m = |E(C)|$. The Transition-Driven Stochastic Hybrid Automaton associated with C with respect to κ is $\mathcal{S}(C, \kappa) = (Q, \mathbf{Y}, \mathfrak{TC}, \mathfrak{TD}, \mathfrak{TS}, \text{init})$, where

- $Q = S_\kappa(C) = S(C) / \sim_\kappa$;
- $\mathfrak{TC} = \{([\text{exit}(e)], v_{\mathbf{Y}, e}, \text{rate}(e)) \mid e \in E(\kappa, C)\}$;
- $\mathfrak{TD} = \emptyset$;
- $\mathfrak{TS} = \{([s_1], [s_2], \text{guard}(e), \mathbf{Y}' = \bar{f}(\mathbf{Y}), \text{rate}(e)) \mid e = (s_1, s_2) \in E(\neg\kappa, C)\}$;
- $\text{init} = \text{init}_0$.

Example. From the previous definition it is easy to generate the TDSHA relative to our running example above, in which $\kappa = (1, 0, 0, 1, 1, 1)$. Once we have the TDSHA, we can generate the corresponding PDMP (see supplementary material [1]), which is shown in Figure 2.

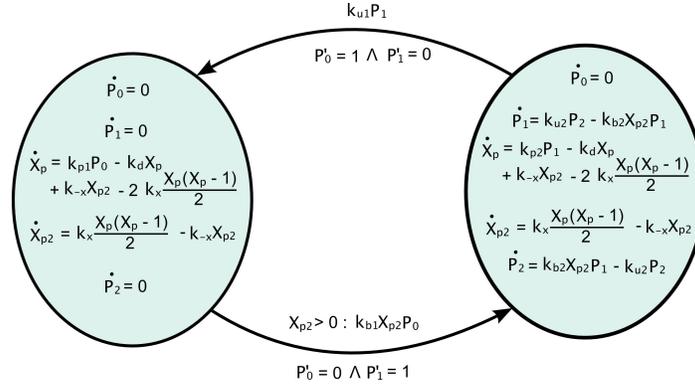


Figure 3: PDMP obtained from the product of TDSHA associated with the three components of the sCCP program of Example 3.1. See also the caption of Figure 2.

Definition 4.1 gives a way to associate a TDSHA with a sequential agent of a sCCP program. In order to define the TDSHA for the whole program, we will use the product construction.

Definition 4.2. Let $\mathcal{A} = (A, \mathcal{D}, \mathbf{X}, \text{init}_0)$ be a simple sCCP program and $\mathcal{A}^+ = (A^+, \mathcal{D}^+, \mathbf{Y}, \text{init}_0^+)$ be its extended version, with $A^+ = C_1 \parallel \dots \parallel C_n$. Fix a boolean vector κ_i for each sequential agent C_i . The Transition-Driven Hybrid Automaton for the sCCP program \mathcal{A} , with respect to $\kappa = (\kappa_i)_{i=1, \dots, n}$ is

$$\mathcal{T}(\mathcal{A}, \kappa) = \mathcal{T}(C_1, \kappa_1) \otimes \dots \otimes \mathcal{T}(C_n, \kappa_n).$$

Example. Consider again the sCCP model of Example 3.1. It has three components: gene, deg and dimer, with 6, 1, and 2 edges respectively. We consider three vectors $\kappa_1 = (1, 0, 0, 1, 1, 2)$, $\kappa_2 = (1)$, and $\kappa_3 = (1, 1)$. The product TDSHA of these three components generates the PDMP depicted in Figure 3.

4.1 Lattice of TDSHA

Definition 4.2 associates a TDSHA with a sCCP agent for a fixed partition, given by vector κ , of the transitions into discrete and continuous. Clearly, different choices of κ correspond to different TDSHA's, with a different degree of approximation of the original sCCP program. The different TDSHA's can be arranged into a lattice according to the following pre-order:

Definition 4.3. Let \mathcal{A} be a sCCP agent, then $\mathcal{T}(\mathcal{A}, \kappa_1) \sqsubseteq \mathcal{T}(\mathcal{A}, \kappa_2)$ if and only if $\kappa_1[e] = 1 \Rightarrow \kappa_2[e] = 1$, for each transition $e \in E(\mathcal{A}) = E(C_1) \cup \dots \cup E(C_n)$, with $A = C_1 \parallel \dots \parallel C_n$ the initial agent of \mathcal{A} .

The bottom element of this lattice is obtained for $\kappa \equiv 0$, while the top element is obtained for $\kappa[e] = 1$ if and only if e is continuously approximable. We remind to the reader that transitions not continuously approximable *must* be kept discrete.

The two “extreme” choices correspond to two particularly important TDSHA's, as shown in the following propositions.

Proposition 4.1. Let \mathcal{A} be a sCCP program. The TDSHA $\mathcal{T}(\mathcal{A}, 0)$ is the CTMC associated with \mathcal{A} by its standard semantics.

Proposition 4.2. Let \mathcal{A} be a sCCP program with initial agent $A = C_1 \parallel \dots \parallel C_n$. If e is continuously approximable for each $e \in E(\mathcal{A}) = E(C_1) \cup \dots \cup E(C_n)$, $\mathcal{T}(\mathcal{A}, \mathbf{1})$ coincides with the system of ODE's associated with \mathcal{A} by its fluid-flow approximation (see [5]).

5 Dynamic Partitioning of Transitions

In the previous sections we have defined a mapping from **sCCP** to TDSHA fixing the level of discreteness and continuity. This choice, however, can be difficult to perform *a priori*, as one does not know if the system will evolve to a state where a different approximation is more accurate.

This is particularly true when one deals with biological systems. In this case, reactions involving large populations of molecular species or having high rates, may be treated as continuous. However, such conditions depend on the state of the system and may change during its evolution. Indeed, there has been a growing attention on hybrid simulation strategies in systems biology, cf. next Section.

In order to have dynamic switching, we can extend the discrete modes, introducing states for each admissible vector κ . New discrete transitions need to be added as well, changing the value of κ according to some user defined conditions.

Intuitively, for each transition e the state space must be partitioned in two—possibly empty—regions: one where the edge e is treated as a continuous transition and one in which it is discrete. In order to define such regions, we consider two predicates:

1. $cont[e](\mathbf{X})$, encoding the condition to change edge e from discrete to continuous;
2. $disc[e](\mathbf{X})$, encoding the condition to change edge e from continuous to discrete.

An obvious requirement is that the regions identified by $cont[e]$ and $disc[e]$ should be disjoint. In order to define such predicates, we will consider a (sufficiently regular, usually continuous) function $f_e : \mathbb{R}^n \rightarrow \mathbb{R}$, whose sign will discriminate between continuous and discrete regions for edge e . More specifically, we define, for a fixed, small $\varepsilon > 0$.

1. $cont[e](\mathbf{x}) := f_e(\mathbf{x}) \geq \varepsilon$;
2. $disc[e](\mathbf{x}) := f_e(\mathbf{x}) \leq -\varepsilon$.

Using ε instead of 0, not only guarantees that the regions in which $cont[e]$ and $disc[e]$ are true are disjoint, but also avoids pathological situations of infinite sequences of instantaneous transitions. See supplementary material [1] for further details.

Now, suppose e is continuous. If the current trajectory enters in a region of the state space in which $disc[e]$ becomes true, then we must trigger an instantaneous transition in order to move from $\kappa_1[e] = 1$ to $\kappa_2[e] = 0$. All the variables must remain unchanged. However, in this case it may happen that the new relation \sim_{κ_2} splits in two the current mode $[s] \in S_{\kappa_1}(C)$, say $[s]_{\kappa_1} = [s_1]_{\kappa_2} \cup [s_2]_{\kappa_2}$. In this case, we need to introduce two instantaneous transitions, one going to $[s_1]$ and the other to $[s_2]$. Now, consider the value of the state variables of $[s]$, $P_{[s]} = \sum_{s_i \in [s]} P_{s_i}$. It can be proved that $P_{[s]} = 1$. Moreover, $P_{[s]} = P_{[s_1]} + P_{[s_2]}$ but, clearly, it is not necessarily the case that the two quantities on the right hand side of the equality are equal. This means that the system may “prefer” to move to states in $[s_1]$ than to those in $[s_2]$. This situation is correctly modeled using priorities, i.e. weighting transition to $[s_i]$ by $P_{[s_i]}$ and re-normalizing variables in $[s_1]$ and $[s_2]$ to maintain the property $P_{[s]} = 1$ for each $[s] \in S_{\kappa_2}(C)$.

We now give a formal definition for this construction, following a similar strategy as in Section 4: first we construct TDSHA for sequential components, then we apply the product construction to combine the local constructions. In order to fix the notation, consider the TDSHA $\mathcal{T}(C, \kappa) = (Q, \mathbf{Y}, \mathcal{T}\mathcal{C}, \mathcal{T}\mathcal{D}, \mathcal{T}\mathcal{S}, init)$ associated with a component C , with respect to a fixed κ . With Q_κ we indicate the set $Q_\kappa = \{([s]_\kappa) \mid [s] \in Q\}$. Moreover, $\mathcal{T}\mathcal{C}_\kappa$, $\mathcal{T}\mathcal{D}_\kappa$, and $\mathcal{T}\mathcal{S}_\kappa$ denote the sets $\mathcal{T}\mathcal{C}$, $\mathcal{T}\mathcal{D}$, and $\mathcal{T}\mathcal{S}$, respectively, with states in Q replaced by the corresponding states (equivalence classes) in Q_κ . A similar rule applies to $init_\kappa$.

Definition 5.1. Let $\mathcal{A} = (A, \mathcal{D}, \mathbf{X}, init_0)$ be a simple **sCCP** program and $\mathcal{A}^+ = (A^+, \mathcal{D}^+, \mathbf{Y}, init_0^+)$ be its extended version. Let C be a sequential agent in parallel in A^+ , with $RTS(C) = (S(C), E(C), \ell)$ and

$|E(C)| = m$. Moreover, let $cont[e], disc[e], e \in E(C)$ be defined as above. The TDSHA with dynamic partitioning associated with C is $\mathcal{T}(C, cont, disc) = (Q, \mathbf{Y}, \mathfrak{T}\mathcal{C}, \mathfrak{T}\mathcal{D}, \mathfrak{T}\mathcal{S}, init)$, with:

1. $Q = \bigcup_{\kappa \in \{0,1\}^m} Q_\kappa$;
2. $\mathfrak{T}\mathcal{C} = \bigcup_{\kappa \in \{0,1\}^m} \mathfrak{T}\mathcal{C}_\kappa$;
3. $\mathfrak{T}\mathcal{S} = \bigcup_{\kappa \in \{0,1\}^m} \mathfrak{T}\mathcal{S}_\kappa$;
4. $\mathfrak{T}\mathcal{D} = \bigcup_{\kappa \in \{0,1\}^m} \mathfrak{T}\mathcal{D}_\kappa \cup \mathfrak{T}\mathcal{D}_{0,1} \cup \mathfrak{T}\mathcal{D}_{1,0}$, where

$$\mathfrak{T}\mathcal{D}_{1,0} = \left\{ \left([s_1]_{\kappa_1}, [s_2]_{\kappa_2}, P_{[s_2]}, disc[e], \mathbf{Y}' = g(\mathbf{Y}) \right) \mid e \in E(C), \right. \\ \left. \kappa_1(e) = 1, \kappa_2(e) = 0, \kappa_1(e') = \kappa_2(e') \text{ for } e \neq e', [s_1]_{\kappa_1} \cap [s_2]_{\kappa_2} \neq \emptyset \right\}$$

where g assigns value $\frac{P_{s'}}{P_{[s]}}$ for $s' \in [s_2]_{\kappa_2}$, 0 to any other P_s , and it is the identity on \mathbf{X} . Moreover

$$\mathfrak{T}\mathcal{D}_{0,1} = \left\{ \left([s]_{\kappa_1}, [s]_{\kappa_2}, 1, cont(e), \mathbf{Y}' = \mathbf{Y} \right) \mid e \in E(C), \right. \\ \left. \kappa_1(e) = 0, \kappa_2(e) = 1, \kappa_1(e') = \kappa_2(e') \text{ for } e \neq e' \right\};$$

5. $init = init_0^+$;

Definition 5.2. Let $\mathcal{A} = (A, \mathcal{D}, \mathbf{X}, init_0)$ be a simple **sCCP** program and $\mathcal{A}^+ = (A^+, \mathcal{D}^+, \mathbf{Y}, init_0^+)$ be its extended version, with $A^+ = C_1 \parallel \dots \parallel C_n$. Moreover, fix predicates $cont_j[e], disc_j[e]$ for each sequential agent C_j of A^+ , according to Definition 5.1. The Transition-Driven Stochastic Hybrid Automata with dynamic partitioning for the **sCCP** program \mathcal{A} , with respect to $(cont_j, disc_j)_{j=1, \dots, n}$ is

$$\mathcal{T}(A, (cont_j, disc_j)_{j=1, \dots, n}) = \mathcal{T}(C_1, cont_1, disc_1) \otimes \dots \otimes \mathcal{T}(C_n, cont_n, disc_n).$$

Remark 5.1 (On the fly simulation). In Definition 5.2, the resulting TDHA has a number of modes exponential in the number of transitions that sequential agents can perform. This combinatorial explosion rules out the possibility of generating all the modes together. However, if we restrict to simulation, this is not a real issue, as we need to record only the current mode: the target mode of a transition can be generated on the fly as soon as the transition has been taken, given the knowledge of *RTS*.

5.1 Hybrid Simulation Strategies

The hybrid simulation algorithms proposed in literature [19, 24, 20, 28, 23, 16, 2] basically differ in two aspects: the kind of continuous dynamics (it can be based on ODE or SDE) and the rules for partitioning reactions into continuous and discrete (usually called fast and slow). More specifically, the partitioning can be static (done at the beginning of the simulation) or dynamic (i.e. recomputed at run-time).

Conditions for separating fast and slow reactions are usually twofold:

1. the *size of species* involved in the reaction must all be bigger than a given threshold. Usually, a fast reaction j must satisfy a condition like $x_i \geq K|v_{i,j}|$ for all species i involved in j , where v is the stoichiometric matrix.

2. the *rate function* of fast reactions must be reasonably bigger than that of slow reactions. Usually, the following constraint is enforced [28]: $\lambda_j(\mathbf{x})\Delta t \geq \Lambda$, which ensures that reaction j fires many times during the time step Δt . In [16] a different partition strategy imposes that rates of fast reactions are Λ times faster than the fastest slow reaction, so as to guarantee a separation of time scales.

Dynamical policies sketched above can be easily accounted for in our setting.

First of all, we need to start from an **sCCP** model of a biochemical network [6], in which reactions are modeled by action capabilities of agents. Then, applying the framework of this paper, we associate a TDSHA with such a model, together with a suitable policy for dynamic partitioning of transitions. All we have to do is define a function f_e for each **sCCP** transition e , such that $f_e(\mathbf{x}) > 0$ when the associated reaction can be considered fast and $f_e(\mathbf{x}) < 0$ when it is slow.

As an example, consider a partition strategy based only on the size of populations, like the one adopted in [24]. In this case, the function f_e for transition e can be the following:

$$f_e(\mathbf{x}) = \min\{x_i - K|v[x_i, e]| \mid v[x_i, e] \neq 0\},$$

where $v[\cdot, e]$ is the stoichiometry of action e , constructed as in Section 4, and K is a constant (that can be tuned for the specific system). Of course, more complex policies can be introduced by suitably modifying the functions f_e .

6 Conclusion and Further Directions

In this paper we provided a specific process algebra, **sCCP**, with a general semantics based on stochastic hybrid systems, parametric with respect to the degree of continuity and discreteness. The different hybrid models generated in this way can be arranged in a lattice, and we provided also a way to dynamically move within the lattice. This allows to formally describe hybrid simulation algorithms, opening up their use as tools to simulate process algebra-based models. Moreover, this approach gives the possibility of using other computational analysis methods than simulation, like reachability computations or model checking.

An interesting problem is how to extend such machinery to other process algebras. First steps have been done to deal with stochastic π -calculus [7], however the peculiarities of each language present specific difficulties to be solved.

The formal treatment developed in the paper, in particular the lattice of TDSHA defined in Section 4.1, can also provide an interesting theoretical framework to study the quality of the approximation and the error introduced. In particular, the mature theory of PDMP [12] can provide interesting tools in this direction.

Another issue we are investigating regards the relationships between discreteness and stochasticity. In particular, we are interested in understanding whether the stochastic ingredient of the dynamics can be dropped in favor of a pure discrete evolution, and at what price [8]. Motivations for this reside in the fact that non-stochastic hybrid systems have a much wider and more efficient set of automated reasoning tools available.

We conclude with a more basic (perhaps philosophical) question: given that a mix of continuous and discrete simulation strategy is the choice, is there a way—other than minimization of computational complexity—to determine which parts of the systems can/may be simulated discretely/continuously? We feel that physical consideration must be taken into account for addressing this issue and that these are probably outside our reach. However we wish to contribute the “computer scientist point of view”:

the level of discreteness/continuity can be established on the ground of a formal specification of the properties to verify/simulate and should guarantee the minimum of computational resources necessary to this task.

References

- [1] Supplementary material to the paper available online at: <http://www.dmi.units.it/~bortolu/sccp.htm>.
- [2] A. Alfonsi, E. Cances, G. Turinici, B. Di Ventura & W. Huisinga (2005): *Adaptive simulation of hybrid stochastic and deterministic models for biochemical systems*. In: *Proceedings of ESAIM*, 14. pp. 1–13.
- [3] L. Bortolussi (2006): *Stochastic Concurrent Constraint Programming*. In: *Proceedings of 4th International Workshop on Quantitative Aspects of Programming Languages (QAPL 2006)*, ENTCS 164. pp. 65–80.
- [4] L. Bortolussi & A. Policriti (2009): *Dynamical systems and stochastic programming - from Ordinary Differential Equations and back*. *Transactions of Computational Systems Biology*, in print.
- [5] L. Bortolussi & A. Policriti (2007): *Stochastic Concurrent Constraint Programming and Differential Equations*. In: *Proceedings of Fifth Workshop on Quantitative Aspects of Programming Languages, QAPL 2007*, ENTCS 167.
- [6] L. Bortolussi & A. Policriti (2008): *Modeling Biological Systems in Concurrent Constraint Programming*. *Constraints* 13(1).
- [7] L. Bortolussi & A. Policriti (2009): *Hybrid Dynamics of Stochastic π -calculus*. *Mathematics in Computer Science* 2(3), pp. 465–491.
- [8] L. Bortolussi & A. Policriti (2009): *Hybrid Dynamics of Stochastic Programs*. Submitted to *Theor. Comp. Sc.*
- [9] L. Bortolussi & A. Policriti (2009): *Stochastic Programs and Hybrid Automata for (Biological) Modeling*. In: *Proceedings of CiE 2009*.
- [10] M.L. Bujorianu & J. Lygeros (2004): *General Stochastic Hybrid Systems: Modeling and Optimal Control*. In: *Proceedings of 43rd IEEE Conference on Decision and Control (CDC 2004)*. pp. 182–187.
- [11] M. Calder, S. Gilmore & J. Hillston (2006): *Modelling the influence of RKIP on the ERK signalling pathway using the stochastic process algebra PEPA*. *Transactions on Computational Systems Biology* 4230, pp. 1–23.
- [12] M.H.A. Davis (1993): *Markov Models and Optimization*. Chapman & Hall.
- [13] D. Gillespie (2000): *The chemical Langevin equation*. *Journal of Chemical Physics* 113(1), pp. 297–306.
- [14] D.T. Gillespie (1976): *A General Method for Numerically Simulating the Stochastic Time Evolution of Coupled Chemical Reactions*. *J. of Computational Physics* 22.
- [15] D.T. Gillespie (1977): *Exact Stochastic Simulation of Coupled Chemical Reactions*. *J. of Physical Chemistry* 81(25).
- [16] M. Griffith, T. Courtney, J. Peccoud & W.H. Sanders (2006): *Dynamic partitioning for hybrid simulation of the bistable HIV-1 transactivation network*. *Bioinformatics* 22(22), pp. 2782–2789.
- [17] V. Gupta, R. Jagadeesan & P. Panangaden (1999): *Stochastic processes as concurrent constraint programs*. In: *Proceedings of POPL'99*.
- [18] V. Gupta, R. Jagadeesan & V.A. Saraswat (1997): *Probabilistic Concurrent Constraint Programming*. In: *Proceedings of CONCUR'97*.
- [19] E.L. Haseltine & J.B. Rawlings (2002): *Approximate simulation of coupled fast and slow reactions for stochastic chemical kinetics*. *Journal of Chemical Physics* 117(15).
- [20] E.L. Haseltine & J.B. Rawlings (2005): *On the origins of approximations for stochastic chemical kinetics*. *J. Chem. Phys.* 123.
- [21] T. A. Henzinger (1996): *The theory of hybrid automata*. In: *LICS '96: Proceedings of the 11th Annual IEEE Symposium on Logic in Computer Science*.

- [22] J. Hillston (2005): *Fluid Flow Approximation of PEPA models*. In: *Proceedings of the Second International Conference on the Quantitative Evaluation of Systems (QEST05)*.
- [23] T.R. Kiehl, R.M. Mattheyses & M.K. Simmons (2004): *Hybrid Simulation of Cellular Behavior*. *Bioinformatics* 20(3), pp. 316–322.
- [24] N. A. Neogi (2004): *Dynamic Partitioning of Large Discrete Event Biological Systems for Hybrid Simulation and Analysis*. In: *Proceedings of 7th International Workshop on Hybrid Systems: Computation and Control, HSCC 2004, LNCS 2993*, pp. 463–476.
- [25] J. Pahle (2009): *Biochemical simulations: stochastic, approximate stochastic and hybrid approaches*. *Brief Bioinform.* 10(1), pp. 53–64.
- [26] A. Di Pierro & H. Wiklicky (1998): *An operational semantics for probabilistic concurrent constraint programming*. In: *Proceedings of IEEE Computer Society International Conference on Computer Languages*.
- [27] A. Regev & E. Shapiro (2002): *Cellular Abstractions: Cells as Computation*. *Nature* 419.
- [28] H. Salis & Y. Kaznessis (2005): *Accurate hybrid stochastic simulation of a system of coupled chemical or biochemical reactions*. *Journal of Chemical Physics* 122.
- [29] V. Saraswat & M. Rinard (1990): *Concurrent Constraint Programming*. In: *Proceedings of 18th Symposium on Principles Of Programming Languages (POPL)*.
- [30] V. A. Saraswat (1993): *Concurrent Constraint Programming*. MIT press.
- [31] S. Strubbe & A. van der Schaft (2007): *Stochastic Hybrid Systems*, chapter *Compositional Modeling of Stochastic Hybrid Systems*, pp. 47–78. CRC Press.
- [32] D. J. Wilkinson (2006): *Stochastic Modelling for Systems Biology*. Chapman & Hall.