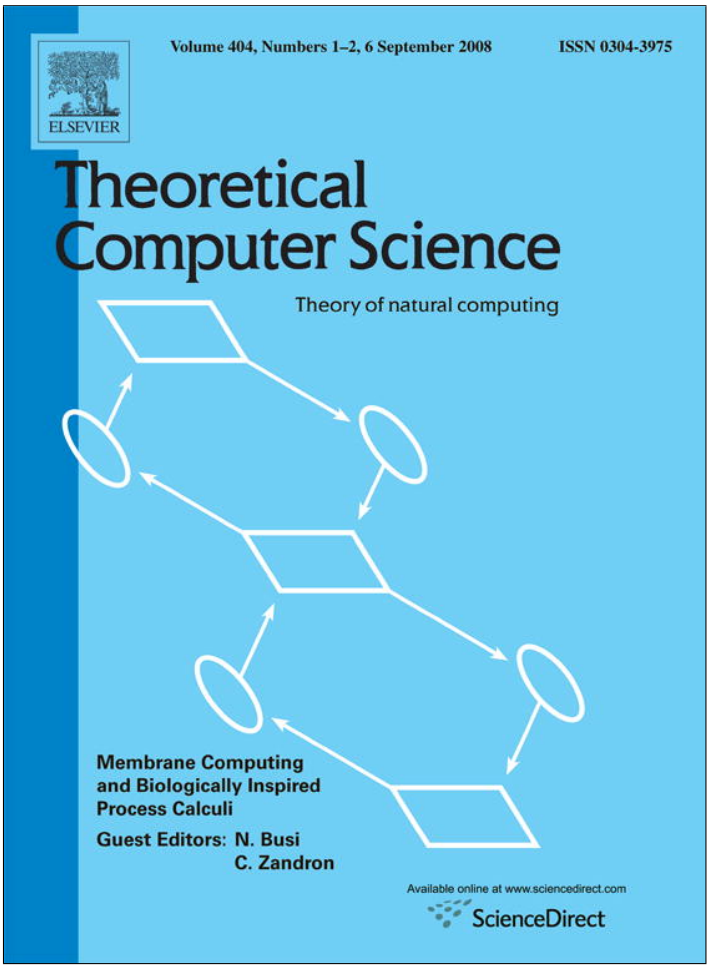

Theoretical Computer Science 404 (2008) 97–111

Contents lists available at ScienceDirect

Theoretical Computer Science

journal homepage: www.elsevier.com/locate/tcs

Relating continuous and discrete PEPA models of signalling pathways

Nil Geisweiller, Jane Hillston *, Marco Stenico

LFCS, School of Informatics, University of Edinburgh, UK

ARTICLE INFO

Keywords:
PEPA
ODEs
CTMC
Kurtz

ABSTRACT

PEPA and its semantics have recently been extended to model biological systems. In order to cope with massive quantities of processes (as is usually the case when considering biological reactions) the model is interpreted in terms of a small set of coupled ordinary differential equations (ODEs) instead of a large state space continuous time Markov chain (CTMC). So far the relationship between these two semantics of PEPA had not been established. This is the goal of the present paper. After introducing a new extension of PEPA, denoted PEPA $+\ \Pi$, that allows models to capture both mass action law and bounded capacity law cooperations, the relationship between these two semantics is demonstrated. The result relies on Kurtz's Theorem that expresses that a set of ODEs can be, in some sense, considered as the limit of pure jump Markov processes.

1. Introduction

In this paper we consider the problem of reconciling the discrete state space continuous time Markov chain (CTMC) representation of a biochemical signalling pathway which is generated by a process algebra model such as PEPA, and the continuous state space ordinary differential equation (ODE) representation which is more usually chosen by biochemists. This resolution is important because using the CTMC model we are able to conduct a variety of analysis which are not accessible from the ODE representation. For example, we can verify properties of the system via model checking, using the stochastic temporal logic CSL. However the validity of the model checking relies on agreement between the discrete and continuous models.

The stochastic process algebra PEPA has previously been shown to be useful for modelling biochemical signalling pathways [3,5,2]. In this work a more abstract mapping between biochemical elements and processes is made than the earlier work using the stochastic π-calculus [21]. In that work a correspondence is drawn between *molecules* and processes. The local states of the processes capture the transformations which the molecule may be subject to during the course of a pathway, e.g. phosphorylation, compound formation, ubiquitination. Since realistic pathways are comprised of thousands of molecules the CTMCs which are generated by models are amenable only to solution via stochastic simulation using Gillespie's algorithm [10]. In contrast, in the PEPA approach a correspondence is drawn between *species* and processes. Now, the local states of the processes capture levels of *concentration* of the species, and distinct biochemical elements, such as compounds, are represented in distinct process algebra components.

In such models the continuous variable, *concentration*, associated with each species in the ODEs, is discretised into a number of *levels*. Thus each component representing a species has a distinct local state for each level of concentration. The more levels that are incorporated into the model, i.e. the finer the granularity of the discretisation, the closer we would expect the results of the CTMC to be to the ODE model. However, this relationship has not previously been established: that is the major contribution of this paper. In earlier work we have shown how an ODE model can be derived from the PEPA

* Corresponding author.
E-mail addresses: ngeiswei@inf.ed.ac.uk (N. Geisweiller), jeh@inf.ed.ac.uk (J. Hillston), m.stenico@sms.ed.ac.uk (M. Stenico).

0304-3975/$ – see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.tcs.2008.04.012

description of a system [16]. Now we are able to show that this set of ODEs captures the limiting behaviour of the CTMCs representing the discretised system. This is based on an earlier result by Kurtz [17] in which the author shows that when certain conditions are satisfied a sequence of CTMCs converge to a set of ODEs. Our contribution is to show that the class of PEPA models used to describe biochemical signalling pathways in the *pathway-centric* style [3] give rise to CTMCs which satisfy these conditions and that the set of ODEs automatically extracted from the PEPA description (following [16]) are the limit of this sequence.

The version of PEPA which we use in this paper is a slight extension based on our experience modelling biochemical pathways. We introduce a new form of the cooperation combinator which represents shared actions with distinct forms of kinetics. The first, based on the notion of *bounded capacity*, is the usual PEPA cooperation at the rate of the slowest participant [14]. The second, based on the notion of *mass action* familiar to biochemists, takes into account the multiple possibilities which may arise and the consequences of this for the observed rate. We demonstrate that this new combinator suits the development of models of biochemical signalling pathways in the pathway-centric style. This is a subsidiary contribution of the paper although we do not show properties of the combinator here due to space limitations. A discussion of its semantics can be found in [9].

The remainder of this paper is structured as follows. Section 2 recalls and extends PEPA with the mass action kinetics. Sections 3 and 4 consider the issue of representing and generating the sets of ODEs of biochemical signalling pathways from PEPA models. In Section 5 we show how to use Kurtz's Theorem to establish the relationship between the two semantics of PEPA considered in this paper, in terms of CTMCs and in terms of sets of coupled ODEs. Then, in Section 6, this result is illustrated by a small example of a biochemical signalling pathway. Finally, Section 7 concludes.

2. PEPA + Π

The PEPA modelling language was introduced in [15] and has been used extensively to represent a wide range of computer and communication systems. In its original form, a rate is associated with each activity, which is assumed to be the parameter of an exponential distribution characterising the delay associated with the activity. Under this interpretation, the structured operational semantics of the language gives rise to a labelled multi-transition system which can be regarded as the state transition diagram of an underlying continuous time Markov chain (CTMC). (The interested reader is referred to [15] for more detail.)

A multi-transition system is used because the multiplicity of actions has an impact on the dynamic behaviour. The state transition diagram of the CTMC is characterised by an *infinitesimal generator* matrix Q in which the entry q_{ij} records the transition rate from state i to state j. From the infinitesimal generator matrix various metrics related to the dynamic behaviour of the CTMC can be derived using linear algebra (the so-called *numerical solution*). These include *transient* measures, for example giving the probability of an event after a given time, or *steady state* measures relating to when the process has reached a limiting or equilibrium behaviour, if this is possible (i.e. if the CTMC is *ergodic*).

The PEPA language has a small number of combinators which we outline informally below; the operational semantics can be found in [15]. As a necessary condition to generate ergodic CTMCs, PEPA models are typically constructed according to the following grammar:

$$S := (\alpha, r).S \mid S + S \mid C$$
$$P := P \underset{L}{\bowtie} P \mid P/L \mid S$$

where S denotes a *sequential component*, P a *model component* and C is a constant defined by a declaration such as

$$C \stackrel{\text{def}}{=} S.$$

$(\alpha, r).S$ carries out activity (α, r), which has action type α and an exponentially distributed duration with parameter r, and it subsequently behaves as S. The component $P + Q$ represents a system which may behave either as P or as Q. The activities of both P and Q are enabled. The first activity to complete distinguishes one of them: the other is discarded. The system will behave as the derivative resulting from the evolution of the chosen component. PEPA supports multi-way cooperations between components: the result of synchronising on an activity α is thus another α, available for further synchronisation. We write $P \underset{L}{\bowtie} Q$ to denote cooperation between P and Q over L. The set which is used as the subscript to the cooperation symbol, the *cooperation set* L, determines those activities on which the *cooperands* are forced to synchronise. For action types not in L, the components proceed independently and concurrently with their enabled activities. We write $P \parallel Q$ as an abbreviation for $P \underset{L}{\bowtie} Q$ when L is empty. The rate of the shared activity which results from cooperation is determined according to the notion of *bounded capacity* [14]. This states that a component cannot be made to go faster than its own capacity by carrying out an action in cooperation. The capacity of a component P to carry out an action of type α is termed the *apparent rate* of α in P and is denoted $r_\alpha(P)$. It is the sum of all the rates of all the α activities enabled in P. It follows that the apparent rate of a shared activity is the *minimum* of the apparent rates of activities of that type in the synchronising components [15]. For example, consider communication through a channel. The rate at which the channel transmits data, the rate at which the sender puts data onto the channel and the rate at which the receiver extracts data from the channel may all differ. But the actual rate of communication will be limited by the slowest one. P/L denotes the component P in

which all actions with types in L are *hidden* meaning that their type is no longer visible but is replaced by the distinguished type τ. We do not consider hiding in the remainder of this paper.

2.1. PEPA with mass action cooperation

In more recent work PEPA has been used to model biochemical signalling pathways. In this context we have found that the bounded capacity kinetics outlined above is not always the most appropriate way to determine the rate of a shared action, although in some cases it is. Thus we have introduced an alternative form of cooperation in PEPA which is designed to allow some cooperations to be made according to *mass action kinetics*, rather than the bounded capacity kinetics explained above [9].

The law of mass action, used by biochemists, states that for a reaction in a homogeneous, free medium, the reaction rate will be proportional to the concentrations of the individual reactants involved. Here we define an extension of PEPA that can capture mass action kinetics, i.e. in which the rate of the shared activity is *multiplied* by the capacity of each component to participate.

The new combinator, which replaces $\underset{L}{\bowtie}$, and encompasses both forms of kinetics, is represented as follows:

$$P \overset{J}{\underset{L}{\bowtie}} Q.$$

The cooperation operator uses two sets of actions, J and L.

- L contains the cooperating action types that follow the standard semantics of PEPA, that is, which respect the definition of the bounded capacity kinetics:
  - when two components P and Q cooperate on an action α, synchronising $P \xrightarrow{(\alpha,r_1)}$ and $Q \xrightarrow{(\alpha,r_2)}$ respectively, then the resulting activity $P \overset{J}{\underset{L}{\bowtie}} Q \xrightarrow{(\alpha,R)} (\alpha \in L)$ has rate R defined by:
    $$R = \frac{r_1}{r_\alpha(P)} \times \frac{r_2}{r_\alpha(Q)} \times \min(r_\alpha(P), r_\alpha(Q)).$$
  - passive actions have rate $\top$ ("top") and have no influence on the rate of the shared activity.
- J contains the cooperation actions that follow the mass action kinetics:
  - when two components P and Q cooperate on an action α, synchronising $P \xrightarrow{(\alpha,r_1)}$ and $Q \xrightarrow{(\alpha,r_2)}$ respectively, then the resulting activity $P \overset{J}{\underset{L}{\bowtie}} Q \xrightarrow{(\alpha,R)} (\alpha \in J)$ has rate R defined by:
    $$R = r_1 \times r_2.$$
  - furthermore, the apparent rate is defined as:
    $$r_\alpha(P \overset{J}{\underset{L}{\bowtie}} Q) = r_\alpha(P) \times r_\alpha(Q) \quad \text{for } \alpha \in J$$
  - passive actions have rate 1 or more generally a constant used as *scaling factor* (see Section 5.2 for more explanations).

We also introduce some additional notation for the case when there is a large number of repeated components. This notation is a shorthand for a description for a number of identical, independent processes in parallel (i.e. without *any* cooperation between them):

$$P[n] = \underbrace{P \parallel \ldots \parallel P}_{n}.$$

In this case it is assumed that the apparent rate of an activity is derived as in the standard PEPA semantics, i.e. $r_\alpha(P[n]) = n \times r_\alpha(P)$, since $\parallel$ is shorthand for $\underset{\emptyset}{\bowtie}$.

We also extend the previous algorithm for constructing a set of coupled ODEs from the PEPA model, capturing the evolution of the number of components in each of the different component states with respect to time. This is presented in detail in Section 4. In conjunction with this we slightly modify the grammar of the model components as follows:

$$P := P \overset{J}{\underset{L}{\bowtie}} P \mid (S_1[n_1] \parallel \ldots \parallel S_k[n_k])$$

where $S_1, \ldots, S_k$ define all local derivatives of the sequential process S.

3. Modelling biochemical signalling pathways

Work on applying formal system description techniques from computer science to biochemical signalling pathways was initially stimulated by Goss and Peccoud [12], Regev [22] and Priami et al. [21]. Subsequently there has been much work in which the stochastic π-calculus is used to model biological systems, for example [8,19,20]. This work is based on a

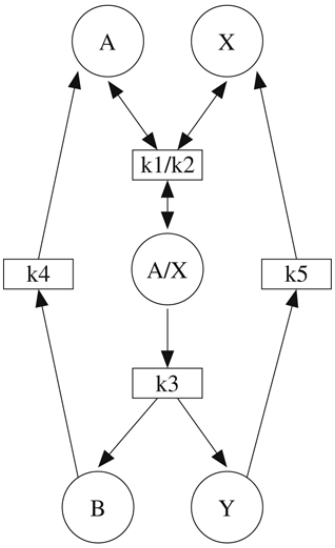

Fig. 1. Small synthetic example pathway.

correspondence between molecules and processes. Each *molecule* in a signalling pathway is represented by a component in the process algebra representation. The local states of the component correspond to the physical changes which a molecule will undergo in the course of biochemical reactions. Thus, if a complex C is formed from molecules A and B, two process algebra components A and B will interact ("communicate") and one will evolve into a C, the other becoming null. In order to represent a system with populations of molecules, many copies of the process algebra components are needed. This leads to underlying CTMC models with enormous state spaces — the only possible solution technique is simulation based on Gillespie's algorithm, as presented in [10].

In contrast, many systems biology models are developed as sets of nonlinear ordinary differential equations. The variables of the equations are *concentrations* of the involved species and there is one equation for each species in the pathway, recording the impact of each reaction on the species. Nonlinear terms arise from interactions between species.

Recent work on PEPA has investigated a more abstract way of mapping biochemical signalling pathways into a process algebra ([3]). Rather than a correspondence between molecules and components, we have proposed a correspondence between *species* or *subpathways* and components (c.f. modelling classes rather than individual objects). Now the components in the process algebra model capture a pattern of behaviour of a whole set of molecules, rather than the identical behaviour of thousands of molecules having to be represented individually. The local states of the components now correspond to the concentrations of species represented in the ODEs but discretised into a number of *"levels"*. Assuming a fixed maximum concentration level for each species, we refer to the distance (in concentration) between two successive levels of concentration distinguished in a model, as the *granularity* of the model.

From such PEPA models we are able to generate underlying mathematical models, suitable for analysis, in a number of different ways. The usual semantics of PEPA gives rise to a CTMC which can be solved numerically (if state space size does not prohibit it) as in [15]. Here each state of the CTMC corresponds to a discrete level of concentration for each chemical species in the pathway. For PEPA models based on modelling species and their concentrations [4] showed how a model with two levels of concentration could be used to generate a set of nonlinear ODEs. Moreover, such models can also be used to derive CTMCs in which molecules are represented individually, suitable for Gillespie simulation. In [16], Hillston showed how a set of nonlinear ODEs can be derived from more general PEPA models (see Section 4 for more detail). These underlying mathematical models have different strengths offering different forms of analysis. The relationship between Gillespie-style molecular simulations and ODEs has been established in the thermodynamic limit [18] but the relationship between the CTMCs with levels of concentrations, such as arise from PEPA models, and ODEs has not previously been established. It is this problem which we address in this paper.

Even within our abstract approach to modelling there are alternative ways of expressing the model [3]. We distinguish these as *reagent-centric* and *pathway-centric*. In a reagent-centric model we treat each distinct reagent or species in the pathway as a distinct component type as described above. The component definition then captures the possible reactions that the reagent may be involved in. The local states of the components correspond to differing levels of concentration and the process definition records the impact of each reaction type on the concentration of the reagent—it will either increase the concentration, moving it up a level, decrease it, moving the state down a level, or leave it unchanged. In a pathway-centric model we focus instead on the transformations which a reagent or species with non-zero initial concentration may undergo through the course of a pathway (phosphorylation, complex formation etc.). Each such subpathway is then represented as a distinct component in the model. Local states now correspond to the states which the physical entity may find itself in through the subpathway. Differing levels of concentration are represented in the global state by having differing multiples of components of a particular pathway type.

As a small example we consider the pathway shown in Fig. 1.

This is comprised of the following kinetic reactions:

$$A + X \underset{k2}{\overset{k1}{\rightleftharpoons}} A/X \xrightarrow{k3} B + Y$$
$$B \xrightarrow{k4} A$$
$$Y \xrightarrow{k5} X.$$

We assume an initial positive concentration of reagents A and X, all other reagents initially being absent.

Reagent-centric model. Modelled in the reagent-centric style with the coarsest possible granularity, i.e. just two levels, the pathway in Fig. 1 is represented by the following declarations:

$$A_H \overset{\text{def}}{=} (k1react, k1).A_L$$
$$A_L \overset{\text{def}}{=} (k2react, k2).A_H + (k4react, k4).A_H$$
$$X_H \overset{\text{def}}{=} (k1react, k1).X_L$$
$$X_L \overset{\text{def}}{=} (k2react, k2).X_H + (k5react, k5).X_H$$
$$A/X_H \overset{\text{def}}{=} (k2react, k2).A/X_L + (k3react, k3).A/X_L$$
$$A/X_L \overset{\text{def}}{=} (k1react, k1).A/X_H$$
$$B_H \overset{\text{def}}{=} (k4react, k4).B_L$$
$$B_L \overset{\text{def}}{=} (k3react, k3).B_H$$
$$Y_H \overset{\text{def}}{=} (k5react, k5).Y_L$$
$$Y_L \overset{\text{def}}{=} (k3react, k3).Y_H.$$

The complete model is the interaction of these components constrained by cooperation to share the appropriate actions:

$$(((A_H \underset{\{k1react,k2react\}}{\bowtie} X_H) \underset{\{k1react,k2react\}}{\bowtie} A/X_L) \underset{\{k3react,k4react\}}{\bowtie} B_L) \underset{\{k3react,k5react\}}{\bowtie} Y_L.$$

More details of this style of representation can be found in [3].

Pathway-centric model. Modelled in the pathway-centric style the pathway in Fig. 1 is represented by the following declarations:

$$A = (k1react, k_1).A/X$$
$$A/X = (k2react, k_2).A + (k3react, k_3).B$$
$$B = (k4react, k_4).A$$
$$X = (k1react, H).X/A$$
$$X/A = (k2react, k_2).X + (k3react, k_3).Y$$
$$Y = (k5react, k_5).X$$

where we have two distinct subpathways, corresponding to A and X respectively. We also need the following system equation to complete the model:

$$(A[n_{1_1}] \parallel A/X[n_{1_2}] \parallel B[n_{1_3}]) \overset{\{k1react\}}{\underset{\{k2react,k3react\}}{\bowtie}} (X[n_{2_1}] \parallel X/A[n_{2_2}] \parallel Y[n_{2_3}]).$$

The rate H of the passive transition of X, $(k1react, H).X/A$ corresponds to the granularity of the discretisation which is determined by the maximal concentration that any reagent can reach during the lifetime of the system divided by the number of levels of concentration (the reason for putting H as rate of the passive transitions is explained in detail in Section 5.2).

In this paper we focus on models presented in the pathway-centric style.

Note that we only model reactions in which the stoichiometric coefficient of all participants in the reaction is one. This restriction is placed by the current syntax of PEPA which cannot express higher order reaction.[1]

3.1. Related work

Whilst a significant body of work is developing on modelling biochemical systems with stochastic process algebras and related formalisms (e.g. [12,22,21,8,19,20]) in most cases the modelling is carried out at a more detailed and less

[1] The current syntax of PEPA + Π does not allow reactions such as $3A + 2B \rightarrow 4C$ to be expressed.

abstract level. Consequently, most analysis of such systems is carried out using Gillespie's stochastic simulation and similar approaches [11,20]. To the best of our knowledge no other authors have considered the relationship between ODE and CTMC models in the context of stochastic process algebras. The original relationship between the two was established by Kurtz in 1970 [17], and considered in the context of chemical reactions in 1972 [18]. The mapping from process algebra models to ODEs has recently been considered by Cardelli [7] and Bortolussi and Policriti [1], but not the relationship with a CTMC.

4. Deriving ODEs from PEPA models

In [16] Hillston presented a compact state representation for PEPA models: the *numerical vector form*. The objective of introducing this new form was to facilitate a fluid approximation of the CTMC underlying a PEPA model when each component type in the model is replicated a large number of times. In this paper we seek to establish the validity of this approximation. As a preliminary we give an overview of how the approximation is formed.

In process algebra models the usual state representation is in terms of the syntactic forms of the model expression. The structured operational semantics define how a model may evolve and these may be applied exhaustively to form a *labelled transition system* (usually termed the *derivation graph* in PEPA) representing the state space of the model. This is a graph in which each node is a distinct syntactic form or *derivative* (or equivalence class of syntactic expressions up to strong equivalence) and each arc represents a possible activity causing the state change.

Rather than the complete syntactic form, since the static cooperation combinators remain unchanged in all states, it is often convenient to represent the states of the model in *vector form*. The state vector records one entry for each sequential component of the PEPA model. These components will be present in each derivative of the model, although they will change their local state or derivative. Thus the global state can be represented as a vector or sequence of local derivatives. For the remainder of this paper we use the term *local derivative* to refer to the local state of a single sequential component, whereas *derivative* will be used to refer to a global state represented in its syntactic form.

In [16] Hillston proposed an alternative vector form for capturing the state information of models with repeated components. In the state vector form there is one entry in the vector for each sequential component in the model. When the number of repeated components becomes large this can be prohibitively expensive in terms of storage. In the alternative vector form there is one entry for each local derivative of each type of component in the model. Two components have the same type if their derivation graphs are isomorphic. The entries in the vector are no longer syntactic terms representing the local derivative of the sequential component, but the *number* of components currently exhibiting this local derivative.

The numerical vector form for an arbitrary PEPA model is defined as follows.

Definition 4.1 (*Numerical Vector Form*). For an arbitrary PEPA model $\mathcal{M}$ with n component types $\mathcal{C}_i$, $i = 1, \ldots, n$, each with N_i distinct derivatives, the *numerical vector form* of $\mathcal{M}$, $\mathcal{V}(\mathcal{M})$, is a vector with $N = \sum_{i=1}^{n} N_i$ entries. The entry v_{i_j} records how many instances of the jth local derivative of component type $\mathcal{C}_i$ are exhibited in the current state.

Since we assume that each component type in the model is replicated a number of times (reflecting the granularity of the discretisation), the domain of values of each entry in $\mathcal{V}(\mathcal{M})$ can be large. If K_i is the number of components of type $\mathcal{C}_i$ in the initial configuration of the model (i.e. the range of concentrations for $\mathcal{C}_i$ is discretised into $K_i + 1$ levels) then each entry in the ith subvector will have domain $0, \ldots, K_i$.

Consider an arbitrary state $\mathcal{M}'$ of the model $\mathcal{M}$ which has the particular numerical vector representation $V(\mathcal{M}')$. When a state change occurs it can happen in two distinct ways:

- A single sequential component, an instance of component type $\mathcal{C}_i$ may engage in an *individual action*. In this case the impact on $V(\mathcal{M}')$ is that within the ith subvector one entry is incremented by one while another is decremented by one, reflecting the evolution of this single component from one local derivative to another.
- Alternatively a *shared action* may be performed resulting in the simultaneous evolution of two or more sequential components of distinct types (since we assume that replicated components are independent of each other). Thus a number of distinct subvectors may need to be updated within $V(\mathcal{M}')$. However in each case one entry is incremented by one and one entry is decremented by one.

The system is inherently discrete with the entries within the numerical vector form always being non-negative integers and always being incremented or decremented in steps of one. When the numbers of components are large (i.e. the granularity of discretisation is fine) these steps are relatively small and we can approximate the behaviour by considering the movement between states to be continuous, rather than occurring in discontinuous jumps. Thus we replace the discrete event system represented by the derivation graph of a PEPA process by a continuous model, represented by a set of coupled ordinary differential equations. The numerical vector form of state representation is an intermediate step towards this objective.

4.1. Assumptions and definitions regarding the PEPA models considered

PEPA allows the same activity to have different rates in different components that share the activity. However, in the context of biological kinetic reactions, the rate characterising a reaction between two species is given *a priori* (obtained by measurements or expertise) but not as the result of the product (or the minimum) of some cooperating rates.

Based on this and the conditions imposed in [16] we make the following set of assumptions for the remainder of the paper. These assumptions simplify the form of the given results without reducing the expressiveness or usability of PEPA in the context of biological pathway modelling. The considered assumptions are listed below:

(1) Any action α used for cooperation is involved only once in a component (i.e. α appears in only one local derivative) and all components containing the action α are synchronised, according either to the bounded capacity kinetics or to the mass action kinetics.
(2) Each component type is declared only once in the system equation, where the declaration of a component $\mathcal{C}_i$ with N_i local derivatives, denoted $\mathcal{C}_{i_1}, \ldots, \mathcal{C}_{i_{N_i}}$, initialised with $n_{i_1}, \ldots, n_{i_{N_i}}$ copies respectively, is denoted:

$$\mathcal{C}_{i_1}[n_{i_1}] \parallel \ldots \parallel \mathcal{C}_{i_{N_i}}[n_{i_{N_i}}].$$

Intuitively this assumption can be interpreted as all copies of each component type are in the same compartment and (as a consequence of the previous assumption) that all copies of all component types are in the same compartment. Moreover, cooperation within groups of components of the same type is not allowed.
(3) Any activity is defined by a single rate. Consequently, when the reaction is the result of the cooperation of several components, either all rates in all transitions are equal (in the case of bounded capacity kinetics), or only one transition defines the rate and the others have rate H, the scaling factor (in the case of mass action kinetics).
(4) All reactions are associated with a visible action, and action hiding is not considered.

These assumptions set, we define the following functions and notation:

(1) Let *coop* be a function that records, for each action type, the form of its cooperation kinetics.

$$coop : Act \mapsto \{\epsilon, \Pi, \min\}$$

where $coop(\alpha) = \epsilon$ means that the action α is not involved in any cooperation; $coop(\alpha) = \Pi$ means that cooperation on α follows mass action kinetics; and $coop(\alpha) = \min$ means that cooperation on α follows bounded capacity kinetics. Note that *coop* is unambiguously defined by the system equation according to the position of action type α with respect to the cooperation operation, i.e. according to whether it is in the upper or lower cooperation sets, or neither.
(2) Each action type α is associated with a unique activity rate, denoted $rate(\alpha)$.
(3) Consider a local derivative D of a sequential component. An action type α is an *exit activity* of D if D enables α, i.e. there is a transition $D \xrightarrow{(\alpha, r_\alpha)}$ in the labelled transition system of D. We denote the set of exit activities of D by $Ex(D)$. Conversely, we denote the set of local derivatives for which α is an exit activity by $pre(\alpha)$.
(4) Similarly, an action type β is an *entry activity* of D if there is a derivative D' which enables β and D is the one-step β-derivative of D', i.e. $D' \xrightarrow{(\beta, r_\beta)} D$ is in the labelled transition system of D'. We use $En(D)$ to denote the set of entry activities of D.

This classification of activities helps us to record the impact of each activity on each local derivative, C_{i_j}, in the model. Recall that in the vector form the numbers of these derivatives, $N(C_{i_j})$, have become our state variables.

Let us consider the evolution of the numerical state vector. Let $v_{i_j}(t) = N(\mathcal{C}_{i_j}, t)$ denote the jth entry of the ith subvector at time t, i.e. the number of instances of the jth local derivative of sequential component $\mathcal{C}_i$. In a short time δt the change to this arbitrary vector entry will be:

$$N(\mathcal{C}_{i_j}, t+\delta t) - N(\mathcal{C}_{i_j}, t) = \underbrace{-\sum_{\alpha \in Ex(\mathcal{C}_{i_j})} rate(\alpha) \times \underset{\mathcal{C}_{k_l} \in pre(\alpha)}{coop(\alpha)}(N(\mathcal{C}_{k_l}, t))\,\delta t}_{\text{exit activities}} + \underbrace{\sum_{\alpha \in En(\mathcal{C}_{i_j})} rate(\alpha) \times \underset{\mathcal{C}_{k_l} \in pre(\alpha)}{coop(\alpha)}(N(\mathcal{C}_{k_l}, t))\,\delta t}_{\text{entry activities}} \tag{1}$$

where $\underset{\mathcal{C}_{k_l} \in pre(\alpha)}{coop(\alpha)}\left(N(\mathcal{C}_{k_l}, t)\right)$ is defined as follows:

$$\underset{\mathcal{C}_{k_l} \in pre(\alpha)}{coop(\alpha)}\left(N(\mathcal{C}_{k_l}, t)\right) = \begin{cases} \Pi_{\mathcal{C}_{k_l} \in pre(\alpha)} N(\mathcal{C}_{k_l}, t) & \text{if } coop(\alpha) = \Pi \\ \min_{\mathcal{C}_{k_l} \in pre(\alpha)} N(\mathcal{C}_{k_l}, t) & \text{if } coop(\alpha) = \min \\ N(\mathcal{C}_{k_l}, t) & \text{if } coop(\alpha) = \epsilon. \end{cases}$$

In Eq. (1) the first term records the impact of exit activities. If the exit activity is an individual activity of this component $coop(\alpha) = \epsilon$ and $pre(\alpha) = \{\mathcal{C}_{i_j}\}$, i.e. there will be $N(\mathcal{C}_{i_j}, t)$ instances of the local derivative each proceeding with the individual activity concurrently. When $pre(\alpha) \neq \{\mathcal{C}_{i_j}\}$ the activity is a shared activity involving local derivatives from two or more component types in a multi-way synchronisation. By the definition of apparent rate in PEPA + Π, if there are N replicated instances of a component enabling a transition (α, r), the apparent rate of the activity will be $N \times r$. By the semantics, the apparent rate of a synchronised activity is either the minimum of the apparent rates of the cooperating components or their product. The second term is explained similarly, noting that the rate of an entry activity will be determined by the number of components for which this is an *exit* activity, in accordance with the semantics of the language.

//Form the preceding set $pre(j)$ *for all activities*
For $j = 1....N_A$
  $pre(j) = \emptyset$
  For $k = 1...N_D$
    If $M_a(k, j) = -1$
      $pre(j) = pre(j) \cup \{k\}$
//Form one ODE for each local derivative/state variable
For $i = 1...N_D$
  //Record the impact of each activity involving this derivative
  For $j = 1....N_A$
    If $M_a(i, j) = +1$
      Add $+r_j \times \underset{k \in pre(j)}{coop}(j)(n_k(t))$ to the equation
    If $M_a(i, j) = -1$
      Add $-r_j \times \underset{k \in pre(j)}{coop}(j)(n_k(t))$ to the equation

Fig. 2. Pseudo-code for generating the set of ODEs.

Dividing by δt and taking the limit, $\delta t \longrightarrow 0$, we obtain:

$$\frac{\mathrm{d}N(C_{i_j}, t)}{\mathrm{d}t} = -\sum_{\alpha \in Ex(C_{i_j})} rate(\alpha) \times \underset{C_{k_l} \in pre(\alpha)}{coop}(\alpha)(N(C_{k_l}, t)) + \sum_{\alpha \in En(C_{i_j})} rate(\alpha) \times \underset{C_{k_l} \in pre(\alpha)}{coop}(\alpha)(N(C_{k_l}, t)).$$

In the following subsection we show how these equations may be derived automatically from the model definition in a straightforward way. To fully specify the system of ODEs it only remains to set the initial values of the state variables, i.e. $N(C_{i_j}, 0)$ for $i_j = 1, \ldots, N$. These are easily recorded from the initial model configuration.

4.2. Automatically deriving ODEs

The impact of activities on derivatives can be recorded in either a graph or a matrix form, easily derived from the syntactic presentation of the model, as defined below.

Definition 4.2 (*Activity Graph*). An *activity graph* is a bipartite graph $(\mathcal{N}, \mathcal{A})$. The nodes $\mathcal{N}$ are partitioned into $\mathcal{N}_A$, the *activities*, and $\mathcal{N}_D$, the *derivatives*. $\mathcal{A} \subset (\mathcal{N}_A \times \mathcal{N}_D) \cup (\mathcal{N}_D \times \mathcal{N}_A)$, where $a = (n_A, n_D) \in \mathcal{A}$ if n_A is an exit activity of derivative n_d, and $a = (n_D, n_A) \in \mathcal{A}$ if n_A is an entry activity of derivative n_D.

The same information can be represented in a matrix, termed the *activity matrix*.

Definition 4.3 (*Activity Matrix*). For a model with N_A activities and N_D distinct local derivatives, the *activity matrix* M_a is an $N_D \times N_A$ matrix, and the entries are defined as follows.

$$(d_i, a_j) = \begin{cases} +1 & \text{if } a_j \text{ is an entry activity of } d_i \\ -1 & \text{if } a_j \text{ is an exit activity of } d_i \\ 0 & \text{otherwise.} \end{cases}$$

In the activity matrix each row corresponds to a single local derivative. In the representation of the model as a system of ODEs there is one equation for each state variable, i.e. for the current number of each local derivative exhibited. This equation details the impact of the rest of the system on the value of that state variable. This can be derived automatically from the activity matrix when we associate a state variable n_i with each row of the matrix and the rate constant $rate(a_j)$ with the column of the matrix corresponding to a_j. The number of terms in the ODE will be equal to the number of non-zero entries in the corresponding row, each term being based on the rate of the activity associated with that column. As explained above, by the semantics of PEPA, the actual rate of change caused by each activity will be the rate multiplied by the minimum or the product of the current number of local derivatives enabling that activity in parallel, for each cooperating component type. The identity of these derivatives can be found in the column corresponding to the activity, a negative entry indicating that this derivative participates in that activity. There will be one ODE in the system for each row of the matrix. Fig. 2 shows the algorithm for generating a set of ODEs with respect to a PEPA model presented in pseudo-formal language.

4.3. Example

Let us now apply this algorithm to generate the set of ODEs of the example pathway given in Fig. 1. Firstly, the activity matrix has to be generated according to Definition 4.3:

M_a	$k1react$	$k2react$	$k3react$	$k4react$	$k5react$
A	-1	1	0	1	0
A/X	1	-1	-1	0	0
B	0	0	1	-1	0
X	-1	1	0	0	1
X/A	1	-1	-1	0	0
Y	0	0	1	0	-1

Then for any activity the set of its preceding local derivatives and the type of the cooperation are defined:

	$k1react$	$k2react$	$k3react$	$k4react$	$k5react$
pre	$\{A, X\}$	$\{A/X, X/A\}$	$\{A/X, X/A\}$	$\{B\}$	$\{Y\}$
$coop$	Π	min	min	ϵ	ϵ

Finally, the algorithm is applied to generate the following set of ODEs. The level of concentration of the local derivative P at time t is denoted $P(t)$:

$$\frac{dA(t)}{dt} = -l_1 \times A(t) \times X(t) + l_2 \times \min(A/X(t), X/A(t)) + l_4 \times B(t)$$
$$\frac{dA/X(t)}{dt} = l_1 \times A(t) \times B(t) - l_2 \times \min(A/X(t), X/A(t)) - l_3 \times \min(A/X(t), A/X(t))$$
$$\frac{dB(t)}{dt} = l_3 \times \min(A/X(t), X/A(t)) - l_4 \times B(t)$$
$$\frac{dX(t)}{dt} = -l_1 \times A(t) \times X(t) + l_2 \times \min(A/X(t), X/A(t)) + l_5 \times Y(t)$$
$$\frac{dX/A(t)}{dt} = l_1 \times A(t) \times X(t) - l_2 \times \min(A/X(t), X/A(t)) - l_3 \times \min(A/X(t), X/A(t))$$
$$\frac{dY(t)}{dt} = l_3 \times \min(A/X(t), X/A(t)) - l_5 \times Y(t).$$

One can notice that the expressions defining $\frac{dA/X(t)}{dt}$ and $\frac{dX/A(t)}{dt}$ are the same. Furthermore since A/X and X/A refer to the same species their initial number of copies is equal as well. So $A/X(t) = X/A(t)$ for all t and $\min(A/X(t), X/A(t))$ can be substituted by $A/X(t)$:

$$\frac{dA(t)}{dt} = -l_1 \times A(t) \times X(t) + l_2 \times A/X(t) + l_4 \times B(t)$$
$$\frac{dA/X(t)}{dt} = l_1 \times A(t) \times B(t) - l_2 \times A/X(t) - l_3 \times A/X(t)$$
$$\frac{dB(t)}{dt} = l_3 \times A/X(t) - l_4 \times B(t)$$
$$\frac{dX(t)}{dt} = -l_1 \times A(t) \times X(t) + l_2 \times A/X(t) + l_5 \times Y(t)$$
$$\frac{dX/A(t)}{dt} = l_1 \times A(t) \times X(t) - l_2 \times A/X(t) - l_3 \times A/X(t)$$
$$\frac{dY(t)}{dt} = l_3 \times A/X(t) - l_5 \times Y(t).$$

5. Applying Kurtz's theorem

In this section we show how to apply Kurtz's Theorem [17] in order to demonstrate the relationship between the discretised CTMC representation of PEPA models and the ODE models which can be derived from the same system description.

5.1. Kurtz's theorem

Kurtz's theorem states that, under certain assumptions, the solutions provided by a set of ODEs can be regarded as the limit of a sequence of "pure jump" Markov processes. As a special case of this general result, Kurtz shows how to obtain the

ODEs as the limit of a sequence of *density dependent* CTMCs, which model discrete numbers of elements in their different states [17]. The density dependent condition means that the rates of the CTMCs may depend on a scaled representation of states. For instance, when states represent number of individuals and are normalised with respect to volume or area, then the rates depend on population densities. Instead in the case of the Markov chains derived from PEPA models we are interested in studying their behaviour when the number of levels increases, or equivalently, when the granularity decreases. Therefore the rates do not contain information on area or volume but on the granularity H, the maximum concentration met by any reagent during the lifetime of the system divided by the number of levels of concentration N.

Definition 5.1 (*Density Dependent Markov Chains*). A family of CTMCs is called *density dependent* if and only if there exists a continuous function $f(x, l),\ x \in \mathbb{R}^h,\ l \in \mathbb{Z}^h$, such that the infinitesimal generators of X_H are given by:

$$q_{k,k+l} = H^{-1} f\,(Hk, l)\,, \quad l \neq 0$$

with $q_{k,k+l}$ denoting an entry of the infinitesimal generator of X_H, k a numerical state vector and l a *transition vector* that contains the modifications for each state of each species (i.e. the number of copies to add or subtract) when the transition is taken.

In [17] Kurtz shows that the ODE system $\frac{\mathrm{d}X(t)}{\mathrm{d}t} = F(X)$ defined by:

$$F(x) = \sum_l l f\,(x, l)$$

is the solution of the limit of X_H when H tends to 0, in the sense that:

$$\lim_{H\to 0} HX_H(0) = X(0) \implies \forall \delta > 0 \lim_{H\to 0} \mathbb{P}\left(\sup_{s\leq t} |HX_H(s) - X(s)| > \delta\right) = 0.$$

The limit expresses that the probability for X_H to take a trajectory different from X tends to 0 when H tends to 0. The result is based on the assumption that the following conditions are met:

There exists an open set $E \subset \mathbb{R}^h$ such that $X(t) \in E$ and

$$\exists M, \forall x, y \in E \quad |F(x) - F(y)| < M\,|x - y| \tag{2}$$

$$\sup_{x\in E} \sum_l |l| f(x, l) < \infty \tag{3}$$

$$\lim_{d\to\infty} \sup_{x\in E} \sum_{|l|>d} |l| f(x, l) = 0. \tag{4}$$

These conditions can be understood as follows:

(2) This says that the function F is Lipschitz continuous, imposing a certain degree of *smoothness* on the function;

(3) This imposes that for each transition the rate of change is bounded;

(4) This ensures that there is a bound for the whole state space which means that the impact of transitions remains bounded.

It is important to note that Kurtz's result does not tell us about the relationship between the Markov chain with granularity H and the system of ODEs. However, it does tell us that in the limit, as H tends to 0, the agreement between the Markov chain and the system of ODEs is complete, in the sense that the behaviour of the two with respect to the state variables will be identical. We can regard this as saying that for density dependent Markov chains the stochasticity is such that when there are large numbers of entities the variability balances in such a way that the process tends to a deterministic limit.

In the context of PEPA models of biochemical signalling pathways we wish to establish that Kurtz's theorem holds. Therefore we seek to prove that the CTMCs generated from PEPA models are density dependent. Furthermore, we have to show how the deterministic distribution obtained in the limit is related to the solution of the system of ODEs derived from the corresponding PEPA models.

5.2. Mass action kinetics and density dependency

In the original definition of PEPA the rates do not depend on some density. However in the context of biochemical reactions they do. It is important to notice that the rate of a reaction does not depend on the actual number of copies (as the apparent rate seems to model) but on their concentration. This is why the number of copies are said to model levels of concentration rather than number of individuals. The idea underlying this demonstration is to study the limit of CTMCs when the levels of concentration are finer but as this number grows the actual transition rate of the PEPA model has to be rescaled appropriately.

Let W, N and H be, respectively, the concentration, the number of levels and the granularity for a system previously defined. W corresponds to the smallest upper bound concentration that any reagent can reach during the lifetime of the system. The relation among these measures is given by the following formula $H = \frac{W}{N}$.

N is an integer denoting the number of levels of concentration and does not have any physical dimension. Therefore H is a concentration corresponding to the width (in absolute value) between the actual concentrations associated with two consecutive levels.

We will show that the rate of change of a given species, once rescaled so that levels of concentration are considered instead concentrations, coincides with the apparent rate defined in Section 2.1.

(1) Let us first consider the rate of change of a species that does not involve interactions with other species. Changing the level of a molecule species M_i takes time Δt to increase or decrease the level of M_i concentration of H. Thus the rate of change is $\frac{1}{\Delta t}$. Consider the deterministic model where $M_i(t)$ is the continuous concentration of molecule species M at time t and α is the name of the reaction that changes the level of molecule species M_i.

$$\frac{\mathrm{d}M_i(t)}{\mathrm{d}t} = rate(\alpha)M_i(t). \tag{5}$$

For small Δt we have

$$M_i(t + \Delta t) = M_i(t) + rate(\alpha)M_i(t)\Delta t. \tag{6}$$

Therefore, if $M_i(t)$ and $M_i(t + \Delta t)$ correspond to two adjacent actual discrete concentrations (i.e. their difference is H), we have

$$\Delta t = \frac{H}{rate(\alpha)M_i(t)}. \tag{7}$$

And the rate of change is given by

$$\frac{1}{\Delta t} = \frac{rate(\alpha)M_i(t)}{H}. \tag{8}$$

Moreover the actual concentration $M_i(t)$ corresponds to the discrete concentration Hx_i where x_i is the level corresponding to molecule M_i.

$$\frac{1}{\Delta t} = rate(\alpha)x_i(t). \tag{9}$$

Thus in that case (species without interaction) the rate of change coincides with the definition of the apparent rate of PEPA + Π.

(2) Let us now consider the rate of change of a species that does involve interactions with another species. We assume that the molecule M_i is created as the result of the interaction of M_j and M_k, via the reaction α.

$$\frac{\mathrm{d}M_i(t)}{\mathrm{d}t} = rate(\alpha)M_j(t)M_k(t). \tag{10}$$

Reasoning as previously we can establish the rate of change of M_i.

$$\frac{1}{\Delta t} = \frac{rate(\alpha)M_j(t)M_k(t)}{H}. \tag{11}$$

Then we can substitute M_j and M_k by Hx_j and Hx_k respectively.

$$\frac{1}{\Delta t} = rate(\alpha)x_j(t)x_k(t)H. \tag{12}$$

Once again the rate of change coincides with the definition of the apparent rate of PEPA + Π.

The generalisation of the rate of change resulting from the cooperations of l kinds of species can easily be drawn in the same manner leading to the following definition of the apparent rate.

$$r_\alpha(P_1[x_1] \overset{\alpha}{\bowtie} \cdots \overset{\alpha}{\bowtie} P_l[x_l]) = H^{-1} \times rate(\alpha) \times \prod_{i=1}^{l} x_i H.$$

5.3. Definitions and application of Kurtz's theorem

Let $\mathcal{M}$ be a PEPA model with n component types $\mathcal{C}_i$, $i = 1, \ldots, n$, each with h_i distinct derivatives. Let $\mathcal{V}(\mathcal{M})$ be its numerical vector form. The total size of $\mathcal{V}(\mathcal{M})$ is $h = \sum_{i=1}^{h} h_i$. Let $x \in \mathbb{N}^h$ denote a numerical state vector of $\mathcal{M}$, that is an instance of $\mathcal{V}(\mathcal{M})$, and represent the number of levels. The entry x_{i_j} records how many instances of the jth local derivative of component type $\mathcal{C}_i$ are exhibited in the current state. Let $l \in \{-1, 0, 1\}^h$ be a vector that represents for each derivative that its number is decreased by 1, unchanged or increased by 1. l, called a *transition vector*, describes the transition between some state x and $x + l$.

To use Kurtz's theorem we need to choose the right rescaling factor such that the discrete Markov chain, once rescaled, approximates the deterministic ODE system. In our situation that rescaling factor is H, that is, if $X_H(t)$ specifies the Markov chain describing the evolution of the number of levels of concentration then $HX_H(t)$ specifies a Markov chain describing the evolution of concentrations that approximates $X(t)$, the deterministic ODE system.

Let Q be the infinitesimal generator of the parametrised Markov chain $X_H(t)$ derived from $\mathcal{M}$ using the numerical vector form as state space. We are interested in entries in Q that can be denoted $q_{x,x+l}$ (entries that cannot be denoted in this way are null).

So x is denoted:

$$x = (x_{1_1}, \ldots, x_{1_j}, \ldots, x_{i_j}, \ldots)$$

with x_{i_j} denoting the number of levels of the species i in the state j. To represent the actual concentration of the different species in their states we introduce the following vector:

$$y = (y_{1_1}, \ldots, y_{1_j}, \ldots, y_{i_j}, \ldots) \in \mathbb{R}_+^h$$

defined by:

$$y = Hx = (Hx_{1_1}, \ldots, Hx_{1_j}, \ldots, Hx_{i_j}, \ldots).$$

For any activity α there is exactly one transition vector, denoted l^α. The entry of the ith component and the jth local derivative of the transition vector l^α is denoted $l^\alpha_{i_j}$. We will now define both $q_{x,x+l^\alpha}$ and $f(x, l^\alpha)$ for all activities (denoted by the set Act) and for each case over $coop(\alpha)$, and show that X_H is density dependent, that is $q_{x,x+l^\alpha} = H^{-1} f(Hx, l^\alpha)$.

$$\forall \alpha \in Act, \quad \left(\forall i, j \quad l^\alpha_{i_j} = \mathbf{1}_{\{\alpha \in En(\mathcal{C}_{i_j})\}} - \mathbf{1}_{\{\alpha \in Ex(\mathcal{C}_{i_j})\}} \right)$$

- $coop(\alpha) = \epsilon$ then $\exists i, j$ such that $\alpha \in Ex\left(\mathcal{C}_{i_j}\right)$ and so $q_{x,x+l^\alpha} = rate(\alpha) \times x_{i_j}$.
  Let
  $$f(x, l^\alpha) = rate(\alpha) \times x_{i_j}$$
  then
  $$q_{x,x+l^\alpha} = rate(\alpha) \times x_{i_j} = H^{-1} rate(\alpha) \times Hx_{i_j} = H^{-1} \times f(Hx_{i_j}, l^\alpha)$$
- if $coop(\alpha) \neq \epsilon$ then
  - if $coop(\alpha) = \min$ then $q_{x,x+l^\alpha} = rate(\alpha) \times \min_{\mathcal{C}_{i_j} \in pre(\alpha)} (x_{i_j})$.
    Let
    $$f(x, l^\alpha) = rate(\alpha) \times \min_{\mathcal{C}_{i_j} \in pre(\alpha)} (x_{i_j})$$
    then
    $$\begin{aligned} q_{x,x+l^\alpha} &= \frac{H}{H} \times rate(\alpha) \times \min_{\mathcal{C}_{i_j} \in pre(\alpha)} (x_{i_j}) \\ &= H^{-1} \times rate(\alpha) \times \min_{\mathcal{C}_{i_j} \in pre(\alpha)} (Hx_{i_j}) \\ &= H^{-1} \times f(Hx, l^\alpha) \end{aligned}$$
  - if $coop(\alpha) = \Pi$ then $q_{x,x+l^\alpha} = H^{-1} \times rate(\alpha) \times \prod_{\mathcal{C}_{i_j} \in pre(\alpha)} Hx_{i_j}$.
    Let
    $$f(x, l^\alpha) = rate(\alpha) \prod_{\mathcal{C}_{i_j} \in pre(\alpha)} x_{i_j}$$
    then
    $$q_{x,x+l^\alpha} = H^{-1} \times f(Hx, l^\alpha).$$

We have shown that for all activities and all types of cooperation the family of CTMCs are density dependent.

Let us assume that the initial concentrations are described by the vector c where

$$c = \left(c_{1_1}, \ldots, c_{1_j}, \ldots, c_{i_j}, \ldots\right).$$

Thus the corresponding initial levels are given by the vector $x_0 = \lfloor \frac{c}{H} \rfloor$.

$\lim_{H \to 0} Hx_0 = c$, thus according to Kurtz's theorem and assuming that the conditions of its applicability are met (see below) $X(t)$ is the solution of the following differential equations $\frac{dX(t)}{dt} = F(X)$ defined by:

$$F(y) = \sum_l l f(y, l)$$

subject to initial conditions $X(0) = c$.

Finally, the three conditions (2), (3) and (4) must be verified.

The trajectory of $X(t)$ is closed, that is, included within a bounded set. Also E can be chosen to be bounded and the inequalities (2) and (3) are trivially verified. The equality (4) is trivially verified because $f(y, l)$ is non-null only when the entries of l have the values -1, 1 or 0. Consequently $f(y, l) = 0$ for all $|l| > C$, where C is a constant which can be chosen as $C = \sqrt{N}$ according to the Euclidean metric, denoting the worst case when all entries of l are 1 or -1.

In conclusion we have proved that the scaled Markov process $HX_H(t)$, that represents the vector of discrete concentrations at time t, converges in probability to the differential equations derived from the same PEPA model.

6. Example

In this section we consider again the example from Section 3 and show that the ODEs derived earlier are those which satisfy Kurtz's theorem. The kinetic reactions are recalled below:

$$A + X \underset{k2}{\overset{k1}{\rightleftharpoons}} A/X \xrightarrow{k3} B + Y$$
$$B \xrightarrow{k4} A$$
$$Y \xrightarrow{k5} X.$$

The five activities and their *coop* values are as follows:

	$k1react$	$k2react$	$k3react$	$k4react$	$k5react$
$coop$	Π	min	min	ϵ	ϵ

These give rise to five transition vectors, denoted $l^{k1react}$, $l^{k2react}$, $l^{k3react}$, $l^{k4react}$ and $l^{k5react}$ respectively. Then $f(x, l)$ is defined for each transition vector as follows:

(1) for $l^{k1react}$:

$$l^{k1react} = (-1, 1, 0, -1, 1, 0)$$
$$\begin{aligned} f(x, l^{k1react}) &= (k1 \times x_{1_1}) \times (1 \times x_{2_1}) \\ &= k1 \times x_{1_1} \times x_{2_1} \end{aligned}$$

(2) for $l^{k2react}$:

$$l^{k2react} = (1, -1, 0, 1, -1, 0)$$
$$\begin{aligned} f(x, l^{k2react}) &= \min(k2 \times x_{1_2}, k2 \times x_{2_2}) \\ &= k2 \times \min(x_{1_2}, x_{2_2}) \end{aligned}$$

(3) for $l^{k3react}$:

$$l^{k3react} = (0, -1, 1, 0, -1, 1)$$
$$\begin{aligned} f(x, l^{k3react}) &= \min(k3 \times x_{1_2}, k3 \times x_{2_2}) \\ &= k3 \times \min(x_{1_2}, x_{2_2}) \end{aligned}$$

(4) for $l^{k4react}$:

$$l^{k4react} = (1, 0, -1, 0, 0, 0)$$
$$f(x, l^{k4react}) = k4 \times x_{1_3}$$

(5) for $l^{k5react}$:

$$l^{k5react} = (0, 0, 0, 1, 0, -1)$$
$$f(x, l^{k5react}) = k5 \times x_{2_3}.$$

It is now possible to express $F(x)$:

$$F(x) = \sum_l f(x, l)$$
$$\begin{aligned} F(x) \;=\; & k1 \times x_{1_1} \times x_{2_1} \times (-1, 1, 0, -1, 1, 0)^T + k2 \times \min(x_{1_2}, x_{2_2}) \times (1, -1, 0, 1, -1, 0)^T \\ & + k3 \times \min(x_{1_2}, x_{2_2}) \times (0, -1, 1, 0, -1, 1)^T + k4 \times x_{1_3} \times (1, 0, -1, 0, 0, 0)^T + k5 \times x_{2_3} \times (0, 0, 0, 1, 0, -1)^T. \end{aligned}$$

After adding all terms we obtain:

$$F(x) = \begin{pmatrix} -k1x_{1_1}x_{2_1} + k2\min(x_{1_2}, x_{2_2}) + k4x_{1_3} \\ k1x_{1_1}x_{2_1} - k2\min(x_{1_2}, x_{2_2}) - k3\min(x_{1_2}, x_{2_2}) \\ k3\min(x_{1_2}, x_{2_2}) - k4x_{1_3} \\ -k1x_{1_1}x_{2_1} + k2\min(x_{1_2}, x_{2_2}) + k5x_{2_3} \\ k1x_{1_1}x_{2_1} - k2\min(x_{1_2}, x_{2_2}) - k3\min(x_{1_2}, x_{2_2}) \\ k3\min(x_{1_2}, x_{2_2}) - k5x_{2_3} \end{pmatrix}.$$

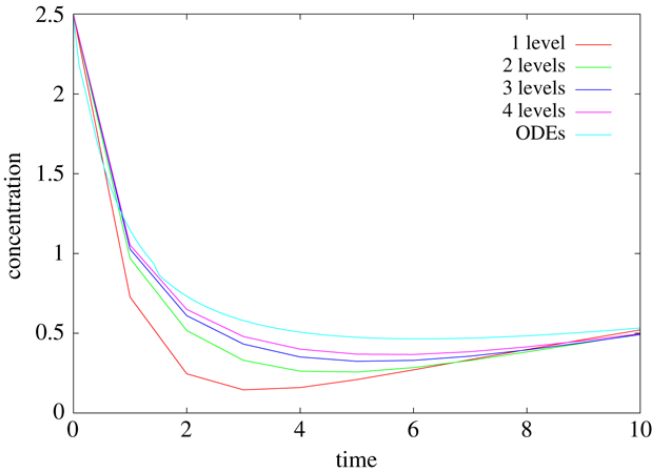

Fig. 3. Convergence of the transient probabilities in the CTMCs with levels to the ODE solution for the concentration of X for the simple model presented in this paper.

We note that the second and fifth rows (corresponding to the components C_{1_2} and C_{2_2} respectively) are exactly the same. These correspond to the two descriptions of the same complex A/X. Thus we can assume that they will be initialised with the same concentration and therefore it follows that $\min(x_{1_2}, x_{2_2}) = x_{1_2} = x_{2_2}$.

Therefore, after renaming appropriately x_{1_1} by A, x_{1_2} by A/X, x_{1_3} by B, x_{2_1} by X, x_{2_2} by X/A, x_{2_3} by Y, and eliminating the min operators we obtain the following set of ODEs:

$$\frac{\mathrm{d}A(t)}{\mathrm{d}t} = -k1 \times A(t) \times X(t) + k2 \times A/X(t) + k4 \times B(t)$$
$$\frac{\mathrm{d}A/X(t)}{\mathrm{d}t} = k1 \times A(t) \times X(t) - k2 \times A/X(t) - k3 \times A/X(t)$$
$$\frac{\mathrm{d}B(t)}{\mathrm{d}t} = k3 \times A/X(t) - k4 \times B(t)$$
$$\frac{\mathrm{d}X(t)}{\mathrm{d}t} = -k1 \times A(t) \times X(t) + k2 \times A/X(t) + k5 \times Y(t)$$
$$\frac{\mathrm{d}X/A(t)}{\mathrm{d}t} = k1 \times A(t) \times X(t) - k2 \times A/X(t) - k3 \times A/X(t)$$
$$\frac{\mathrm{d}Y(t)}{\mathrm{d}t} = k3 \times A/X(t) - k5 \times Y(t).$$

Note that this is identical to the set of ODEs generated by the algorithm given in Fig. 2 from the same example.

In Fig. 3 we show the transient trajectory of the first four CTMCs corresponding to the simple example, and the ODE solution. The convergence of the sequence of CTMCs towards the deterministic model as the number of levels increases can be plainly seen. Furthermore this simple model reaches the limit distribution very fast: the transient solutions of Markov chains with more than four levels are indistinguishable from the solutions of ODEs.

7. Conclusions and future work

We have established the relationship between the discretised, CTMC representation of PEPA models of biochemical signalling pathways, and the ODE models which can be derived from the same system description. A variety of analysis techniques are available for the CTMC model which are not possible on the ODEs. For example, standard process algebra analysis techniques can be used to establish that a model is deadlock free before expensive numerical experimentation is started. Furthermore more sophisticated model checking techniques can be applied to prove properties about the model. For instance, in [13] the authors apply Computational Stochastic Logic model checking on a CTMC model to assess the probability that one binding site is used before another.

It should be noted that our result establishes the relationship between the set of ODEs and the limit of the sequence of CTMCs, without any indication of how many levels of granularity are necessary in order to get good agreement. Of course, finer granularity means that there will be more states in the CTMC and the alternative analysis techniques may become prohibitively expensive. Thus we are faced with a trade-off between accuracy and tractability. However, empirical evidence is that, for many systems, a relatively low number of levels (steps along the sequence of CTMCs) is sufficient for the CTMC and ODEs to exhibit the same behaviour. For example, in [6], the authors show that results become indistinguishable with just seven levels.

An important area for future work is to investigate the relationship between the level of granularity in the discretisation and the accuracy achieved with respect to the ODEs. In particular, from a practical point of view it would be useful to establish a characterisation of the accuracy of the analysis with a given level of granularity.

Currently, the syntax of PEPA does not allow higher order reactions and elements with stoichiometric coefficient greater than one to be represented. As a result we are working on a new stochastic process algebra, more closely tailored to the needs of modelling biochemical signalling pathways. This is an area of on-going work and it is hoped to extend the results in this paper to this new formalism when it is established.

Acknowledgements

Nil Geisweiller was supported by a visiting fellowship from the School of Informatics at the University of Edinburgh, during part of this work. Jane Hillston is supported by the EPSRC Advanced Research Fellowship *Process Algebra for Collective Dynamics* (EP/c543696/01a) and EU FET-IST Global Computing 2 project, SENSORIA (IST-3-016004-IP-09). Marco Stenico is supported by EPSRC studentship (EP/c54370x/01).

References

[1] L. Bortolussi, A. Policriti, Stochastic concurrent constraint programming and differential equations, in: Proceedings of Fifth Workshop on Quantitative Aspects of Programming Languages, Braga, Portugal, 2007.
[2] M. Calder, A. Duguid, S. Gilmore, J. Hillston, Stronger computational modelling of signalling pathways using both continuous and discrete-state methods, in: Proceedings of 4th International Workshop on Computational Methods in Systems Biology, Trento, Italy, 18–19th October 2006, pp. 63–77.
[3] M. Calder, S. Gilmore, J. Hillston, Modelling the influence of RKIP on the ERK signaling pathway using the stochastic process algebra PEPA, BioConcur, 2004.
[4] M. Calder, S. Gilmore, J. Hillston, Automatically deriving ODEs from process algebra models of signalling pathways, in: Computational Methods in Systems Biology, 2005, pp. 204–215.
[5] M. Calder, J. Hillston, What do scaffold proteins really do? in: Proceedings of the 5th International Workshop on Process Algebras and Stochastically Timed Activities, PASTA'06, Imperial College, London, UK, June 2006, pp. 96–101.
[6] M. Calder, V. Vyshemirsky, D. Gilbert, R. Orton, Analysis of signalling pathways using continuous time Markov chains, Transactions on Computational Systems Biology VI (2006) 44–67.
[7] L. Cardelli, From processes to ODEs by chemistry, Unpublished manuscript, 2006.
[8] D. Chiarugi, M. Curti, P. Degano, R. Marangoni, VICE: A VIrtual CEll, in: Proceedings of the 2nd International Workshop on Computational Methods in Systems Biology, Springer, Paris, France, Apr 2004.
[9] N. Geisweiller, An attempt to give a clear semantics of the extension of PEPA for massively parallel processes and biological modelling, in: Process Algebra and Stochastically Timed Activities, 2006, pp. 36–43.
[10] D. Gillespie, Exact stochastic simulation of coupled chemical reactions, Journal of Physical Chemistry 81 (25) (1977) 2340–2361.
[11] D. Gillespie, L. Petzold, Numerical simulation for biochemical kinetics, 2006.
[12] P. Goss, J. Peccoud, Quantitative modeling of stochastic systems in molecular biology by using stochastic Petri nets, Proceedings of National Academy of Science, USA 95 (12) (1998) 7650–6755.
[13] J. Heath, M. Kwiatkowska, G. Norman, D. Parker, O. Tymchynshyn, Computer assisted biological reasoning: The simulation and analysis of FGF signalling pathway dynamics, in: Proceedings of the Winter Simulation Conference, 2006.
[14] J. Hillston, The nature of synchronisation, in: U. Herzog, M. Rettelbach (Eds.), Proceedings of the Second International Workshop on Process Algebras and Performance Modelling, Erlangen, Nov. 1994, pp. 51–70.
[15] J. Hillston, A Compositional Approach to Performance Modelling, Cambridge University Press, 1996.
[16] J. Hillston, Fluid flow approximation of PEPA models, in: Proceedings of the 2nd International Conference on Quantitative Evaluation of Systems, QEST, IEEE Computer Society Press, Torino, Italy, September 2005, pp. 33–42.
[17] T. Kurtz, Solutions of ordinary differential equations as limits of pure jump markov processes, J. Appl. Prob. 7 (1970) 49–58.
[18] T. Kurtz, The relationship between stochastic and deterministic models for chemical reactions, Journal of Chemical Physics 57 (7) (1972) 2976–2978.
[19] C. Kuttler, J. Niehren, Gene regulation in the pi calculus: simulating cooperativity at the lambda switch, in: Transactions on Computational Systems Biology, in: LNCS, vol. 4230, 2006, pp. 24–55.
[20] P. Lecca, C. Priami, C. Laudanna, G. Constantin, A Biospi model of lymphocyte-endothelial interactions in inflamed brain venules, in: Pacific Symposium of Biocomputing, 2004, pp. 521–532.
[21] C. Priami, A. Regev, W. Silverman, E. Shapiro, Application of a stochastic name passing process calculus to representation and simulation of molecular processes, Information Processing Letters 80 (2001) 25–31.
[22] A. Regev, Computational systems biology: A calculus for biomolecular knowledge, 2002.