INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

Exact simulation of hybrid stochastic and deterministic models for biochemical systems

Aurélien Alfonsi — Eric Cancès — Gabriel Turinici — Barbara Di Ventura — Wilhelm Huisinga

N° 5435

December 2004

Thèmes NUM et BIO

Rapport de recherche

N 0249-6399 ISRN INRIA/RR--5435--FR+ENG

Exact simulation of hybrid stochastic and deterministic models for biochemical systems

Aurélien Alfonsi*, Eric Cancès †, Gabriel Turinici‡, Barbara Di Ventura§, Wilhelm Huisinga¶

Thèmes NUM et BIO — Systèmes numériques et Systèmes biologiques
Projet MicMac

Rapport de recherche n° 5435 — December 2004 — 20 pages

Abstract:

Motivation: Over the last years it has become evident that stochastic effects play an important role in biological processes leading to an increase in stochastic modelling attempts. Despite the availability of exact algorithms to numerically solve the chemical master equation that entirely describes a stochastic system, stochastic simulations are most of the times very computationally expensive. Hybrid methods that treat some processes as in the deterministic framework and others as stochastic are a promising way to speed up simulations for those cases involving different time scales, e.g., systems integrating metabolic pathways and gene regulatory networks.

Results: We present a sound mathematical basis for hybrid stochastic and deterministic models, and provide easy to handle algorithmic schemes for exact simulation of such hybrid models. We finally show numerical results obtained for a model describing bacteriophage T7 intracellular growth that illustrate the power of and the computational speed gained by the presented hybrid modelling approach.

Availability: An implementation of the hybrid algorithm for the T7 model system is available in Fortan 77 or C++ on request from the authors.

Key-words: molecular biology, stochastic deterministic coupling, bacteriophage T7

* CERMICS, ENPC,6-8 Avenue Blaise Pascal, Cité Descartes - Champs sur Marne, 77455 Marne la Vallée Cedex 2, France

† CERMICS, ENPC,6-8 Avenue Blaise Pascal, Cité Descartes - Champs sur Marne, 77455 Marne la Vallée Cedex 2, France, INRIA Rocquencourt, Domaine de Voluceau, 78153 Le Chesnay Cedex, France

‡ CERMICS, ENPC,6-8 Avenue Blaise Pascal, Cité Descartes - Champs sur Marne, 77455 Marne la Vallée Cedex 2, France, INRIA Rocquencourt, Domaine de Voluceau, 78153 Le Chesnay Cedex, France

§ Structural Biology and Biocomputing program, EMBL, Meyerhofstrasse 1, 69117 Heidelberg

¶ Free University Berlin, Department of Mathematics and Computer Science, Arnimallee 2-6, D-14195 Berlin/Germany, and DFG Research Center, Berlin

Simulation exacte hybride stochastique-détérministe pour systèmes biochimiques

Résumé : La reconnaissance de l'importance des effets stochastiques en biologie a donné lieu à un fort développement des modèles incorporant de l'aléa. Malgré l'existence d'algorithmes permettant de résoudre exactement l'équation maîtresse d'un système de réactions biochimiques, les simulations stochastiques restent très coûteuses en termes de temps de calcul. Les méthodes hybrides consistant à considérer certains processus comme déterministes et d'autres comme stochastiques, constituent une approche très prometteuse pour accélérer les simulations de modèles faisant intervenir différentes échelles de temps, comme par exemple ceux décrivant des réseaux de régulation de gènes.

Nous établissons dans cet article une formulation mathématique rigoureuse de ces modèles hybrides et nous proposons des algorithmes numériques permettant de simuler ces modèles. Nous illustrons l'efficacité de ces concepts par une application au cas du système bactériophage T7.

Mots-clés : biologie moléculaire, couplage stochastique déterministe, bactériophage T7

1 Introduction

Stochastic models have gained considerable attention when experiments conducted at the level of single cells showed the existence of a non-negligible level of noise in intracellular processes, like transcription and translation [7, 22, 23]. Although in most cases regulatory circuits use feedback loops, redundancy and other mechanisms to be robust against inherent fluctuations and give a deterministic outcome nonetheless [2, 24, 30, 34], noise is, for instance, used as source of phenotypic variability, which is favorable in evolutionary terms [21, 27]. In the past few years, therefore, a great number of stochastic models have appeared to correctly deal with extremely low number of molecules and large fluctuations in reaction kinetics [1, 15, 29, 31]. The dynamics of a stochastic system is described by the chemical master equation which only seldom posses analytical solutions. Fortunately, Gillespie, back in 1976, devised two exact algorithms to numerically simulate the stochastic time evolution of coupled chemical reactions, which are equivalent to solving the chemical master equation [10, 11, 33]. Although a more efficient exact method has been proposed by Gibson and Bruck, based on reuse of random numbers and intelligent data structures [9], purely stochastic simulation are very computationally expensive, when there are many reactions, some molecular species are present in relative large amounts and/or reaction rates are high. Only recently, modifications to the original chemical master equation have been proposed to further speed up simulations. These involve either averaging over fast reactions [18], application of quasi-steady-state theory [26], or grouping together reactions that occur in fast succession (the so-called tau leaping methods) [3, 4, 14, 28]. Another strategy is to model those processes that either involve large number of particles or have fast rates, in a deterministic way, keeping stochastic the remaining ones. Very recently, two algorithms to simulate biochemical systems in such hybrid framework have been proposed [20, 32]. They are based on a prediction correction type heuristics for the realization of the stochastic part. In both cases, the main idea is to first predict the time in which a stochastic event should occur and then evolve the system of ordinary differential equations. At specific instances in time, the system is updated, and it is checked whether the stochastic event has to be performed or not. The algorithm, then, proceeds recursively. Although these algorithms seem to give good results, provided the time-step of the ODE solver is kept small, they lack a mathematical justification, on which basis their accuracy could be judged. In the following, we propose a rigorous mathematical ground for hybrid stochastic and deterministic modelling in a very elegant and natural way. We also present three algorithms for its implementation and finally show first intriguing numerical results for the bacteriophage T7 model system.

2 Purely stochastic or deterministic models

Consider N chemical species $S_1, \ldots, S_N$ involved in M reactions $R_1, \ldots, R_M$. Chemical species are modelled in terms of number of molecules $X(t) = (X_1(t), \ldots, X_N(t))$. The reaction rate for each reaction R_j is specified by a so-called propensity function $a_j = a_j(X(t), t)$, which is equal to the product of a rate constant c_j and the number of possible combinations of reactant molecules involved in reaction R_j. For the most frequently used reaction types, $S_a \to *$, $S_a + S_b \to *$ and $S_a + S_a \to *$, we get $a_j = c_j X_a(t)$, $a_j = c_j X_a(t) X_b(t)$ and $a_j = c_j X_a(t)(X_a(t)-1)/2$, respectively [10, 11]. Once a reaction R_j is performed, the number of molecules for each species is updated according to the state change vector ν_j, i.e., $X(t) \leftarrow X(t) + \nu_j$.

Deterministic model. Based on the law of mass action, a system of coupled ordinary differential equations (ODEs) is established for the time evolution of the *number* of molecules $X(t) \in \mathbb{R}_+^N$ (see, e.g.,[5, 19])

$$\frac{\mathrm{d}}{\mathrm{d}t} X(t) \quad = \quad \sum_{j=1}^{M} \nu_j a_j(X(t), t) \tag{1}$$

with some initial value $X(t_0) \in \mathbb{R}_+^N$. While the system should be described as a vector of integers, this model needs real values for $X(t)$. This is however acceptable under the assumption of large number of molecules ($X_i(t) \gg 1$) so that the relative error can be neglected.

Stochastic model. Based on physical laws and the idea that chemical reactions are essentially random processes, the stochastic formulation of chemical reactions is given in terms of a Markov jump process $X(t) \in \mathbb{N}^N$ [13, 33]. Its characterization is based on the probability $a_j(X(t), t)\mathrm{d}t$ of reaction R_j occurring in the next infinitesimal time interval $[t, t+\mathrm{d}t]$. Denoting by $T_j(t)$ the time at which reaction R_j first occurs after t, this amounts to write that

$$\mathbf{P}[T_j(t) \in [t, t+\mathrm{d}t] \,|X(t)] = a_j(X(t), t)\,\mathrm{d}t.$$

It is usually assumed that reactions are locally independent implying that $\mathbf{P}[\{T_j(t), T_k(t)\} \in [t, t+\mathrm{d}t] \,|X(t)] = a_j(X(t), t) a_k(X(t), t)\,(\mathrm{d}t)^2$.

In order to establish an evolution equation for $X(t)$ as well as justifying hybrid stochastic and deterministic models, the following time transformation

$$g_j(s|t) \quad = \quad \int_t^s a_j(X(\tau), \tau)\,\mathrm{d}\tau$$

is of key importance. The function $s \mapsto g_j(s|t)$ is non-decreasing for $s > t$, since the propensities a_j are non-negative by definition. Denote by $\mathrm{Exp}(1)$ the exponential random variable of parameter 1, i.e., $\xi \sim \mathrm{Exp}(1)$ if $\mathbf{P}[\xi \in [x, x+\mathrm{d}x]] = e^{-x}\mathrm{d}x$ for all $x \geq 0$ (here $\sim$ denotes equality in law between two random variables); ξ has survival probability $\mathbf{P}[\xi \geq x] = e^{-x}$. Now, let us consider a sequence (ξ_{jk}) of independent random variables $\xi_{jk} \sim \mathrm{Exp}(1)$, with $j = 1, \ldots, M$ and $k \in \mathbb{N}$. Define the random variables $S_j(n) = \sum_{k=1}^{n} \xi_{jk}$ for $j = 1, \ldots, M$ and

$$N_j(t) = \sum_{n=1}^{\infty} \mathbf{1}_{\{S_j(n) \leq g_j(t|t_0)\}}. \tag{2}$$

Then, it can be easily show that for all $j \in \{1, .., M\}$

$$\mathbf{P}[N_j(t + \mathrm{d}t) - N_j(t) = 1 | X(t)] = a_j(X(t), t)\mathrm{d}t$$

so that the next jump-time of the processes $N_j(t)$ have the same law than $T_j(t)$. Therefore, we get the important relation

$$T_j(t) \quad \sim \quad g_j^{-1}(\mathrm{Exp}(1)|t), \tag{3}$$

and the law of the evolution equation of the number of molecules is given by:

$$\mathrm{d}X(t) \quad = \quad \sum_{j=1}^{M} \nu_j \, \mathrm{d}N_j(t). \tag{4}$$

Note that $N_j(t)$ counts the number of times that reaction R_j occurred from the initial time t_0 up to time t. Evolution equation (4) corresponds to the infinitesimal generator

$$\begin{aligned} \mathcal{A}f(x) &= \lim_{s \to t+} \frac{\mathrm{d}}{\mathrm{d}s} \mathbf{E}[f(X(s)) \,| X(t) = x] \\ &= \sum_{j=1}^{M} f(x + \nu_j) a_j(x, t) - a_j(x, t) f(x). \end{aligned}$$

The dual point of view (Chapman-Kolmogorov) leads to the well-known chemical master equation [12, 33]

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}t} \mathbf{P}[X(t) = x] &= \sum_{j=1}^{M} \Big(a_j(x - \nu_j, t) \mathbf{P}[X(t) - \nu_j = x] - \\ &\qquad a_j(x, t) \mathbf{P}[X(t) = x] \Big). \end{aligned}$$

holding for all $x \in \mathbb{N}^N$.

3 Hybrid stochastic and deterministic models

Fully stochastic simulations become quite slow when many molecules and fast reactions are involved, due to the fact that the effort is proportional to the number of reactions performed. However, for each time-changed Poisson process $N_j(t)$ we have

$$\begin{aligned} \mathbf{E}[N_j(t)] &= \mathbf{E}[(N_j(t) - g_j(t|t_0))^2] \\ &= \int_0^t \mathbf{E}[a_j(X(s), s)] \mathrm{d}s \end{aligned}$$

for $t > t_0$. Since the relative fluctuation between $N_j(t)$ and $g_j(t|t_0)$ is given by

$$\frac{\mathbf{E}[(N_j(t) - g_j(t|t_0))^2]^{1/2}}{\mathbf{E}[N_j(t)]} = \frac{1}{\mathbf{E}[N_j(t)]^{1/2}},$$

it is reasonable to neglect it and approximate the stochastic dynamics by its continuous counterpart $g_j(t|t_0)$ (as in the deterministic model) when the propensity a_j is large and the numbers of molecules involved in the reaction are not too small. This motivates mixed stochastic and deterministic models.

Consider a partition of the reactions $R_1, \ldots R_M$ into those modelled stochastically (with index $j \in \mathcal{S}$) and those modelled deterministically (with index $j \in \mathcal{D}$). For a given model, there are at least three options for getting such a partition: (i) run a fully stochastic realization and analyze the frequencies/propensities of each reaction (note that the major computational cost is in computing many realizations); (ii) use biological insight. It seems reasonable, for instance, to model gene regulatory parts stochastically, while metabolic reactions deterministically; (iii) for each reaction choose adaptively between the two approaches using a criterion based on the number of molecules and its propensity function (see Section 6). The evolution equation for $X(t) \in \mathbb{R}^N$ is now given by the hybrid system

$$\mathrm{d}X(t) = \sum_{j \in \mathcal{D}} \nu_j \, a_j(X(t), t) \, \mathrm{d}t + \sum_{j \in \mathcal{S}} \nu_j \, \mathrm{d}N_j(t) \tag{5}$$

with initial value $X(t_0) \in \mathbb{R}_+^N$. Due to the partition of the reactions, a species can belong to the stochastic as well as the deterministic part of the hybrid system. As a consequence, it may theoretically happen that some $X_j(t)$ becomes negative. We note first that this situation never occurred in our simulations, and it is rather due to an unsuitable modelling choice (the deterministic equations should not act on small quantities). Secondly, this situation can be solved by an adaptive state-dependent choice of the reactions to be modelled stochastically or deterministically. Such an approach is currently under study.

The hybrid system (5) corresponds to the infinitesimal generator

$$\begin{aligned} \mathcal{A}f(x) &= \lim_{s \to t+} \frac{\mathrm{d}}{\mathrm{d}s} \mathbf{E}[f(X(s)) | X(t) = x] \\ &= \sum_{j \in \mathcal{D}} a_j(x,t) \frac{\mathrm{d}}{\mathrm{d}x} f(x) + \\ &\quad \sum_{j \in \mathcal{S}} f(x + \nu_j) a_j(x,t) - a_j(x,t) f(x). \end{aligned}$$

The dual point of view (Chapman-Kolmogorov) leads to the chemical master equation coupled to a Liouville type equation (cf. [8, Chap. 3.4]).

4 Algorithmic realization of the hybrid models

At least three different algorithmic approaches for simulating the stochastic part in the hybrid model (5) are available: (i) the first reaction method [10, 11], (ii) the direct method [10, 11] and (iii) the next reaction method [9]. In the following, we will discuss algorithmic realizations of the hybrid model based on each of these.

To illustrate the basic idea, consider a single chemical species S_1 being involved both in a reaction R_1 modelled stochastically and a reaction R_2 modelled deterministically. Set the initial time $\tau = t_0$ and the number of molecules $X(\tau) = X_1(t_0)$. Draw a random number $\xi_1 \sim \text{Exp}(1)$ and initialize $g_1(\tau|t) = 0$. As long as $g_1(\tau|t) < \xi_1$ no stochastic event occurs and the dynamics is simply given by the deterministic part

$$\frac{\mathrm{d}X}{\mathrm{d}\tau}(\tau) = \nu_2\, a_2(X(\tau), \tau) \tag{6}$$

(cf. eq. (5)). In the course of the (deterministic) evolution, however, the value of $g_1(\tau|t)$ increases following the differential equation

$$\frac{\mathrm{d}}{\mathrm{d}\tau} g_1(\tau|t) = a_1(X(\tau), \tau). \tag{7}$$

Hence, algorithmically, we simply simultaneously solve the system of ODEs (6) and (7). The numerical solution of such ordinary differential equations is very well documented (e.g., [5, 16, 17, 25]). We therefore suppose that we are able to compute the solution of the ODEs up to any desired accuracy and neglect the discretization error in what follows. Due to eq. (3), we know that the first stochastic reaction occurs according to the random variable $T_1(t_0)$. Consequently, we integrate the system of ODEs (6) and (7) until time $\tau = s$ such that $g_1(s|t) = \xi_1$. Thus $T_1(t_0) = s$ and the stochastic reaction is performed. The entire procedure is then repeated until a specified final time.

The above algorithmic scheme requires the use of numerical integrators that allow to stop integration when some so-called event function (in our case $\tau \mapsto g_1(\tau|t) - \xi_1$) vanishes. A few numerical ODE integrators already include event handling. However, the special nature of our event function makes it easy to implement the event detection, since $g_1(\tau|t) - \xi_1$ increases in τ (it is negative before the event time and positive thereafter). Thus one numerically integrates the system of ODEs until the first time τ_+ when the event function $g_1(\tau_+|t) - \xi_1$ becomes positive, while it is still negative at the previous time $\tau_- = \tau_+ - \delta t$ (δt being the time step). Then, a polynomial interpolation between $(t_-, g_1(\tau_-|t) - \xi_1)$ and $(t_+, g_1(\tau_+|t) - \xi_1)$ can be used in order to evaluate the event time $s \in [\tau_-, \tau_+]$ at which $g(s|t) - \xi_1 = 0$. This procedure is the same used to generate a dense output [16, 17], and is a well established method. Alternatively, with some additional computational cost, a finer time step could be used to solve the ODE in the interval $[\tau_-, \tau_+]$.

4.1 The direct hybrid method

The direct method explicitly calculates which reaction occurs next and when it occurs [10, 11]. The reaction time is given by the jump of the Poisson process $N_\sigma(t) = \sum_{j\in\mathcal{S}} N_j(t)$ with intensity given by the cumulative time transformation $g_\sigma(\tau|t) = \sum_{j\in\mathcal{S}} g_j(\tau|t)$. Thus the algorithmic realization of the direct hybrid method is as follows:

1. Set initial time $t = t_0$ and initial numbers of molecules $X(t_0)$;
2. Generate a random variable $\xi \sim \mathrm{Exp}(1)$;
3. Set $g_\sigma(t|t) = 0$ and solve the system of ODEs starting at time $\tau = t$

$$\frac{\mathrm{d}X}{\mathrm{d}\tau}(\tau) = \sum_{j\in\mathcal{D}} \nu_j a_j(X(\tau),\tau) \tag{8}$$

$$\frac{\mathrm{d}g_\sigma}{\mathrm{d}\tau}(\tau|t) = \sum_{j\in\mathcal{S}} a_j(X(\tau),\tau) \tag{9}$$

   until time $\tau = s$ such that $g_\sigma(s|t) = \xi$;

4. Generate a discrete random variable with values in $\mathcal{S}$ and probabilities $(a_j(X(s),s))_{j\in\mathcal{S}}$ in order to determine the reaction R_m to be performed;
5. Update $X(s)$ according to reaction R_m, hence set $X(s) \leftarrow X(s) + \nu_m$; set $t \leftarrow s$ and go to Step 2.

4.2 The first and next reaction hybrid methods

In the first reaction method, a putative reaction time is generated for each reaction; the reaction corresponding to the smallest time is chosen to occur, the state vector $X(t)$ is accordingly updated and the process repeated [10, 11]. The next reaction method is an efficient and economic (from the point of view of use of random variables) variant of the first reaction method. It is based on reuse of random variables and optimized data structures [9]. This reuse allows us to sample only one random variable at each iteration (instead of two like in the direct method). The algorithmic realizations of the first and the next reaction hybrid methods are very similar, thus we state them in a compact form:

1. Set initial time $t = t_0$ and initial numbers of molecules $X(t_0)$;
2. Generate independent random variables $(\xi_j)_{j\in\mathcal{S}}$, one for each reaction R_j in $\mathcal{S}$, with $\xi_j \sim \mathrm{Exp}(1)$;
3. Set $g_j(t|t) = 0$ for $j \in \mathcal{S}$;

No.	reaction	propensity	rate (day^{-1})	state change
R_1	gen $\xrightarrow{c_1}$ tem	$a_1 = c_1 \cdot \mathrm{gen}$	$c_1 = 0.025$	ν_1=(1,-1,0)
R_2	tem $\xrightarrow{c_2}$ $\emptyset$	$a_2 = c_2 \cdot \mathrm{tem}$	$c_2 = 0.25$	ν_2=(-1,0,0)
R_3	tem $\xrightarrow{c_3}$ tem+ gen	$a_3 = c_3 \cdot \mathrm{tem}$	$c_3 = 1.0$	ν_3= (0,1,0)
R_4	gen+struc $\xrightarrow{c_4}$ "virus"	$a_4 = c_4 \cdot \mathrm{gen} \cdot \mathrm{struc}$	$c_4 = 7.5 \cdot 10^{-6}$	ν_4=(0,-1,-1)
R_5	tem $\xrightarrow{c_5}$ tem+struc	$a_5 = c_5 \cdot \mathrm{tem}$	$c_5 = 1000$	ν_5=(0,0,1)
R_6	struc $\xrightarrow{c_6}$ $\emptyset$	$a_6 = c_6 \cdot \mathrm{struc}$	$c_6 = 1.99$	ν_6=(0,0,-1)

Table 1: T7 model equations

4. Solve the system of ODEs starting at time $\tau = t$

$$\frac{\mathrm{d}X}{\mathrm{d}\tau}(\tau) = \sum_{j\in\mathcal{D}} \nu_j a_j(X(\tau),\tau) \tag{10}$$

$$\frac{\mathrm{d}g_j}{\mathrm{d}\tau}(\tau|t) = a_j(X(\tau),\tau); \qquad j \in \mathcal{S} \tag{11}$$

until time $\tau = s$ such that for the first time $g_m(s|t) = \xi_m$ for some $m \in \mathcal{S}$;

5. Update $X(s)$ according to reaction R_m, hence set $X(s) \leftarrow X(s) + \nu_m$;

6. Set $t \leftarrow s$ and
   (i) First reaction hybrid method: go to Step 2.
   (ii) Next reaction hybrid method: For reaction R_m generate a new random variable $\xi_m \sim \mathrm{Exp}(1)$ and set $g_m(t|t) = 0$, while keeping all other values $g_j(t|t)$ for $j \neq m$ as initial values for the system of ODEs (10) and (11); go to Step 4.

For an alternative algorithmic realization of the hybrid next reaction method based on rescaling of the random variables ξ_j, we simply replace Step 6 by

6' (ii) For reaction R_m, generate a new random variable $\xi_m \sim \mathrm{Exp}(1)$; for the remaining reactions R_j, with $j \in \mathcal{S}$ and $j \neq m$, rescale ξ_j according to $\xi_j \leftarrow \xi_j - g_j(s|t)$. Set $t \leftarrow s$ and go to Step 3.

5 Numerical studies for the T7 model

We now wish to demonstrate the power of our hybrid method using a model derived by Srivastava et al. [29] describing the intracellular growth of bacteriophage T7. We choose this model since it clearly shows the difference between deterministic and stochastic modelling (Figure 1) and thus allows to check whether the hybrid method is able to speed up simulation without compromising the results.

The bacteriophage T7 test model comprises three chemical components: viral nucleic acids classified into genomic (gen) and template (tem) and viral structural proteins (struc). The infection process is modelled by six reactions (Table 1). In the sequel, we will focus on the low infection level corresponding to the initial numbers of molecules tem $= 1$, gen $=$ struc $= 0$.

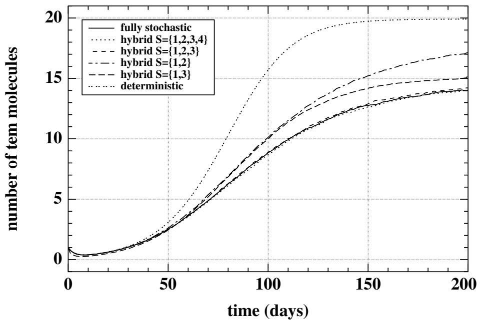

Figure 1: Comparison of the deterministic model with the means of the fully stochastic and different hybrid models (based on 10^4 realizations).

As was shown in [29], the deterministic model of the above system possesses two stationary points: (1) the point tem $=$ gen $=$ struc $= 0$, which is unstable (the system will move away from it after small perturbations) and (2) the point tem $= 20$, gen $= 200$ and struc $= 10000$, which is stable (the system will return to it after small perturbations) and attractive (the system tends to reach this state, if possible).

As an analysis of the propensities (e.g., by running a single realization) reveals, reactions R_5 and R_6 happen much more frequently than the others, suggesting that we model them deterministically, while treating the remaining ones stochastically. For our T7 model, this can even be made more precise: The propensities at the stable steady state are:

$$\begin{array}{lll} a_1 = 5 & a_2 = 5 & a_3 = 20 \\ a_4 = 15 & a_5 = 20000 & a_6 = 19900 \end{array} \tag{12}$$

which clearly show a separation of time scale between the two sets of reactions. The results for 10^4 realizations of the corresponding hybrid model are shown in Figure 2. As in [29], we display results related to the number of tem molecules. It can be seen that the distributions obtained with the hybrid model are almost indistinguishable from those obtained with the fully stochastic model. The advantage of the hybrid model is a considerable saving of CPU time in numerical simulations. For the sake of comparison, we have simulated the fully stochastic and the hybrid models based on either the direct method [10, 11] or the next reaction method [9]. The observed CPU times are reported in Table 2. It appears that the hybrid simulations are about 100 times as fast as the fully stochastic ones. The hybrid simulation is based on a Runge Kutta integrator of order 4 with constant time step

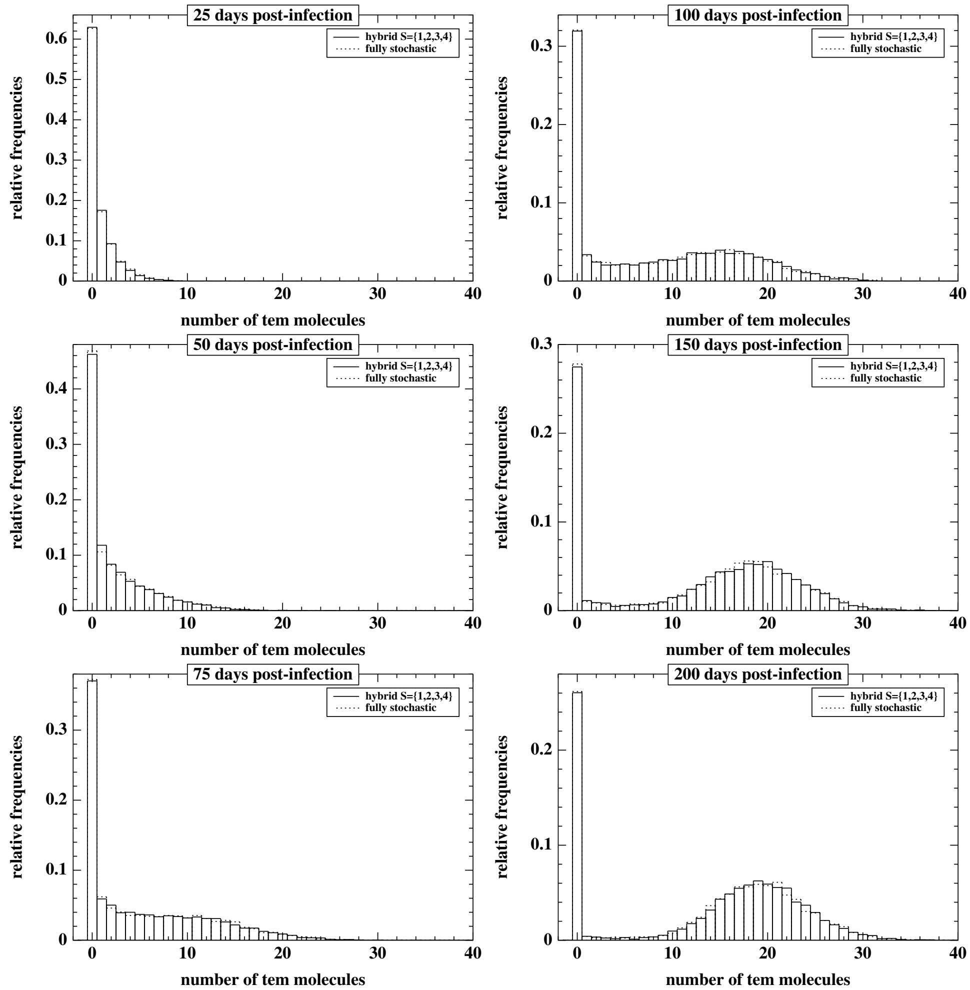

Figure 2: Hybrid kinetics for the bacteriophage T7 model (reactions R_1, R_2, R_3 and $R4$ modelled stochastically, reactions R_5, R_6 modelled deterministically) compared to the reference fully stochastic model: post-infection distribution of tem molecules (based on 10^4 realizations).

$\delta t = 0.01$ day [5, 17, 16]. In order to verify the accuracy of the event detection, we exploit the existence of an analytical solution for hybrid equations for $\mathcal{S} = \{1, 2, 3, 4\}$. Hence solving the event equations (9) or (11) can be accomplished by a few Newton iterations [6]. The results obtained are very similar suggesting that the event detection is performed with sufficient accuracy.

Model	Direct method	Next reaction method
Fully stochastic	15600 s	21200 s
Hybrid model:		
$\mathcal{S} = \{1, 2, 3, 4\}$	183 s	201 s

Table 2: Comparison of fully stochastic and hybrid method in terms of CPU time (10^4 realizations, Fortran 77 code run on a Pentium 1.4 GHz processor).

It is instructive to perform simulations under different choices of $\mathcal{S}$ and $\mathcal{D}$ in order to identify the reactions whose stochastic effects contribute most to the overall dynamical behavior of the T7 model system (as was also pointed out in [4]). As can be inferred from Figure 3 (left), reducing the number of reactions modelled stochastically to $\mathcal{S} = \{1, 2, 3\}$ hardly changes the results in terms of distribution profiles. Interestingly, the simulation times are somewhat longer, for the time step has to be reduced by a factor 10 to maintain the desired accuracy. On the other hand, a further reduction of the number of reactions modelled stochastically dramatically modifies the overall dynamical behavior of the system (see Figure 3). However, the mean values are still closer to the fully stochastic than to the deterministic model (Figure 1). Hence, the interplay of the three reactions R_1, R_2 and R_3, and the order of their occurrence seems to be an important part of the regulatory network. Neglecting these stochastic effects induces (large) deviations from the reference fully stochastic model.

6 Discussion

We presented a mathematical derivation for hybrid stochastic and deterministic modelling and three exact simulation algorithms that are easy to implement. The power and applicability of the hybrid modelling approach has been demonstrated on a model for bacteriophage T7 model [29] that was especially designed to analyze the influence of stochastic fluctuations on the overall dynamical behavior. As a result, we gained a speed up of the simulations by two orders of magnitude without compromising the statistics.

In the past years several approaches have been proposed to speed up stochastic simulations. Most of the times, the starting point is an approximation of the stochastic process $N_j(t)$ (see eq. (2)) counting the number of times reaction R_j occurs. In the case of tau-leaping methods [3, 4, 14, 28] $N_j(t)$ is approximated by a Poisson process, while in the case of [18, 26] averaging techniques are applied to those $N_j(t)$ that correspond to fast reactions. Very recently, two algorithms to simulate hybrid models have been proposed [20, 32]. The main difference to our approach is that they lack a mathematical justification and are not exact algorithmic realizations of the hybrid system (5). They are based on a prediction correction heuristic for the realization of the stochastic part that can be seen as an approximation to the simultaneous solution of the system of ODEs (8) and (9), or (10) and (11) of our hybrid methods. Gibson and Bruck [9] already presented an exact solution to simulate stochastic systems with a linear increase of the reaction volume (treated deterministically);

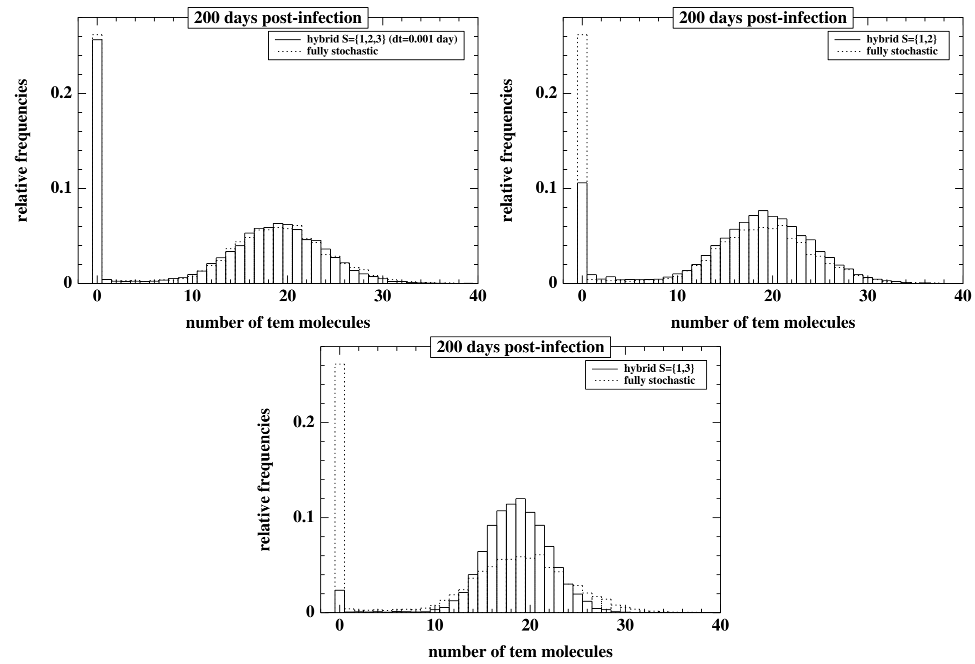

Figure 3: Results for different hybrid systems for the bacteriophage T7 model (solid lines) compared to the reference fully stochastic system (dotted line) based on 10^4 realizations.

however, to the best of our knowledge, the general strategy presented herein has not be considered before.

The first promising results of our hybrid approach indicate some further directions of research. The numerical solutions of the hybrid system were obtained with some constant step size integrator. Here, an adaptive step size control is likely to further speed up simulations, which should be combined with a more detailed analysis of event detection. It was explicitly not our aim to theoretically justify any partitioning of the set of reactions into stochastic and deterministic ones. A mathematical analysis of the approximation error of the hybrid model compared to fully stochastic one is needed, in particular when aiming at adaptively partitioning reactions based on propensity functions and number of reactant molecules. For the T7 model different partitions have been analyzed. Interestingly, the computational speed up does not simply increase with the number of reactions modelled deterministically. This interplay between speed up and partitioning also need further analysis that could eventually lead to a rationale for setting up hybrid models.

Acknowledgements

This work has been initiated during the CEMRACS 2004 research center, held in Marseilles (France) from July 26th to September 3rd, 2004. E.C. and G.T. acknowledge financial support from the ARC Prestissimo (INRIA). B.D.V. acknowledges Dr. Luis Serrano for his precious advices and the EU STREP Grant (COMBIO:LSHG-CT-2004-503568) for funding. W.H. acknowledges financial support by the DFG Research Center Matheon, Berlin. We all thank Frédéric Legoll for fruitful discussions on this subject.

References

[1] Adam P. Arkin, John Ross, and Harley H. McAdams. Stochastic kinetic analysis of developmental pathway bifurcation in phage λ-infected *Escherichia coli* cells. *Genetics*, 149:1633–1648, 1998.

[2] N. Barkai and S. Leibler. Robustness in simple biochemical networks. *Nature*, 387:913–917, 1997.

[3] Kevin Burrage and Tianhai Tian. Poisson-Runge-Kutta methods for chemical reaction systems. *in: Advances in Scientific Computing and Applications, Y. Lu W. Sun and T. Tang eds, Science Press*, pages 82–96, 2004.

[4] Kevin Burrage, Tianhai Tian, and Pamela Burrage. A multi-scale approach for simulating chemical reaction systems. *Progress Biophys Mol Biol*, 85:217–234, 2004.

[5] Peter Deuflhard and Folkmar Bornemann. *Scientific Computing with Ordinary Differential Equations*, volume 42 of *TAM*. Springer, 2002.

[6] Peter Deuflhard and Andreas Hohmann. *Introduction to Scientific Computing*. Springer, 2002.

[7] Michael B. Elowitz, Eric D. Siggia, Peter S. Swain, and Arnold J. Levine. Stochastic gene expression in a single cell. *Science*, 297:1183–1186, 2002.

[8] C. W. Gardiner. *Handbook of Stochastic Methods—for Physics, Chemistry and the Natural Sciences*. Springer, 2nd edition, 6th printing 2002.

[9] Michael A. Gibson and Jehoshua Bruck. Efficient exact stochastic simulation of chemical systems with many species and many channels. *J Phys Chem*, 104:1876–1889, 2000.

[10] Daniel T. Gillespie. A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. *J Comput Phys*, 22:403–434, 1976.

[11] Daniel T. Gillespie. Exact stochastic simulating of coupled chemical reactions. *J Phys Chem*, 81:2340–2361, 1977.

[12] Daniel T. Gillespie. The chemical Langevin and Fokker-Planck equation for reversible isomerization. *J Phys Chem A*, 106:5063–5071, 1992.

[13] Daniel T. Gillespie. *Markov Processes—An Introduction for Physical Scientists.* Academic Press, 1992.

[14] Daniel T. Gillespie. Approximate accelerated stochastic simulation of chemically reacting systems. *J Chem Phys*, 115:1716–1733, 2001.

[15] D. Gonze, J. Halloy, and A. Goldbeter. Deterministic versus stochastic models for circadian rhythms. *J Biol Phys*, 28:637–653, 2002.

[16] E. Hairer, P. Nørsett, and G. Wanner. *Solving ordinary differential equations: Non-stiff problems.* Springer, 1993.

[17] E. Hairer and G. Wanner. *Solving ordinary differential equations: Stiff and differential-algebraic problems.* Springer, 1996.

[18] Eric L. Haseltine and James B. Rawlings. Approximate simulation of coupled fast and slow reactions for stochastic chemical kinetics. *J Chem Phys*, 117:6959–6969, 2002.

[19] James Keener and James Sneyd. *Mathematical Physiology*, volume 8 of *IAM.* Springer, 2001.

[20] Thomas R. Kiehl, Robert M. Mattheyses, and Melvin K. Simmons. Hybrid simulation of cellular behavior. *Bioinformatics*, 20:316–322, 2004.

[21] E. Korobkova, T. Emonet, J. M. G. Vilar Thomas, S. Shimizuß, and P. Cluzel. From molecular noise to behavioural variability in a single bacterium. *Nature*, 428:574–578, 2004.

[22] H. H. McAdams and A. P. Arkin. Stochastic mechanisms in gene expression. *PNAS*, 94:814–819, 1997.

[23] H. H. McAdams and A. P. Arkin. It's a noisy business! Genetic regulation at the nanomolar scale. *Trends Genet*, 15:65–69, 1999.

[24] M. Morohashi, A. E. Winn, M. T. Borisuk, H. Bolouri, J. Doyle, and H. Kitano. Robustness as a measure of plausibility in models of biochemical networks. *J Theor Biol*, 216:19–30, 2002.

[25] William H. Press, Brian P. Flannery, Saul A. Teukolsky, and William T. Vetterling. *Numerical Recipes in C, The Art of Scientific Computing.* Cambridge University Press, 2nd edition, 1992. www.nr.com.

[26] Christopher V. Rao and Adam P. Arkin. Stochastic chemical kinetics and the quasi-steady-state assumption: Application to the Gillespie algorithm. *J Chem Phys*, 118:4999–5010, 2003.

[27] Jonathan M. Raser and Erin K. O'Shea. Control of stochasticity in eukaryotic gene expression. *Science*, 304:1811–1814, 2004.

[28] Muruhan Rathinam, Linda R. Petzold, Yang Cao, and Daniel T. Gillespie. Stiffness in stochastic chemically reaction systems: the implicit tau-leaping method. *J Chem Phys*, 119:12784–12794, 2003.

[29] R. Srivastava, L. You, J. Summer, and J. Yin. Stochastic vs. deterministic modeling of intracellular viral kinetics. *J theor Biol*, 218:309–321, 2002.

[30] Jorg Stelling, Uwe Sauer, Zoltan Szallasi, Francis J. Doyle III, and John Doyle. Robusteness of cellular functions. *Cell*, 118:675–685, 2004.

[31] Ralf Steuer. Effects of stochasticity in models of the cell cycle: from quantized cycle times to noise-induced oscillations. *J Theor Biol*, 228:293–301, 2004.

[32] Kouichi Takhashi, Kazunari Kaizu, Bin Hu, and Masuro Tomita. A multi-algorithm, multi-timescale method for cell simulation. *Bioinformatics*, 20:538–546, 2004.

[33] T. E. Turner, S. Schnell, and K. Burrage. Stochastic approaches for modelling in vivo reactions. *Comp Bio Chem*, 28:165–178, 2004.

[34] T. Wilhelm, J. Behre, and S. Schuster. Analysis of structural robustness of metabolic networks. *Systems biology*, 1:114–120, 2004.

A Appendix

Stochastic model.

Here, we derive the infinitesimal generator and the chemical master equation for the Markov jump process $X(t)$ defined in (4). Consider the system $X(t)$ at time t and consider a time increment $\Delta t > 0$. We calculate
(i) the probability that there is no reaction between t and $t + \Delta t$ as

$$\begin{aligned}\mathbf{P}[N_j(t+\Delta t) - N_j(t) = 0;\, j = 1,..,M] &= \prod_{j=1}^{M} \exp(-a_j(X(t),t)\Delta t) \\ &= 1 - \sum_{j=1}^{M} a_j(X(t),t)\Delta t + \mathcal{O}(\Delta t^2);\end{aligned}$$

(ii) the probability that only the reaction R_k occurs between t and $t + \Delta t$

$$\mathbf{P}[N_k(t+\Delta t) - N_k(t) = 1 \text{ and } N_j(t+\Delta t) - N_j(t) = 0;\, j = 1,\dots,M, j \neq k] = a_k(X(t),t)\Delta t + \mathcal{O}(\Delta t^2).$$

Therefore, the probability that there are two or more reactions between t and $t + \Delta t$ is $\mathcal{O}(\Delta t^2)$. Now, for any bounded function $f : \mathbb{N}^N \to \mathbb{R}$ we obtain

$$\begin{aligned}\mathbf{E}[f(X(t+\Delta t))\,|X(t)] &= f(X(t))(1-\sum_{j=1}^{M} a_j(X(t),t)\,\Delta t) \\ &\quad + \sum_{j=1}^{M} f(X(t)+\nu_j)a_j(X(t),t)\Delta t + \mathcal{O}(\Delta t^2),\end{aligned}$$

so we get that $X(t)$ is Markovian, and

$$\lim_{s\to t^+}\frac{\mathrm{d}}{\mathrm{d}s}\mathbf{E}[f(X(s))\,|X(t)] = \sum_{j=1}^{M} f(X(t)+\nu_j)a_j(X(t),t) - a_j(X(t),t)f(X(t)). \tag{13}$$

The right hand side of the above equation gives us the infinitesimal generator

$$\mathcal{A}f(x) = \sum_{j=1}^{M} f(x+\nu_j)a_j(x,t) - a_j(x,t)f(x)$$

of the Markov process $X(t)$ which entirely characterizes it. Considering $f(X(t)) = \mathbf{1}_{\{X(t)=x\}}$ in (13), we obtain the Chemical Master Equation

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbf{P}[X(t)=x] = \sum_{j=1}^{M}\mathbf{P}[X(t)=x-\nu_j]a_j(x-\nu_j,t) - a_j(x,t)\mathbf{P}[X(t)=x].$$

The hybrid model.

Consider the hybrid system $X(t)$ at time t and consider a time increment $\Delta t > 0$. Considering a bounded, differentiable observable $f : (\mathbb{R}_+)^N \to \mathbb{R}$ we obtain

$$\begin{aligned}\mathbf{E}[f(X(t+\Delta t))\,|X(t)] &= f\Big(X(t)+\sum_{j\in\mathcal{D}} a_j(X(t),t)\Delta t + \mathcal{O}(\Delta t^2)\Big)\times \\ &\quad \Big(1-\sum_{j\in\mathcal{S}} a_j(X(t),t)\Delta t\Big) \\ &\quad + \sum_{j\in\mathcal{S}} f(X(t)+\nu_j+\mathcal{O}(\Delta t))a_j(X(t),t)\Delta t + \mathcal{O}(\Delta t^2)\end{aligned}$$

and thus

$$\begin{aligned}\lim_{s\to t^+}\frac{\mathrm{d}}{\mathrm{d}s}\mathbf{E}[f(X(s))\,|X(t)] &= \sum_{j\in\mathcal{D}} a_j(X(t),t)\frac{\mathrm{d}}{\mathrm{d}t}f(X(t)) \\ &\quad + \sum_{j\in\mathcal{S}} f(X(t)+\nu_j)a_j(X(t),t) - a_j(X(t),t)f(X(t))\end{aligned}$$

implying the infinitesimal generator

$$\mathcal{A}f(x) = \sum_{j\in\mathcal{D}} a_j(x,t)\frac{\mathrm{d}}{\mathrm{d}t}f(x) + \sum_{j\in\mathcal{S}} f(x+\nu_j)a_j(x,t) - a_j(x,t)f(x),$$

which fully characterizes the hybrid process $X(t)$. Thus, we have $\frac{\mathrm{d}}{\mathrm{d}t}\mathbf{E}[f(X_t)] = \mathbf{E}[\mathcal{A}f(X_t)]$. We can then take the dual point of view and get the Chapman-Kolmogorov equation for the grand probability measure

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}\mathbf{P}[X(t)\in \mathrm{d}x] &= -\sum_{j\in\mathcal{D}}\frac{\mathrm{d}}{\mathrm{d}x}\Big(a_j(x,t)\mathbf{P}[X(t)\in\mathrm{d}x]\Big) \\ &\quad + \sum_{j\in\mathcal{S}} a_j(x-\nu_j,t)\mathbf{P}[X(t)-\nu_j\in\mathrm{d}x] - a_j(x,t)\mathbf{P}[X(t)\in\mathrm{d}x],\end{aligned} \tag{14}$$

where the derivation with respect to x must be understood in the sense of distribution theory. The second term of the right hand side is simply the fully stochastic Chemical Master Equation, while the first term is due to the deterministic model.

Reuse of random numbers for the next reaction method.

Here we justify the reuse of the exponential random variables in the setting of the next reaction method. Consider M' independent exponential random variables $\xi^1,\dots,\xi^{M'}$ of parameter 1 and nondecreasing functions $f^1,\dots,f^{M'}:\mathbb{R}^+\to\mathbb{R}^+$ with inverse $(f^j)^{-1}(y) = \inf\{x\ge 0 : f^j(x)\ge y\}$.

Then, the law of $(\xi^1,\dots,\xi^{M'-1})$ conditioned to $\{f^{M'}(\xi^{M'}) = \min_{1\le j\le M'} f^j(\xi^j)\}$ is equal in law to $(\xi^1 + (f^1)^{-1}[f^{M'}(\xi^{M'})],\dots,\xi^{M'-1} + (f^{M'-1})^{-1}[f^{M'}(\xi^{M'})])$.

To prove this, consider M' continuous bounded functions g^1,..,$g^{M'}$. We have to calculate

$$\begin{aligned} E &= \mathbf{E}\Big(g^1(\xi^1)g^2(\xi^2)\dots g^{M'-1}(\xi^{M'-1})|f^{M'}(\xi^{M'}) = \min_{1\le j\le M'} f^j(\xi^j)\Big) \\ &= \frac{\mathbf{E}\Big[\prod_{j=1}^{M'-1} g^j(\xi^j)\mathbf{1}_{\xi^j\ge(f^j)^{-1}(f^{M'}(\xi^{M'}))}\Big]}{\mathbf{E}\Big[\prod_{j=1}^{M'-1}\mathbf{1}_{\xi^j\ge(f^j)^{-1}(f^{M'}(\xi^{M'}))}\Big]} = \frac{\prod_{j=1}^{M'-1}\mathbf{E}\Big[g^j(\xi^j)\mathbf{1}_{\xi^j\ge(f^j)^{-1}(f^{M'}(\xi^{M'}))}\Big]}{\prod_{j=1}^{M'-1}\mathbf{P}\Big[\xi^j\ge(f^j)^{-1}(f^{M'}(\xi^{M'}))\Big]}\end{aligned}$$

using the independence. Using the lack of memory property of the exponential variable (i.e. $\forall x\ge a,\ \mathbf{P}[\xi^j > x|\xi^j > a] = \mathbf{P}[\xi^j + a > x]$), we get

$$\mathbf{E}\Big[g^j(\xi^j)\mathbf{1}_{\xi^j\ge(f^j)^{-1}(f^{M'}(\xi^{M'}))}\Big] = \mathbf{E}\Big[g^j\big(\xi^j + (f^j)^{-1}(f^{M'}(\xi^{M'}))\big)\Big]\mathbf{P}\Big[\xi^j\ge(f^j)^{-1}(f^{M'}(\xi^{M'}))\Big]$$

and finally

$$E = \prod_{j=1}^{M'-1} \mathbf{E}\Big[g^j(\xi^j + (f^j)^{-1}(f^{M'}(\xi^{M'})))\Big]$$

which proves the statement.

Analytical solution of the hybrid system for $\mathcal{S} = \{1, 2, 3, 4\}$.

Let us remark that for $\mathcal{S} = \{1, 2, 3, 4\}$ and $\mathcal{D} = \{5, 6\}$, system of ODEs (8)-(9) reads

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}\tau}\mathrm{tem} &= 0\\
\frac{\mathrm{d}}{\mathrm{d}\tau}\mathrm{gen} &= 0\\
\frac{\mathrm{d}}{\mathrm{d}\tau}\mathrm{struc} &= c_5 \cdot \mathrm{tem} - c_6 \cdot \mathrm{struc}\\
\frac{\mathrm{d}}{\mathrm{d}\tau}g_\sigma(\tau|t) &= c_1 \cdot \mathrm{gen} + (c_2 + c_3) \cdot \mathrm{tem} + c_4 \cdot \mathrm{gen} \cdot \mathrm{struc}
\end{aligned}$$

with initial conditions $\mathrm{tem}(t), \mathrm{gen}(t), \mathrm{struc}(t)$ at $\tau = t$. It has the analytical solution

$$\begin{aligned}
\mathrm{tem}(\tau) &= \mathrm{tem}(t)\\
\mathrm{gen}(\tau) &= \mathrm{gen}(t)\\
\mathrm{struc}(\tau) &= \Big(\mathrm{struc}(t) - \frac{c_5}{c_6} \cdot \mathrm{tem}(t)\Big)\, e^{-c_6(\tau - t)} + \frac{c_5}{c_6} \cdot \mathrm{tem}(t)\\
g_\sigma(\tau|t) &= \Big(c_1 \cdot \mathrm{gen}(t) + (c_2 + c_3) \cdot \mathrm{tem}(t) + \frac{c_4 c_5}{c_6} \cdot \mathrm{gen}(t) \cdot \mathrm{tem}(t)\Big) \cdot (\tau - t)\\
&\quad + \frac{c_4 \cdot \mathrm{gen}(t) \cdot \Big(\mathrm{struc}(t) - \frac{c_5}{c_6} \cdot \mathrm{tem}(t)\Big)}{c_6}\Big(1 - e^{-c_6(\tau - t)}\Big).
\end{aligned}$$

Contents

Unité de recherche INRIA Rocquencourt
Domaine de Voluceau - Rocquencourt - BP 105 - 78153 Le Chesnay Cedex (France)

Unité de recherche INRIA Futurs : Parc Club Orsay Université - ZAC des Vignes
4, rue Jacques Monod - 91893 ORSAY Cedex (France)
Unité de recherche INRIA Lorraine : LORIA, Technopôle de Nancy-Brabois - Campus scientifique
615, rue du Jardin Botanique - BP 101 - 54602 Villers-lès-Nancy Cedex (France)
Unité de recherche INRIA Rennes : IRISA, Campus universitaire de Beaulieu - 35042 Rennes Cedex (France)
Unité de recherche INRIA Rhône-Alpes : 655, avenue de l'Europe - 38334 Montbonnot Saint-Ismier (France)
Unité de recherche INRIA Sophia Antipolis : 2004, route des Lucioles - BP 93 - 06902 Sophia Antipolis Cedex (France)

Éditeur
INRIA - Domaine de Voluceau - Rocquencourt, BP 105 - 78153 Le Chesnay Cedex (France)
http://www.inria.fr
ISSN 0249-6399