ADAPTIVE SIMULATION OF HYBRID STOCHASTIC AND DETERMINISTIC MODELS FOR BIOCHEMICAL SYSTEMS

Aurélien Alfonsi[1], Eric Cancès[2], Gabriel Turinici[3], Barbara Di Ventura[4] and Wilhelm Huisinga[5]

Abstract. In the past years it has become evident that stochastic effects in regulatory networks play an important role, leading to an increasing in stochastic modelling attempts. In contrast, metabolic networks involving large numbers of molecules are most often modelled deterministically. Going towards the integration of different model systems, gen-regulatory networks become part of a larger model system including signalling pathways and metabolic networks. Thus, the question arises of how to efficiently and accurately simulation such coupled or hybrid systems. We present an algorithmic approach for the simulation of hybrid stochastic and deterministic reaction models that allows for adaptive step-size integration of the deterministic equations while at the same time accurately tracing the stochastic reaction events. We present a mathematical derivation of the hybrid system on the stochastic process level, and present numerical examples that outline the power of hybrid simulations.

Résumé. Au cours des dernières années, il est devenu clair que les effets aléatoires jouaient un rôle important dans les réseaux de régulation, et les modèles employés aujourd'hui pour décrire ces réseaux sont de nature stochastique. En revanche, les réseaux métaboliques, qui mettent en jeu un grand nombre de molécules, sont le plus souvent décrits par des modèles déterministes. Dans la modélisation de systèmes complexes, réseaux régulateurs de gènes, chemins de signaux et réseaux métaboliques sont intégrés dans un même modèle. Se pose alors la question de simuler efficacement et avec précision de tels modèles couplés (on parle aussi de modèles hybrides). Nous présentons ici une approche pour la simulation de modèles de réactions hybrides stochastiques/déterministes permettant à la fois d'avoir recours à des pas de temps adaptatifs dans l'intégration des équations déterministes et de simuler précisément les réactions décrites par des processus stochastiques. Des simulations numériques illustrent la puissance de ces simulations hybrides.

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

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

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

[3] INRIA Rocquencourt, Domaine de Voluceau, 78153 Le Chesnay Cedex, France and CERMICS, ENPC, 6-8 Avenue Blaise Pascal, Cité Descartes, 77455 Marne la Vallée Cedex 2, France

[4] Structural Biology and Biocomputing program, EMBL, Meyerhofstrasse 1, 69117 Heidelberg, Germany

[5] Corresponding author: Free University Berlin, Department of Mathematics and Computer Science, Arnimallee 2-6, D-14195 Berlin/Germany, and DFG Research Center Matheon, Berlin, `huisinga@math.fu-berlin.de`

translation [10, 26, 27]. 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 [3,28,35,40], noise is, for instance, used as source of phenotypic variability, which is favorable in evolutionary terms [23, 31]. 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 [2, 18, 34, 36]. 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 [13,14,39]. Although a more efficient exact method has been proposed by Gibson and Bruck, based on reuse of random numbers and intelligent data structures [12], 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 application of quasi-steady-state theory [30], or grouping together reactions that occur in fast succession (the so-called tau leaping methods) [4, 5, 17, 32]. Another strategy is to model those processes that either involve large number of particles or have fast rates, in a continuous (stochastic or deterministic) way, keeping discrete the remaining ones. These approaches are based on a partition of the set of reactions into fast and slow ones. Haseltine and Rawlings [21] were the first proposing to model the fast reactions by a continuous Markov process being coupled to the Markov jump process for the slow reactions. The resulting coupling is equivalent to ours, however, their algorithmic realization requires to introduce artificial "no reaction events" to ensure accuray and is thus inappropriate for adaptive integration. Very recently, two algorithms to simulate biochemical systems in a hybrid framework have been proposed [22,37]. They are based on a prediction correction type heuristics for the realization of the discrete stochastic part. However, since the coupling of the discrete and continuous models is not stochastically exact, these algorithms only give good results, if the time-step of the continuous solver is kept smal. In the following, we propose a mathematical ground for a hybrid stochastic and deterministic modelling in a elegant and natural way. We also present three adaptive algorithms for its implementation and finally show first intriguing numerical results for the bacteriophage T7 model system.

1. STOCHASTIC AND DETERMINISTIC MODELS

Consider a well-mixed system of volume V with N chemical species $S_1, \ldots, S_N$ involved in M reactions $R_1, \ldots, R_M$. The state of the system is modelled in terms of number of molecules $X(t) = (X_1(t), \ldots, X_N(t))$ of the chemical species. The reaction probability for each reaction R_j is specified in terms of the 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 [13, 14]. 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., $X(t) \leftarrow X(t) + \nu_j$.

The stochastic system. 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 an $\mathbb{N}^N$-valued Markov jump process $\{X(t), t \geq t_0\}$, where t_0 denotes the initial time [16, 39]. The probability that the reaction R_j occurs in the next infinitesimal time interval $[t, t + \mathrm{d}t]$ is given in terms of the propensity function and equals $a_j(X(t), t)\mathrm{d}t$. In order to state the time evolution equation for $X(t)$, denote by $\{N_j(t), t \geq t_0\}$ the stochastic process counting the number of times that reaction R_j occurs in the time interval $[t_0, t]$ with $N(t_0) = 0$. Then, obviously, we obtain

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

with some initial value $X(t_0)$. In order to specify the law of $N_j(t)$, we note that $N_j(t)$ is an inhomogeneous Poisson process specified by

$$\mathbf{P}[N_j(t + \mathrm{d}t) - N_j(t) = 1 | X(t)] = a_j(X(t), t)\mathrm{d}t. \tag{2}$$

Now, it is easy to show that the evolution equation (1) 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 [15, 39]

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

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

For coupling stochastic and deterministic models, we exploit the fact that any time inhomogeneous Poisson process can be transformed into a time homogeneous Poisson process with parameter 1 (e.g., [38, Chapt. 2.4], [7, Thm. 7.4.I]). Denote by $T_j(t)$ the time at which reaction R_j first occurs after t, then from eq. (2) follows

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

Note that 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

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

is fundamental. 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 $j = 1, \ldots, M$. Then, we get the following representation

$$N_j(t) = \sum_{n=1}^{\infty} \mathbf{1}_{\{S_j(n) \leq g_j(t|t_0)\}}, \tag{3}$$

which can be deduced from $\mathbf{P}[N_j(t + \mathrm{d}t) - N_j(t) = 1 | X(t)] = a_j(X(t), t)\mathrm{d}t$.

The following consequence of the above derivation will be exploited. Given that the system is in state $X(t)$ at time t, the next reaction time $T_j(k)$ equals the first (smallest) time $s > t$ satisfying

$$g_j(s|t) = \int_t^s a_j(X(\tau), \tau)\mathrm{d}\tau = \xi, \tag{4}$$

where ξ is a random number exponentially distributed with parameter 1.

The deterministic system. Under certain conditions the mathematical description of the system can be embedded into a deterministic framework given by the following evolution equation (ODE)

$$\mathrm{d}x(t) = \sum_{j=1}^{M} \nu_j \, a_j(x(t), t) \mathrm{d}t \tag{5}$$

with initial value $x_0 \in \mathbb{R}_+^N$. This approach can be justified by the system entering a specific region of its state space or to a limiting process (so-called thermodynamic limit, [24, 25]).

2. Hybrid 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 below). Based on a given partition, the evolution equation for $X(t) \in \mathbb{R}^N$ is now given by the hybrid system

$$\mathrm{d}X(t) \;\; = \;\; \sum_{j \in \mathcal{D}} \nu_j \, a_j(X(t), t) \, \mathrm{d}t + \sum_{j \in \mathcal{S}} \nu_j \, \mathrm{d}N_j(t) \tag{6}$$

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. The hybrid system (6) corresponds to the infinitesimal generator

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

The dual point of view (Chapman-Kolmogorov) leads to the chemical master equation coupled to a Liouville type equation (cf. [11, Chap. 3.4]).

Adaptive partitioning. Up to now, we assumed that the partition of the system was given a-priori, being fixed for the entire evolution. However, in practice, the choice of the stochastic or deterministic description is determined by whether the underlying assumptions of the models are satisfied, which are likely to change in time. In addition, from the computational point of view, it is also more efficient to use the deterministic description as soon as it is appropriate. This motivates to consider a time-depending partitioning of the reactions into the two classes.

Here, as was introduced previously [6, 21, 33], we use a criterion based on two conditions. A reaction is modelled by the deterministic equation, if (1) the reaction occurs many times in an small time interval, and

(2) the number of reactant molecules is sufficiently large. More quantitatively, the reaction R_j is modelled deterministically, if

(1) the expected next reaction time τ_j is smaller than some specified time increment Δt, i.e., for $s > t \geq t_0$ it is

$$\mathbf{E}[\tau_j | T_j = t] = 1/a_j(X(t), t) < \Delta t.$$

(2) the number of reactant molecules $X_r(t)$ involved in reaction R_j are larger than some specified number Λ, i.e.,

$$X_r(t) > \Lambda.$$

For our purpose, we set $\Delta t = 0.1$ and $\Lambda = 10$, but, of course, these numbers might be system-dependent.

In the light of our presentation, the first of the two criteria can be justified as follows: The fully stochastic simulations become quite slow when many molecules and/or 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

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

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

$$N_j(t) \approx g_j(t|t_0) = \int_{t_0}^t a_j(X(s), s) \mathrm{d}s \tag{7}$$

when the propensity a_j is large. This is the precisely what the first criterion states. The second criterion ensures that the quantity

$$\frac{a_j(X(t) + 1, t) - a_j(X(t), t)}{a_j(X(t), t)}$$

is small in modulus [17]. Although frequently used in practice, no complete studies on its practical relevance are as yet available. In particular, examples exist where this criterion can be relaxed while still maintaining good properties of the deterministic description.

3. Algorithmic realization of the hybrid models

At least three different algorithmic approaches for simulating the stochastic part in the hybrid model (6) are available: (i) the first reaction method [13, 14], (ii) the direct method [13, 14] and (iii) the next reaction method [12]. In the following, we will discuss algorithmic realizations of the hybrid model based on each of these.

To illustrate the basic idea, consider the hypothetical situation of 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 $\xi_1 \sim \mathrm{Exp}(1)$ 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

$$\frac{\mathrm{d}X}{\mathrm{d}\tau}(\tau) = \nu_2 \, a_2(X(\tau), \tau) \tag{8}$$

(cf. eq. (6)). In the course of the (deterministic) evolution, however, the value of $g_1(\tau|t)$ increases following the differential equation

$$\frac{\mathrm{d}}{\mathrm{d}\tau}g_1(\tau|t) \quad = \quad a_1(X(\tau),\tau). \tag{9}$$

Hence, algorithmically, we simply simultaneously solve the system of ODEs (8) and (9). The numerical solution of such ordinary differential equations is very well documented (e.g., [8,19,20,29]). 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. (4), we know that the first stochastic reaction occurs according to the random variable $T_1(t_0)$. Consequently, we integrate the system of ODEs (8) and (9) 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 $F(\tau) = g_1(\tau|t) - \xi_1$) 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 $F(\tau)$ 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 $F(\tau_+)$ becomes positive, while it is still negative at the previous time $\tau_- = \tau_+ - \delta t$ (δt being the time step). Then, one applies, e.g., the be-section method or the Newton method [9] to accurately determine the event time. For the Newton method, the next reaction time s is approximated by a sequence of times $\{s_k\}_{k\in\mathbb{N}}$ defined iteratively by

$$s_{k+1} - s_k = \Delta s_k = -\frac{F(s_k)}{F'(s_k)} = -\frac{F(s_k)}{a(X(s_k),s_k)}.$$

The iteration is stopped, if $|\Delta s_k| < \text{TOL}$, where TOL is some predefined tolerance. Both approaches requires the successive solution of the original system of ODEs. This, however, could be too computationally costly, if there are many species involved in the system (large N). However, the dimension of the ODE to be solved for the event detection can be reduced to a fixed, low dimension, if a special type of numerical integrator supplying a so-called dense output is used (see, e.g., [19, Chapt. II.5], [20, Chapt. IV.6] for Dormand and Prince Runge-Kutta (RK) type methods).

In what follows we present the algorithmic schemes for a fixed partition of the set of reactions. It is straightforward to include the time-dependent partitioning.

3.1. The direct hybrid method

The direct method explicitly calculates which reaction occurs next and when it occurs [13, 14]. 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 $\xi \sim \text{Exp}(1)$;
(3) Set $g_\sigma(t|t) = 0$ and solve the system of ODEs starting at time $\tau = t$

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

$$\frac{\mathrm{d}g_\sigma}{\mathrm{d}\tau}(\tau|t) \quad = \quad \sum_{j\in\mathcal{S}} a_j(X(\tau),\tau) \tag{11}$$

Table 1: T7 model equations.

No.	reaction	propensity	rate (day^{-1})	state change
R_1	gen$\xrightarrow{c_1}$ tem	$a_1 = c_1 \cdot$ gen	$c_1 = 0.025$	$\nu_1 = (1,-1,0)$
R_2	tem$\xrightarrow{c_2}$ $\emptyset$	$a_2 = c_2 \cdot$ tem	$c_2 = 0.25$	$\nu_2 = (-1,0,0)$
R_3	tem$\xrightarrow{c_3}$ tem$+$ gen	$a_3 = c_3 \cdot$ tem	$c_3 = 1.0$	$\nu_3 = (0,1,0)$
R_4	gen$+$struc$\xrightarrow{c_4}$ "virus"	$a_4 = c_4 \cdot$ gen $\cdot$ struc	$c_4 = 7.5 \cdot 10^{-6}$	$\nu_4 = (0,-1,-1)$
R_5	tem$\xrightarrow{c_5}$ tem$+$struc	$a_5 = c_5 \cdot$ tem	$c_5 = 1000$	$\nu_5 = (0,0,1)$
R_6	struc$\xrightarrow{c_6}$ $\emptyset$	$a_6 = c_6 \cdot$ struc	$c_6 = 1.99$	$\nu_6 = (0,0,-1)$

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.

3.2. **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 [13, 14]. 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 [12]. 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 $\xi_j \sim \text{Exp}(1)$;

(3) Set $g_j(t|t) = 0$ for $j \in \mathcal{S}$;

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

$$\frac{\mathrm{d}X}{\mathrm{d}\tau}(\tau) = \sum_{j \in \mathcal{D}} \nu_j a_j(X(\tau),\tau) \tag{12}$$

$$\frac{\mathrm{d}g_j}{\mathrm{d}\tau}(\tau|t) = a_j(X(\tau),\tau); \qquad j \in \mathcal{S} \tag{13}$$

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 $\xi_m \sim \text{Exp}(1)$ 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 (12) and (13); 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 $\xi_m \sim \text{Exp}(1)$; 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.

4. Numerical studies for the bacteriophage T7 model

We now wish to demonstrate the power of our hybrid method using a model derived by Srivastava et al. [34] 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.
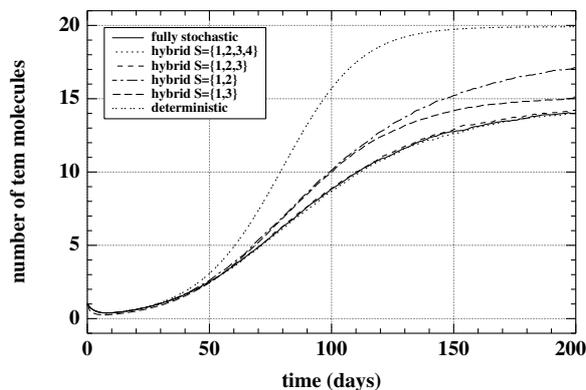
Figure 1: Comparison of the deterministic model with the means of the fully stochastic and different hybrid models (based on 10^4 realizations).

As was shown in [34], 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:

$$a_1 = 5 \quad a_2 = 5 \quad a_3 = 20$$
$$a_4 = 15 \quad a_5 = 20000 \quad a_6 = 19900 \tag{14}$$

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 [34], 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 [13, 14] or the next reaction method [12]. 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 [8, 19, 20]. 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 (11) or (13) can be accomplished by a few Newton iterations [9]. The results obtained are very similar suggesting that the event detection is performed
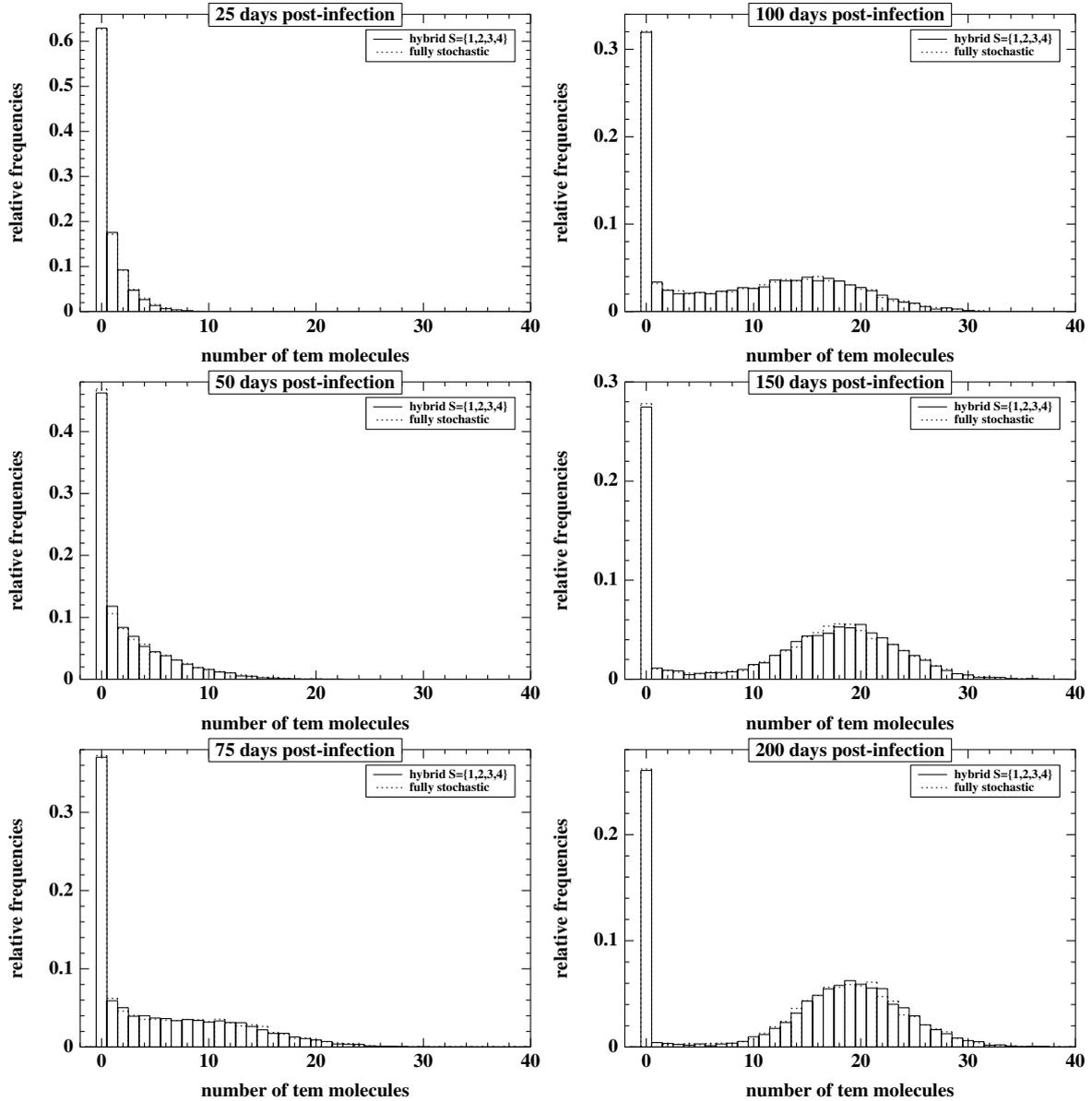
Figure 2: Hybrid kinetics for the bacteriophage T7 model (reactions R_1, R_2, R_3 and $R4$ modelled stochastically, reactions R_5, R_6 modelled deterministically) compared to the reference fully stochastic model: post-infection distribution of tem molecules (based on 10^4 realizations).

with sufficient accuracy.

Further speed up can be obtained by adaptively choosing the partition of the reactions into those modelled stochastically and those approximated deterministically. We based our partitioning on the expected next reaction time and thus on size of the propensity function (see Adaptive partitioning, p. 4). No restriction on the number of reactant molecules was imposed. The approximation quality is as excellent as for the fixed

Table 2: Computational savings. 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).

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

partitioning into $\mathcal{S} = \{1, 2, 3, 4\}$ and $\mathcal{D} = \{5, 6\}$ (cp. Fig. 2), while further saving CPU time. In our model, a restriction on the number of molecules would drastically slow down the simulation, since, as a result, reaction R_5 would be simulated stochastically. However, due to its catalytic nature, approximating reaction R_5 by the deterministic model is no problem. While this seems obvious for the T7 model, less obvious situations exist, where it is sufficient to base the partitioning on the size of the propensity only. On the other side, there are also models, where the partitioning cannot be determined on the basis of the propensity only. Regarding the criteria for adaptive partitioning, further investigation should be performed.

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 [5]). 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.

5. DISCUSSION

We presented a mathematical derivation for hybrid stochastic and deterministic modelling and three easy to implement simulation algorithms. The power and applicability of the hybrid modelling approach has been demonstrated on a model for bacteriophage T7 model [34] 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. (3)) counting the number of times reaction R_j occurs. In the case of tau-leaping methods [4, 5, 17, 32] $N_j(t)$ is approximated by a Poisson process, while in the case of [30] the quasi-steady state approximation is applied to those $N_j(t)$ that correspond to fast reactions. Very recently, two algorithms to simulate hybrid models have been proposed [22, 37]. The main difference to our approach is that they lack a mathematical justification and that the coupling is not statistically exact. 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 (10) and (11), or (12) and (13) of our hybrid methods.

The first promising results of our hybrid approach indicate some further directions of research. 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 the fully stochastic one is needed in order to better understand, under what circumstances the hybrid approach gives reliable results. Moreover, an application of the hybrid approach to a larger realistic biological systems is desirable. Only after our work on the hybrid method has been completed (cmp. [1]), we got aware of the paper by Salis and
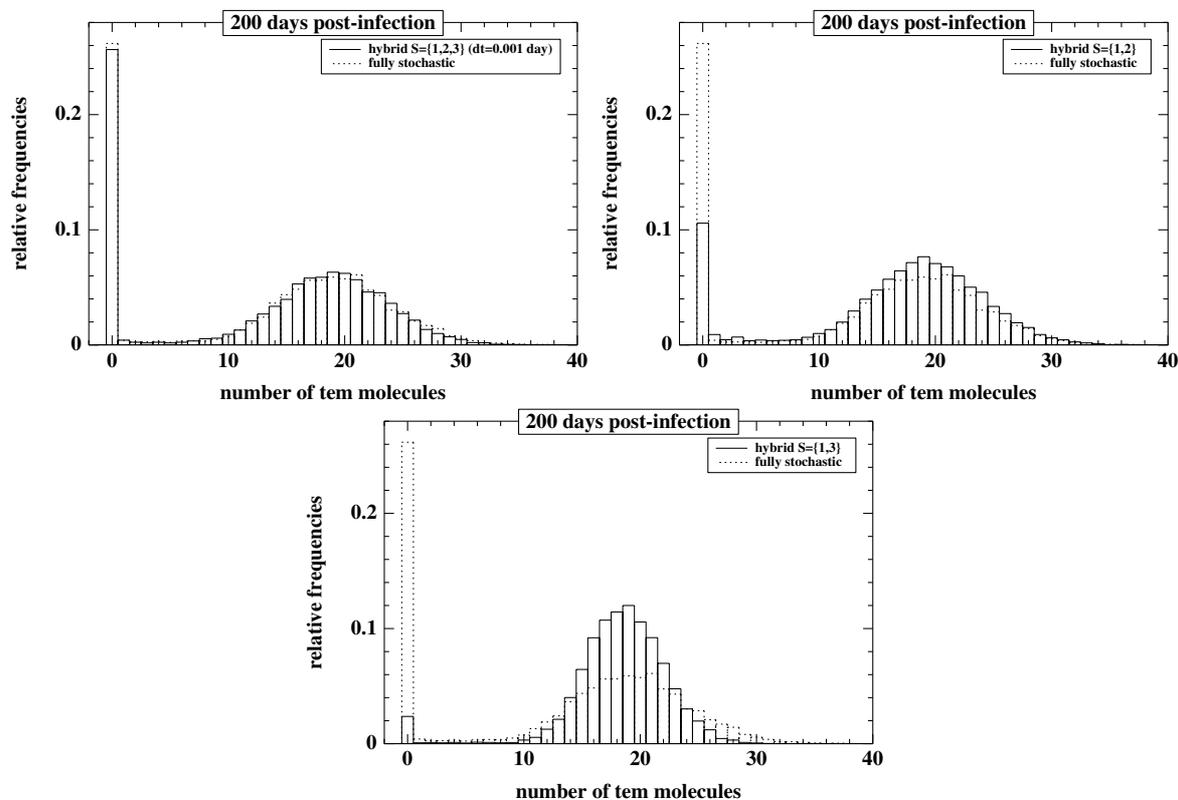
Figure 3: Results for different hybrid systems for the bacteriophage T7 model (solid lines) compared to the reference fully stochastic system (dotted line) based on 10^4 realizations.

Kaznessis [33], which also addresses the problem of coupling Markov jump processes with continuous stochastic or deterministic processes with similar mathematical techniques.

ACKNOWLEDGEMENTS

REFERENCES

[1] Aurélien Alfonsi, Eric Cancès, Gabriel Turinici, Barbara Di Ventura, and Wilhelm Huisinga. Exact simulation of hybrid stochastic and deterministic models for biochemical systems. Rr-5435, INRIA-Rocquencourt, Dec 2004.

[2] 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.

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

[4] 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.

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

[6] Yang Cao, Daniel T. Gillespie, and Linda R. Petzold. The slow-scale stochastic simulation algorithms. *J chem Phys*, 122, 2005.

[7] D. J. Daley and D. Vere-Jones. *An Introduction to the Theory of Point Processes*, volume I: Elementary Theory and Methods of *Probability and its Applications*. Springer, 2nd edition, 2003.

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

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

[10] 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.

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

[12] 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.

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

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

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

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

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

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

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

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

[21] 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.

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

[23] 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.

[24] Thomas G. Kurtz. Limit theorems for sequences of jump markov porcesses approximating ordinary differential equations. *J Appl Prob*, 8:344–356, 1971.

[25] Thomas G. Kurtz. The relationship between stochastic and determinsitc models of chemical reactions. *J Chem Phys*, 57:2976–2978, 1973.

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

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

[28] 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.

[29] 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.

[30] 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.

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

[32] 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.

[33] Howard Salis and Yiannis Kaznessis. Accurate hybrid simulation of a system of coupled chemical or biochemical reactions. *J chem Phys*, 122, 2005.

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

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

[36] 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.

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

[38] Petar Todorovic. *An Introduction to Stochastic Processes and Their Applications*. Springer Series in Statitics. Springer, 1992.

[39] T. E. Turner, S. Schnell, and K. Burrage. Stochastic approaches for modelling in vivo reactions. *Comp Bio Chem*, 28:165–178, 2004.
[40] T. Wilhelm, J. Behre, and S. Schuster. Analysis of structural robustness of metabolic networks. *Systems biology*, 1:114–120, 2004.