Research article

Hybrid stochastic simplifications for multiscale gene networks

Alina Crudu[1], Arnaud Debussche[2,3] and Ovidiu Radulescu*[1,3]

Address: [1]IRMAR UMR CNRS 6625 Université de Rennes1, Campus de Beaulieu, 35042 Rennes, France, [2]ENS Cachan, Antenne de Bretagne Avenue Robert Schumann, 35170, BRUZ, France and [3]INRIA-IRISA, Campus de Beaulieu, 35042 Rennes, France

Email: Alina Crudu - alina.crudu@univ-rennes1.fr; Arnaud Debussche - arnaud.debussche@bretagne.ens-cachan.fr; Ovidiu Radulescu* - ovidiu.radulescu@inria.fr

* Corresponding author

Abstract

Background: Stochastic simulation of gene networks by Markov processes has important applications in molecular biology. The complexity of exact simulation algorithms scales with the number of discrete jumps to be performed. Approximate schemes reduce the computational time by reducing the number of simulated discrete events. Also, answering important questions about the relation between network topology and intrinsic noise generation and propagation should be based on general mathematical results. These general results are difficult to obtain for exact models.

Results: We propose a unified framework for hybrid simplifications of Markov models of multiscale stochastic gene networks dynamics. We discuss several possible hybrid simplifications, and provide algorithms to obtain them from pure jump processes. In hybrid simplifications, some components are discrete and evolve by jumps, while other components are continuous. Hybrid simplifications are obtained by partial Kramers-Moyal expansion [1-3] which is equivalent to the application of the central limit theorem to a sub-model. By averaging and variable aggregation we drastically reduce simulation time and eliminate non-critical reactions. Hybrid and averaged simplifications can be used for more effective simulation algorithms and for obtaining general design principles relating noise to topology and time scales. The simplified models reproduce with good accuracy the stochastic properties of the gene networks, including waiting times in intermittence phenomena, fluctuation amplitudes and stationary distributions. The methods are illustrated on several gene network examples.

Conclusion: Hybrid simplifications can be used for onion-like (multi-layered) approaches to multi-scale biochemical systems, in which various descriptions are used at various scales. Sets of discrete and continuous variables are treated with different methods and are coupled together in a physically justified approach.

Background

At a molecular level, the functioning of cellular processes is unavoidably stochastic. Recent advances in real-time single cell imaging, micro-fluidic manipulation and synthetic biology have shown that gene expression and protein abundance in single cells are dynamic random variables submitted to significant variability [4-7].

Markov processes approaches to gene networks dynamics, originating from the pioneering ideas of Delbrück [8], capture diverse features of the experimentally observed variability, such as transcription bursts [4,6,7,9], various types of steady-state distributions of RNA and protein numbers [10,11], noise amplification or reduction [12,13]. In the case of molecular switches, random transitions between two or several metastable states with different transcriptional activities, limit the possibility of the corresponding genetic circuits to store cellular memory and generate variability [6].

Even the simplest stochastic model, such as the production module of a protein [11,14] contains tens of variables and biochemical reactions, and at least the same number of parameters. Direct simulation of such networks by Stochastic Simulation Algorithm (SSA) [15] is extremely time consuming. Network reverse engineering, model parameter identification, and design principles studies suffer from the same drawbacks when the full models are attacked with traditional tools. Various approximation methods were proposed, based on time discretization such as binomial, Poisson or Gaussian time leaping [16,17], or based on fast/slow partitions [18-21].

The time complexity of exact simulation algorithms scales with the number of jumps (reactions) per unit time. Currently available hybrid algorithms treat fast reactions as continuous variables, which significantly reduces the number of jumps [18-21]. Although computationally efficient, the use of slow reactions as discrete variables, and of rapid reactions as continuous variables, does not always provide a convenient description of the hybrid stochastic process. Indeed, the discrete/continuous structure of the state space of a hybrid molecular system is better expressed in terms of species components. Activity of some promoters and production of the corresponding proteins occurs in bursts [6,7]. In a bursting event, a steep increase of the number of transcripts is followed by relatively smoother exponential degradation. In this example, the natural candidates for continuous variables are the gene transcripts and products concentrations, while the states of the DNA promoters are the discrete variables. This picture also emphasizes the possibility for noise transfer from discrete, microscopic components, to continuous, mesoscopic, and closer to the phenotype, variables. In such processes, both microscopic and mesoscopic variables are noisy. At the mesoscopic scale, the stochastic fluctuations result from the intermittency of the dynamics and do not have a simple dependence on the numbers of molecules of the continuous components. Counter-intuitively, the noise amplitude could increase with the number of transcripts or proteins in a burst [6].

Stochastic intermittence or bursting occurs not only in transcription, but also in many other biochemical processes, such as action potential firing [22], blips in calcium signaling [23], possibly happloinsufficiency [24], etc.

Thus, a large class of molecular stochastic models can be approximated by stochastic hybrid processes in which some state components are discrete, while others are continuous. Processes of this kind have been intensively studied in relation to applications in technology (air traffic management, flexible manufacturing systems, fault tolerant control systems, storage models, etc.) [25-28]. Several classes of stochastic hybrid processes were studied, among which the most important are the switching diffusion processes and the piecewise deterministic processes. For switching diffusions, the continuous state component has continuous trajectories governed by stochastic differential equations, punctuated by jumps controlled by the discrete component. For piecewise deterministic processes, the continuous component follows deterministic ordinary differential equations between two jumps. The jumps can be changes of the parameters in the (stochastic) differential equations (switch of regime with no discontinuity in the trajectory), or instantaneous changes of the continuous variables (discontinuities of the trajectory), or both. Piecewise deterministic processes have been recently proposed as models for various biological systems [29,30]. Diffusion approximations (without jumps) of Markov processes justify approximate versions of the Gillespie algorithm, such as the τ-leap method [31]. A general approach allowing to obtain and classify various hybrid approximations that can be obtained from molecular Markov jump processes is still needed and represents a first result of this paper. Hybrid stochastic processes can be obtained by applying the law of large numbers (or the central limit theorem) to the continuous components. This will be performed here by using a partial Kramers-Moyal expansion of the chemical master equation. In many cases, this procedure provides only a moderate reduction of the number of stochastic jumps to be simulated. Indeed, remaining discrete components can jump rapidly, in which case the simulation by a hybrid process performs ineffectively.

The second result of this paper allows us to cope with frequent jumps of the discrete components by using averaging. Averaging techniques are widely used in physics and chemistry to simplify models by eliminating fast microscopic variables [32-34]. In our case, the microscopic variables are the discrete state components (for instance species present in low numbers) and the resulting approximations are averaged switched diffusion or averaged (piecewise) deterministic processes. Standard averaging techniques for discrete Markov chains [35] use state aggregation. However, state aggregation algorithms are effec-

tive only for Markov chains with a finite, small number of states. A more effective averaging method, that we present here, is cycle averaging in chemical reaction networks. Our method exploits the formal analogy between the quasi-stationarity assumption for fast deterministic linear cycles [36] and the stochastic averaging assumption for the same type of submodels.

Stochastic quasi-steady-state approximations presented in [37] combine singular perturbations for the master equations and Van Kampen's Ω-expansion [38] (first order Kramers-Moyal expansion) and are very close to our approach. However, we provide here a much more general setting and completely characterize the hybrid limits that result from this setting.

The stochastic dynamics of our simplified hybrid models provide good approximations of the un-reduced chemical master equation. We show rigorously elsewhere [39] that the simplified hybrid processes are weak approximations of the un-reduced Markov process. This means that all the statistical properties of the trajectories are approximated in distribution. In particular, steady state distributions of molecular species concentrations should be accurately reproduced. Also, distributions of intermittence times like the times between bursts in the production of mRNA or proteins, or the commutation time for stochastic switches are expected to be similar for the reduced model and for the initial model.

The structure of this paper is the following. In the Methods section we obtain hybrid simplifications for pure jump Markov processes and also propose a new averaging algorithm for these models. In the "Results and Discussion" section we present several examples of hybrid models obtained from stochastic chemical kinetics models, including a medium scale protein production model for which we demonstrate averaging.

Methods
Introductory notes on stochastic hybrid systems
Stochastic hybrid processes are couples of the type $(\theta(t), x(t))$ where $\theta(t)$ is a discrete component (piecewise constant), $x(t)$ is a vector with piecewise continuous components, and t is time. The discrete component $\theta(t)$ can be considered as a controlled Markov chain, whose transition matrix may depend on the continuous variable $x(t)$. The discrete state space Θ (values of θ) can be finite or infinite. For instance, supposing that the discrete variables are r "rare" molecular species, whose numbers never go over a small value N, then $\Theta \subset \{0,..., N\}^r$ has at most $(N + 1)^r$ states. Different values of θ may also correspond to various states of a single molecule. For proteins with multiple phosphorylation sites the number of states is 2^r where r is the number of sites. In the case of competitive transcrip-

tional regulation, when N transcription factors are competing on r binding sites, the number of states is at most $(N + 1)^r$.

There are several possible ways to specify the transitions of the discrete variables. Generally, these can be given by the stochastic matrix $\lambda_{i,j}(x, t)$ where $\lambda_{i,j} dt + o(dt)$ is the probability of a transition from the state i to the state j, between t and $t + dt$. However, this representation is not handy when the matrices have very large dimension. In such cases, the usual molecular Markov jump process description is more appropriate. Then, possible transitions are specified by the stoichiometry vectors γ_i. The intensity (propensity) function $\lambda_i (\theta, x)$ represents the probability per unit time that a transition from θ to $\theta + \gamma_i$ takes place between t and $t + dt$ [26]. It follows that the probability for a transition (of any kind) between t and $t + dt$ is $\lambda(\theta, x) dt + o(dt)$, where $\lambda = \sum_i \lambda_i$ is the total intensity. The inverse of the total intensity represents the mean waiting time between two successive transitions. The transition $\theta \to \theta + \lambda_i$ is chosen with probability λ_i/λ.

Jumps can also occur in the continuous variables, in which case the size of the jump can be continuously distributed.

There are several possibilities for the evolution of the continuous variables leading to various classes of hybrid processes.

Piecewise deterministic processes
Piecewise deterministic processes were first formalized by [40] and found many applications in various areas ranging from industry to mathematical finance and biology.

The state of a piecewise deterministic process (PDP) is a couple $\zeta = (\theta, x)$ where $\theta \in \Theta$ denotes the discrete variable and $x \in \mathbb{R}^n$ denotes the continuous variable. A PDP is given by three local characteristics namely:

- a vector field $\chi_\theta(x)$ (flow function),

- a jump intensity $\lambda(\theta, x)$ and

- a transition measure Q^T such that $Q^T(\theta', x'; \theta, x)$ is the probability distribution of the post-jump positions $(\theta', x') \in \Theta \times \mathbb{R}^n$. Related to this is the jump distribution Q^J giving the probability distribution of the jumps: $Q^J(\delta\theta, \delta x; \theta, x) = Q^T(\theta + \delta\theta, x + \delta x; \theta, x)$ (Q^J is obtained from Q^T by phase space translation).

Such a process involves deterministic motion punctuated by jumps. The lengths of the time intervals between successive jumps are random variables called waiting times τ_i. On the time interval $[t_i, t_{i+1} = t_i + \tau_i]$ the discrete variable

remains constant $\theta = \theta_i$, while the continuous variable evolves according to the differential equation:

$$\begin{cases} dx/dt = \chi_{\theta_i}(x), \\ x(t_i) = x_i \end{cases} \qquad (1)$$

Let $\varphi(t, (x_i, \theta_i))$ be the solution of the differential equation (1). We suppose that this solution exists and is unique for all $t \geq t_i$. The waiting time is a random variable whose distribution reads [40]:

$$P[\tau_i > t] = F(\theta_i, x_i) = \exp[-\int_0^t \lambda(\theta_i, \phi(s; x_i, \theta_i))ds]. \qquad (2)$$

The reader can easily obtain Eq.(2). The probability not to jump in the time interval $[s, s + ds]$ is $1 - \lambda ds + o(ds)$. Thus $P[\tau_i > s + ds] = P[\tau_i > s](1 - \lambda ds) + o(ds)$. By taking the logarithm we get $\log P[\tau_i > s + ds] - \log P[\tau_i > s] = -\lambda ds + o(ds)$. Summing over ds leads to Eq.(2). The more familiar expression used in the Gillespie algorithm $P[\tau_i > t] = \exp[-\lambda(\theta_i) t]$ corresponds to the particular case when λ does not depend on the continuous variable.

After a deterministic evolution on the $(i + 1)$-th interval, the discrete variable and the continuous variable can change according to:

$$\theta_i \rightarrow \theta_i + \gamma_i^D, x_i \rightarrow x_i + \gamma_i^C,$$

where γ_i^D, γ_i^C are sampled from the jump distribution Q^J.

The discrete variables are parameters of the differential equations describing the dynamics of the continuous variables. Notice that if $\gamma_i^C = 0$ the only effect of a jump on the continuous variables is a change of regime (change of parameter in the differential equation). The trajectories of the continuous variables are globally continuous. However, the velocities of the continuous variables are discontinuous at jumps of the discrete variables. We call this possibility "switching". If $\gamma_i^C \neq 0$, then continuous variables can have instantaneous jumps and their trajectories are only piecewise continuous. We call this possibility "breakage". It is possible to have both "switching" and "breakage".

A PDP can be also defined by its generator $\mathcal{A}$ [40,41], acting on functions f defined on the phase space:

$$\mathcal{A}f(\theta, x) = \chi_\theta(x).\nabla_x f + \lambda(x, \theta)\int [f(\theta + \gamma^D, x + \gamma^C) - f(\theta, x)]Q^J(d_\gamma^D, d_\gamma^C; \theta, \boldsymbol{x}) \qquad (3)$$

The adjoint of the generator is used to obtain the Fokker-Planck equation [3,38,42] describing the time evolution of the probability distribution $p(\theta, x, t)$ of the process:

$$\frac{\partial p(\theta, x, t)}{\partial t} = \mathcal{A}^* p(\theta, x, t) = -\nabla_x \cdot [\chi_\theta(x)p] - \lambda(\theta, x)p(\theta, x, t) +$$
$$+ \int \lambda(\theta - \gamma^D, x - \gamma^C)p(\theta - \gamma^D, x - \gamma^C, t)Q^t(d\gamma^D, d\gamma^C; \theta - \theta^C, x - \gamma^C) \qquad (4)$$

The structure of the generic PDP corresponds to the following simulation algorithm:

(1) Set the initial state condition $x(t_0) = x_0$, $\theta = \theta_0$,

(2) Generate a random variable u uniformly distributed in $[0, 1]$,

(3) Integrate the system of differential equations

$$\begin{cases} dx/dt = \chi_\theta(x(t)), \\ dF/dt = -\lambda(\theta, x)F, \\ x(t_i) = x_i, F(t_i) = 1, \end{cases}$$

between t_i and $t_i + \tau_i$ with the stopping condition $F(t_i + \tau_i, \theta_i, x_i) = u$,

(4) Generate a second uniform variable v and use it to randomly choose the jumps (γ_i^C, γ_i^D) (the decision is made in the same way as in the Gillespie algorithm [15]),

(5) Change the system state $(\theta_i, x(t_i))$ into $(\theta_i + \gamma_i^D, x(t_i) + \gamma_i^C)$,

(6) Reiterate the system from 2) with the new state until a time t_{max} previously defined is reached.

Hybrid diffusions
A rather general class of hybrid diffusions has been proposed in [26]. However, in that setting, jumps of continuous variables are commanded by the hitting of some predefined phase space sets, which is not our case. We propose a simpler and different setting, which is natural for biochemical systems. In our setting, a hybrid diffusion is defined by:

- a vector field $\chi(\theta, x)$ (drift or flow function),

- a diffusion matrix $\sigma(\theta, x)$,

• a jump intensity and a transition measure defined like for PDPs.

Between two consecutive jumps at t_i and $t_i + \tau_i$, the continuous variables $x(t)$ satisfy the Itô stochastic differential equations [43]:

$$dx_k(t) = \chi_k(\theta, x)dt + \sum_j \sigma_{kj}(\theta, x)dW_j(t)$$

where $W_j(t)$ are independent one-dimensional Wiener processes.

The waiting time τ is determined like for PDPs. One should integrate the system:

$$\begin{cases} dx_k(t) = \chi_k(\theta, x)dt + \sum_j \sigma_{kj}(\theta, x)dW_j(t) \\ dF = -\lambda(\theta, x)F dt \end{cases}$$

with the initial condition $F(t_i) = 1$, $x(t_i) = x_i$ and the stopping condition $F(t_i + \tau_i) = u$, where u is a random variable, independent from (θ_i, x_i), uniformly distributed on [0, 1].

The generator for such a process is:

$$\mathcal{A}f(\theta, x) = \chi_\theta(x).\nabla_x f + \frac{1}{2}\sigma(\theta, x) : \nabla_x^2 f + \lambda(\theta, x)\int [f(\theta + \gamma^D, x + \gamma^C) - f(\theta, x)]Q(d\gamma^D, d\gamma^C; \theta, x) \quad (5)$$

where "." stands for the scalar product and ":" stands for the double contracted tensor product, i.e.

$$\sigma : \nabla_x^2 f = \sum_{ij}\sigma_{ij}\frac{\partial^2 f}{\partial x_i \partial x_j}.$$

Notice that hybrid diffusions contain PDPs as the particular case when diffusion is zero ($\sigma = 0$).

Hybrid simplification of Markov pure jump processes (MPJP)

Markovian stochastic chemical systems

Most of the existing numerical methods for stochastic chemical systems are based on the representation of the chemical reaction system as a Markov pure jump process (MPJP), that corresponds to Gillespie's [15,44] stochastic simulation algorithm (SSA). The state of a system of n chemical species $A_1,..., A_n$ is a vector $X(t)$ whose components are the numbers of molecules from each species. The state space of the process is $E = {}^n$. From now on, we shall use upper case letters X for numbers of molecules, and lower case letters $x = X/\mathcal{V}$ for concentrations, where $\mathcal{V}$ is the reaction volume, typically the volume of the cell's compartment.

For each reaction,

$$\alpha_{i1}A_1 + ... + \alpha_{in}A_n \rightleftharpoons \beta_{i1}A_1 + ... + \beta_{in}A_n, \quad (6)$$
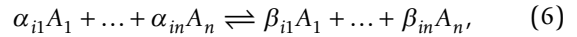
we have the jump $\gamma_i = \beta_i - \alpha_i \in {}^n$, $i = 1,..., n_r$ (n_r is the number of reactions). We consider equally the reverse reaction, corresponding to the jump $-\gamma_i$. The intensities and the transition measures for the Markov jump process are:

$$\lambda(X) = \sum_{i=1}^{n_r}[V_i(X) + V_{-i}(X)] \quad (7)$$

$$Q^T(X', X) = \sum_{i=1}^{n_r}[q_i(X)\delta_{X+\gamma_i}(X') + q_{-i}(X)\delta_{X-\gamma_i}(X')] \quad (8)$$

V_i, V_{-i} are the reaction rates and probabilities $q_i = V_i / \sum_{j=1}^{n_r}(V_j + V_{-j})$ satisfy $\sum_{j=1}^{n_r}(q_{-i} + q_i) = 1$. Here δ_X is the Dirac measure centered on X, and X' is the post-jump position.

Several choices are possible for the reaction rates [45]. The most popular is the mass action law, when rates are polynomial functions of concentrations:

$$V_i(X) = \mathcal{V}v_i(x) = \mathcal{V}k_i\prod_{s=1}^n x_s^{\alpha_{is}}, \quad V_{-i}(X) = \mathcal{V}v_{-i}(x) = \mathcal{V}k_{-i}\prod_{s=1}^n x_s^{\beta_{is}} \quad (9)$$

As for any homogeneous Markov process, we can characterize this process by its generator [41], whose action on test functions f defined on the state space E is:

$$\mathcal{A}f(X) = \sum_{\pm i\in[1,n_i]}[f(X + \gamma_i) - f(X)]V_i(X) \quad (10)$$

The adjoint of the generator defines the chemical master equation which describes the time evolution of the probability distribution $p(X, t)$ in phase space.

$$\frac{\partial p}{\partial t}(X,t) = \mathcal{A}^*p(X,t) = -p(X,t)\lambda(X) + \sum_{\pm i\in[1,n_r]}V_i(X - \gamma_i)p(X - \gamma_i, t) \quad (11)$$

Notice that hybrid diffusions or PDPs with no continuous components are Markov pure jump processes.

Partial Kramers-Moyal expansion

Diffusion approximations (Langevin dynamics) for stochastic chemical kinetics was proposed by [31] and was

rigorously justified by [46,47]. This approximation is valid when the numbers of molecules of all species are much larger than one. It concerns for instance the thermodynamic limit $\mathcal{V} \to \infty$ where $\mathcal{V}$ is the volume of a well stirred reactor. Then, the diffusion approximation applies to intermediate scales $1/\sqrt{\mathcal{V}}$ in phase space. Formally, this approximation can be obtained by the Kramers-Moyal expansion [1-3], which is the second order Taylor expansion of the master equation with respect to jumps divided by the volume; in the first order one gets the deterministic dynamics, and in the second order the Langevin dynamics. In [29], we have used partial Kramers-Moyal expansions to obtain hybrid approximations of MPJP. The possibilities of this method are examined here in more generality.

First, let us partition the species of the system into two sets $X = (X_D, X_C)$, where X_D represents the species in small numbers and X_C the species in large numbers. Note that this partition is different from reaction based partitions used in other works [18,20,21,48-50]. All these works have as objective the reduction of simulation time. With respect to this objective, many schemes that avoid individual simulation of rapid reactions by regrouping them and by treating the corresponding channels as continuous variables in "fluid" limits, are more or less equivalent. We fix ourself also another objective, which is to find a natural representation of the simplified processes. A species based partition is more suitable for this end. Let M be the set of reactions in which every reversible reaction contributes with two reactions, one for each direction. The species partition determines a partition of M into three sets $M = \mathcal{R}_D \bigcup \mathcal{R}_{DC} \bigcup \mathcal{R}_C$. The reaction partition distinguishes between the reactions whose reactants and products are in X_C ($\mathcal{R}_C$), reactions whose reactants and products are in X_D ($\mathcal{R}_D$) and reactions with at least a reactant or a product in X_D and at least a reactant or product in X_C ($\mathcal{R}_{DC}$). We should emphasize that although $\mathcal{R}_C$ reactions are rapid, some other rapid reactions can be contained in $\mathcal{R}_D$ or $\mathcal{R}_{DC}$.

The second step towards the hybrid processes is to rescale the continuous variables by dividing them with the volume and obtaining $x_C = X_C / \mathcal{V}$. The X_D variables will not be rescaled and will be considered discrete. Their evolution is given by discrete transitions.

The third step is to consider a second order Taylor expansion of the master equation, for the continuous variables x_c and with respect to the small parameters $\gamma_i / \mathcal{V}$. This expansion can be performed equivalently on the generator (10) or on the master equation (11). For simplicity, we use the generator. The test functions depend now on two variables (x_C, X_D) and are considered to be twice differentiable with respect to the continuous variables x_C.

Let us consider that rates of reactions in $\mathcal{R}_C$ are proportional to the volume (this follows from mass action law), $V_i = \mathcal{V} v_i(x_C)$, $i \in \mathcal{R}_C$. The second order Taylor expansion of (10) with respect to x_C reads:

$$\mathcal{A}^{(2)} f(X_D, x_C) = \sum_{i \in \mathcal{R}_D} [f(X_D + \gamma_i^D, x_C) - f(X_D, x_C)] V_i(X_D) +$$

$$\sum_{i \in \mathcal{R}_{DC}} [f(X_D + \gamma_i^D, x_C + \gamma_i^C / \mathcal{V}) - f(X_D, x_C)] V_i(X_D, x_C) +$$

$$[\sum_{i \in \mathcal{R}_C} \gamma_i^C v_i(x_C)] \cdot \nabla_{x_C} f(X_D, x_C) +$$

$$\frac{1}{2\mathcal{V}} [\sum_{i \in \mathcal{R}_C} v_i(x_C) \gamma_i^C \otimes \gamma_i^C] : \nabla_{x_C, x_C}^2 f(X_D, x_C)$$

$$(12)$$

where $\otimes$ is for tensor product $((\gamma \otimes \gamma)_{kl} = \gamma_k \gamma_l)$.

The approximated generator corresponds to hybrid diffusions with drift function $\chi(x_C) = \sum_{i \in \mathcal{R}_C} \gamma_i^C v_i(x_C)$ and diffusion matrix $\sigma(x_C) = \frac{1}{\mathcal{V}} \sum_{i \in \mathcal{R}_C} v_i(x_C) \gamma_i^C \otimes \gamma_i^C$. As it can be easily seen, the drift and diffusion coefficients for the variables x_C do not depend on the discrete variables X_D. Thus, switching is not present in this approximation and discrete variables can be just forgotten if not observed. Furthermore, jumps $\gamma_i^C / \mathcal{V}$ vanish in the limit $\mathcal{V} \to \infty$, implying that continuous variables have no instantaneous jumps (no breakage). However, switching and breakage is present in many examples in molecular biology. How is this possible?

In order to cope with the possibility of switching and breakage we need to consider that some reactions from $\mathcal{R}_{DC}$ exhibit particular properties. We can consider two types of special reactions:

1. reactions in $\mathcal{R}_{DC}$ that are frequent enough to change the continuous variables and induce switching,

2. reactions in $\mathcal{R}_{DC}$ with a large stoichiometric vector, that induce breakage.

These reactions will be referred to as "super-reactions". The first type of super reactions are denoted by the set $\mathcal{S}_1 \subset \mathcal{R}_{DC}$ and the second type of super-reactions are denoted $\mathcal{S}_2 \subset \mathcal{R}_{DC}$. As a consequence, we distinguish two types of hybrid approximations following which one of the two "super-reactions" are present in the mechanism.

A. Hybrid approximations with switching

This first case of approximate process is induced by "super-reactions" $\mathcal{S}_1 \subset \mathcal{R}_{DC}$ with stoichiometric vectors of order one, but with rates that scale with the volume: $V_i = \mathcal{V}v_i(X_D, x_C)$, $i \in \mathcal{S}_1$. Reactions from $\mathcal{S}_1$ induce rapid transitions of the continuous variables. The set $\mathcal{S}_2$ is considered to be empty. For this approximation, even if the rates of the "super-reactions" depend on the discrete variables, we suppose that the "super-reactions" do not change the discrete variables. In other words

$$\gamma_i^D = 0, \text{ for all } i \in \mathcal{S}_1. \tag{13}$$

If this condition fails, another type of approximation arises: some discrete variables change rapidly. The resulting process will be an averaged hybrid process.

The evolution of the discrete variables is given by slow jumps (reactions from $\mathcal{R}_D$). The component x_C has rapid jumps but their amplitude is relatively small. Thus, between two consecutive transitions of the discrete variables, the continuous variables can be approximated by a smooth deterministic trajectory. The continuous variables have continuous trajectory, however, discontinuous velocities. The first order of the Kramers-Moyal expansion corresponds to the *piecewise deterministic approximation* defined by the flow function

$$\chi(X_D, x_C) = \sum_{i \in \mathcal{S}_1} \gamma_i^C v_i(X_D, x_C) + \sum_{i \in \mathcal{R}_C} \gamma_i^C v_i(x_C), \tag{14}$$

the jump intensity

$$\lambda(X_D, x_C) = \sum_{i \in \mathcal{R}_D} V_i(X_D) + \sum_{i \in \mathcal{R}_{DC} \setminus \mathcal{S}_1} V_i(X_D, x_C), \tag{15}$$

and the jump probability

$$P[X_D \to X_D + \gamma_i^D] = \frac{1}{\lambda(X_D, x_C)} \begin{cases} V_i(X_D), \text{ if } i \in \mathcal{R}_D, \\ V_i(X_D, x_C), \text{ if } i \in \mathcal{R}_{DC} \setminus \mathcal{S}_1, \end{cases} \tag{16}$$

where γ_i^D, γ_i^C are the projections of γ_i on continuous and on discrete components, respectively. The relation (15) clearly shows that the jump intensity depends both on continuous and on discrete variables. Nevertheless, only the discrete variables are changed by jumps.

The second order of the Kramers-Moyal expansion corresponds to the *hybrid diffusion approximation* whereas the diffusion matrix reads:

$$\sigma(X_D, x_C) = \frac{1}{\mathcal{V}}[\sum_{i \in \mathcal{S}_1} v_i(X_D, x_C)\gamma_i^C \otimes \gamma_i^C + \sum_{i \in \mathcal{R}_C} v_i(x_C)\gamma_i^C \otimes \gamma_i^C], \tag{17}$$

B. Hybrid approximations with breakage

This type of approximation is induced by the "super-reactions" with a large stoichiometric vector (i.e. reactions $\mathcal{S}_2$). We suppose also, that the set $\mathcal{S}_1$ is empty. In other words, to have this approximation, we suppose that at least one reaction i in the set $\mathcal{R}_{DC}$ has a large stoichiometric vector $(\lambda_i)_j / \mathcal{V}$ is comparable with $|x_j|$ for $i \in \mathcal{S}_2$ and for some species $j \in C$. A reaction $\mathcal{S}_2$ can change both continuous and discrete components. The main difference consists in the fact that a reaction $\mathcal{S}_2$ significantly change, in one step, the concentration of the x_C component.

The reactions of the type $\mathcal{S}_2$ do not contribute to the deterministic flow. They only appear as jumps. As before, we write down the main characteristics of the approximated PDP process (first order Kramers-Moyal expansion), namely the flow function

$$\chi(x_C) = \sum_{i \in \mathcal{R}_C} \gamma_i^C v_i(x_C), \tag{18}$$

the jump intensity

$$\lambda(X_D, x_C) = \sum_{i \in \mathcal{R}_D} V_i(X_D) + \sum_{i \in \mathcal{R}_{DC}} V_i(X_D, x_C),$$

(19)

the jump probabilities for the discrete variables

$$P[X_C \to X_C + \gamma_r^C] = \frac{1}{\lambda(X_D, x_C)} \begin{cases} V_i(X_D), & \text{if } i \in \mathcal{R}_D, \\ V_i(X_D, x_C), & \text{if } i \in \mathcal{R}_{DC} \setminus \mathcal{S}_2, \end{cases}$$

(20)

and the jump probability for the continuous variables

$$P[x_C \to x_C + \gamma_i^C / \mathcal{V}] = \frac{1}{\lambda(X_D, x_C)} V_i(X_D), \text{ for } i \in \mathcal{S}_2.$$

(21)

If the super-reactions $i \in \mathcal{S}_2$ have $\gamma_i^{(D)} \neq 0$, jumps occur simultaneously in the continuous and in the discrete variables.

The deterministic flow is defined by the flow function $\chi(x_C)$ (Eq.(18)) which does not depend on the discrete variable X_D. In this case there is no switching. In the second order of the Kramers-Moyal expansion, we include phase space diffusion, which contains contributions only from reactions of $\mathcal{R}_C$:

$$\sigma(x_C) = \frac{1}{\mathcal{V}} \sum_{i \in \mathcal{R}_C} v_i(X_D, x_C) \gamma_i^C \otimes \gamma_i^C,$$

(22)

If both sets $\mathcal{S}_1$, $\mathcal{S}_2$ are not empty, then both types of discontinuities can appear in the trajectory of x_C:

- switching: discontinuity of velocity (jump of X_D), with no discontinuity of trajectory,

- breakage: discontinuity of trajectory (jump in some variable x_C).

The first discontinuities are induced by reactions from $\mathcal{S}_1$, while the latter are induced by reactions from $\mathcal{S}_2$.

C. Hybrid approximations with singular switching
Although mathematically distinct, breakage and switching could be physically indistinguishable in certain cases. Indeed, the jump of the continuous variable can result from the repeated application of one or several very fast reactions from $\mathcal{S}_1$. We can obtain processes with breakage as singular limits of processes with switching when

very short lasting steep variations alternate with long lasting slower variations of the continuous variables.

Let us consider that there is a subset of extremely rapid super-reactions $\mathcal{S}_3$ such that:

$$V_i = \begin{cases} (\mathcal{V} / \varepsilon) v_i(X_D, x), & \text{if } X_{Di} \neq 0 \\ 0, & \text{if } X_{Di} = 0 \end{cases}$$

where $0 < \ll 1$ is a small positive parameter, and where X_{D_i} are discrete species, substrates of the reaction $i \in \mathcal{S}_3$.

Like for reactions $\mathcal{S}_1$, we consider that $\gamma_i^D = 0$, for all $i \in \mathcal{S}_3$.

For each reaction $i \in \mathcal{S}_3$ we consider that there are two subsets of reactions in $\mathcal{R}_D$: $\mathcal{R}_{D_i}^+$ contains reactions that produce the species X_{D_i} and $\mathcal{R}_{D_i}^-$ contains reactions that consume the species X_{D_i}. We define $\mathcal{R}_D^- = \cup_i \mathcal{R}_{D_i}^-, \mathcal{R}_D^+ = \cup_i \mathcal{R}_{D_i}^+$. We consider that all reactions in $\mathcal{R}_D^-$ are rapid:

$$V_j = \frac{1}{\varepsilon} v_j(X_D), \text{ for all } j \in \mathcal{R}_D^-$$

First, we apply the Kramers-Moyal expansion for large $\mathcal{V}$ and obtain a switched PDP. The flow function of this PDP reads:

$$\chi(X_D, x) = \chi_0(x) + \frac{1}{\varepsilon} \chi_1(X_D, x)$$

where

$$\chi_0(x) = \sum_{i \in \mathcal{R}_C} \gamma_i^C v_i(x), \chi_1(X_D, x) = \sum_{i \in \mathcal{S}_3} \gamma_i^C v_i(X_D, x).$$

The jump intensity of the PDP reads

$$\lambda(X_D, x) = \sum_{i \in \mathcal{R}_D \setminus \mathcal{R}_D^-} V_i(X_D) + \frac{1}{\varepsilon} \sum_{j \in \mathcal{R}_D^-} v_j(X_D).$$

We show in the Additional File 1 that, in the limit $\to 0$, the switched PDP converges to a PDP having the following generator:

$$\mathscr{A}f(X_D, x) = \chi_0(x).\nabla_x f + \sum_{i \in \mathscr{R}_D \setminus (\mathscr{R}_D^- \cup \mathscr{R}_D^+)} V_i(X_D)[f(X_D + \gamma_i, x) - f(X_D, x)] +$$

$$+ \sum_{i \in \mathscr{R}_D^-} V_i(X_D) \sum_{j \in \mathscr{R}_D^-} \int_0^\infty [f(X_D + \gamma_i^D + \gamma_j^D, \Phi(s; x, X_D + \gamma_i^D)) - f(X_D, x)] \rho_{ij}(s) ds$$

$$(23)$$

where $\rho_{ij}(s) = v_j(X_D + \gamma_i^D) \exp[-v_j(X_D + \gamma_i^D)s]$ is a probability density satisfying $\int_0^\infty \rho_{ij}(s) ds = 1$ and $\Phi(s; x, X_D)$ is the solution of the differential equation $\frac{d\Phi}{ds} = \chi_1(X_D, \Phi), \Phi(0) = x$.

We recognize above the generator of a PDP with breakage. Contrary to the previous case of breakage resulting from super-reactions $\mathscr{S}_2$, the breakage size is now continuously distributed. The switch transitions $X_{D_i} = 0 \rightleftharpoons X_{D_i} = 1$ disappear in the singular limit (the substrates of the reactions $\mathscr{S}_3$ remain practically all the time in the state $X_{D_i} = 0$).

Practical criteria for identifying small parameters and super-reactions leading to piecewise deterministic approximations

The law of large numbers is applicable in the limit $\mathscr{V} \to \infty$. Certainly, in cell biology, the idea of infinite volumes should be considered with care. For this reason we will replace this condition by a set of easy to check criteria concerning relative orders of parameters of the models. These criteria will concern piecewise deterministic approximations. Criteria for hybrid diffusion approximation, involving central limit theorem, will not be discussed in this paper.

Our criteria are relative to two reference quantities. The first reference is a large number N representing a lower bound for the numbers of molecules of continuous species. The second reference is a time τ, which is a lower bound both for the characteristic times of the deterministic dynamics of the continuous variables τ_C and for the average time between two successive jumps of the discrete variables τ_D, namely $\tau = \min(\tau_C, \tau_D)$. τ can be related to kinetic parameters by methods exposed in [51,52].

The applicability of the law of large numbers to continuous variables implies two conditions. The first condition is that jump sizes should be small with respect to the numbers of molecules, i.e. $\gamma_i^C \ll N$ for all $i \in \mathscr{R}_C$. The second condition is that the number of jumps of the continuous variables on the timescale τ should be big, i.e. $V_i \tau \gg 1$ for all $i \in \mathscr{R}_C$. The two conditions are fulfilled simultaneously in the limit $\mathscr{V} \to \infty$ because N, V_i scale with $\mathscr{V}$ and γ_i^C, τ do not depend on $\mathscr{V}$.

A similar condition must be satisfied by super-reactions of the type 1: $V_i \gg \tau^1$ also for all $i \in \mathscr{S}_1$. The super-reactions of the type 2 do not satisfy the small jump condition, i.e. γ_i^C must be comparable to N (not much smaller) for all $i \in \mathscr{S}_2$.

In order to have singular switching we need a new small parameter $\ll 1$ and the following set of conditions:

a) There are super-reactions of the type 3. These are just very quick super-reactions of the type 1.

Table 1: Practical criteria to be satisfied by various reactions.

Reaction set	Approximation	Condition
$\mathscr{R}_C$	hybrid all	$\gamma_i^C \ll N$ and $V_i \gg (\tau)^{-1}$
$\mathscr{S}_1$	hybrid with switching	$\gamma_i^C \ll N$, $V_i \gg \tau^1$
$\mathscr{S}_2$	hybrid with breaking	$\gamma_i^C \sim N$
$\mathscr{R}_{D_i}^-$	hybrid with breaking as singular limit	$V_i = (\tau)^{-1} \gg \tau^1$
$\mathscr{S}_3$	hybrid with breaking as singular limit	$V_i \gg (\tau)^{-1}$
cycling reactions	averaged hybrid	$k_{lim} \gg (\tau_C)^{-1}$
branching reactions	averaged hybrid	$k \ll k_j$, j cycling

See text for definitions of various reaction sets. τ_C, N are typical timescale and number of molecules of the continuous components, respectively. $\tau = \min(\tau_C, \tau_D)$, where τ_D is the mean time between two successive jumps of the discrete components. V_i are the average numbers of reactions per unit time (intensities of jump Markov processes), k are first order kinetic constants and γ_i are the stoichiometric vectors.

b) Super-reactions of the type 3 act only on a laps of time $\tau = \tau$ i.e. $V_i = (\tau)^{-1} >> \tau^1$ for $i \in \mathcal{R}^-_{D_i}$. Reactions in $\mathcal{R}^-_{D_i}$ that inactivate super-reactions of type 3 could thus be as frequent as reactions in $\mathcal{R}_C$ and in $\mathcal{S}_1$ (but not as quick as reactions in $\mathcal{S}_3$).

c) Super-reactions of the type 3 are quick enough to produce breakages comparable to *N* during the time τ, i.e. $V_i >> (\tau)^{-1}$ for $i \in \mathcal{S}_3$.

The conditions of the type $V_i >> \tau^1$ can be simplified if the reactions are first order with respect to discrete species. In this case $V_i = k_i X_D$ with $X_D \approx 1$, therefore the condition is $k_i >> \tau^1$. All these criteria are summarized in the Table 1.

Averaging for stochastic chemical kinetics
The performance of the hybrid algorithm can be very bad when the discrete mechanism contains rapid cycles which effectuate many reactions on the deterministic timescale τ_C (this is the timescale on which the continuous variables have significant variation). Indeed, in this situation the deterministic solver is artificially sampling the interval between two discrete cycle reactions. First, this leads to unreasonable increase of the simulation time. Second, the condition on the number of jumps of continuous variables is not satisfied and the hybrid approximation is not accurate. In this case, the hybrid scheme performs worse than SSA. It is therefore important to eliminate rapid cycling from the system before implementing numerical schemes for the hybrid approximations. This can be done by averaging.

Averaging principles are widely used for deterministic (see classical textbooks [53,54] more recently revisited in the non-autonomous case by [55]) and stochastic (stochastic differential equations [56], Markov chains [35,57]) systems.

The classical averaging idea is to identify fast ergodic (that, starting in any value, can reach in finite, small time any other value in a given phase space set) variables and to average the dynamics of the slow variables with respect to the quasi-stationary distribution of the fast variables. An important difficulty is to identify the right slow and fast variables. For perturbed Hamiltonian dynamical systems the action-angle variables provide the natural framework for averaging [53].

In the case of chemical kinetics, there are two kinds of slow variables that should be taken into account for averaging. The first kind are just slow components (discrete components that change infrequently or continuous components with large deterministic timescale). The second kind are linear combinations of components that are conserved by the fast cycling dynamics. These new slow variables, that play a role similar to the action variables in Hamiltonian systems, provide new "aggregated" or "lumped" species in the averaged system. Notice that aggregation of states has been used for averaging Markov chains [35,57]. The aggregation of species that we propose here adapts this type of argument to the case of chemical kinetics. Variable aggregation for simple deterministic reactive models has been used in relation to applications in ecology (see for instance [58]). In [36,52] we proposed a general solution for simplification of reaction networks, that works for any linear mechanism with well separated constants. In this algorithm, species aggregation is systematically applied to hierarchies of rapid cycles. We show here that the same algorithm can be effectively used also in the stochastic case.

Averaging principles for reaction mechanisms
Averaging can be applied to multi-scale reaction mechanisms. This procedure leads to averaged hybrid models. In order to apply averaging, one should first identify a sub-model of discrete components, satisfying the following conditions:

C1) the dynamics of the sub-model is much faster than the dynamics of continuous and of other discrete, slower modes.

C2) the dynamics of the sub-model is ergodic: starting with any state, the system can reach any other state in finite time.

The general procedure to obtain averaged hybrid simplifications is described in the Additional File 1. Some cases are particularly interesting.

The first case corresponds to *fast discrete cycles producing continuous species*. In this case rapid super-reactions $\mathcal{S}_1$ change both discrete and continuous components (Eq.(13) is not valid). The discrete components that are affected by such reactions are fast discrete variables. In the resulting hybrid model, both the continuous dynamics and the slow discrete reaction rates should be averaged with respect to the fast discrete variables.

The second case corresponds to *fast, purely discrete cycles*. In this case, some of the cycles of the sub-mechanism $\mathcal{R}_D$ are at least as fast or faster than reactions $\mathcal{R}_C$. The other reactions of $\mathcal{R}_D$ are much slower. In the resulting model,

rates of the slow reactions of $\mathcal{R}_D$ should be averaged with respect to the fast discrete variables.

In both cases, one needs the steady probability distribution for the fast discrete sub-model. All the slow processes will be averaged with respect to this distribution. The calculation of the steady probability distribution can be easily done only when the sub-model is linear. Considering linear sub-models has also another advantage. In this case, averaging and aggregation lead to a coarse-grained reaction mechanisms. For these reasons, we have developed an algorithm for linear sub-models. Of course, gene network models are generally nonlinear. However, this does not mean that all the parts of such models behave nonlinearly. Many sub-models, in particular monomolecular cycles, can be simplified by averaging methods that we designed for linear networks.

Cycle averaging for linear sub-models
Linear sub-models
The are two types of linear reaction networks: monomolecular networks and first order networks. The structure of monomolecular reaction networks can be completely defined by a simple digraph, in which vertices correspond to chemical species A_i, edges correspond to reactions $A_i \rightarrow A_j$ with rate constants $k_{ji} > 0$. For each vertex, A_i, a positive real variable c_i (concentration) is defined.

The deterministic kinetic equation is

$$\frac{dc_i}{dt} = \sum_j k_{ij} c_j - (\sum_j k_{ji}) c_i, \qquad (24)$$

First order reaction networks include monomolecular networks as a particular case, and are characterized by a single substrate and by reaction rates that are proportional to the concentration of the substrate. First order reaction networks can contain reactions that are not monomolecular, such as $A \rightarrow A + B$, or $A \rightarrow B + C$. We shall restrict ourselves to pseudo-conservative first order reactions, ie reactions that do not change the total number of molecules in a given submechanism ($A \rightarrow A + B$ reactions are allowed, provided that B is external to the submechanism; similarly $A \rightarrow B + C$ reactions are allowed, provided that either B or C is external to the submechanism). With such constraints, the total number of molecules in the sub-mechanism is conserved and the kinetic equations are the same as (24). Degradation reactions can be studied in this framework by considering a special component (sink), that collects degraded molecules. Further release of the constraints is possible. For instance, the system can be opened by allowing constant (or slowly variable) production terms in Eq.(24). These terms will change the steady state, but will not influence the relaxation times of the system. The linear sub-mechanisms can be considered as part

of a nonlinear network, given fixed (or slowly changing) values of external inputs (boundaries). For instance, even in systems of binary reactions, one can define pseudo-monomolecular reactions when one of the substrates of the binary reaction is not changing (or changing slowly). This condition can be fulfilled if the substrate is in excess, for instance.

The stochastic dynamics of a unique molecule in such linear reaction network is given by the probability $p(j, t)$ that the molecule is in A_j at the time t. We can easily show that the master equation for $p(j, t)$ is the same as the deterministic kinetic equation (24). Considering only one molecule does not restrict generality because when several molecules are present in a linear network, these behave independently. Thus, the simplification algorithm proposed for deterministic networks [36,52] can be also applied to stochastic networks [59]. The algorithm is based on a set of operations transforming the reaction graph into an acyclic digraph without branching (no graph component is the substrate of more than one reaction). For such type of graphs the eigenvectors and eigenvalues of the kinetic matrix can be straightforwardly calculated, which completely solves the problem of deterministic dynamics. We could follow precisely the same approach to simplify and then solve the stochastic dynamics. However, we argue here that applying only a few steps of the algorithm is enough for effectively reducing computational time of the SSA method.

Cycle averaging algorithm
Let us recall that the reduction method presented in [36,52] uses two types of operations acting on the reaction graph.

I. Construction of an auxiliary network (dominance). For each node A_i of a linear sub-mechanism, let us define κ_i as the maximal kinetic constant for reactions $A_i \rightarrow A_j$: $k_i = \max_j\{k_{ji}\}$. For the corresponding j we use the notation $\varphi(i)$: $\varphi(i) = \arg \max_j\{k_{ji}\}$. An auxiliary reaction network $\mathcal{W}$ is the set of reactions $A_i \rightarrow A_{\varphi(i)}$ with kinetic constants κ_i. In such a network there is no branching: if several reactions consume the same component A_i, only the quickest one is kept and all the other discarded.

II. Glueing cycles (aggregation). Rapid cycles are replaced by a single node. Constants of reactions leaving these cycles are renormalized according to an averaging principle (see the Additional File 1).

In order to present the simplification algorithm let us use two simple examples.

First, let us consider a chain of molecular reactions $A_1 \rightarrow A_2 \rightarrow ... A_m$. The reaction rate constant for $A_i \rightarrow A_{i+1}$ is k_i. All rate constants are considered well separated, i.e. either $k_i << k_j$ or $k_i >> k_j$ for any $i \neq j$.

The smallest rate constant in the chain is called limiting, and denoted by $k_{\lim}$. If $1/k_{\lim} << \tau_C$ (rapid chain), then all molecules A_1 are transformed into molecules A_m on a timescale much quicker than the deterministic time τ_C. We can thus ignore the chain reactions and consider that the entire mass of the chain is practically always in A_m. This is equivalent to considering the chain at quasi-stationarity because the steady state probability distribution of a chain is a Dirac delta measure localized at the end of the chain. However, if we do not simplify chains, then simulating them by Gillespie's SSA will not be computationally expensive because the mass of the chain is transferred to the end of the chain A_m in a number of steps that is relatively small (this is bounded by the total mass of discrete species multiplied by the longest chain length).

As a second example, let us consider the cycle C be the following sequence of mono-molecular reactions $A_1 \rightarrow A_2 \rightarrow ... A_m \rightarrow A_1$. Let all rate constants be well separated and the smallest one be k_{lim} like before.

We add to the cycle one branching reaction; this transforms A_j a component of the cycle into B a component exterior to the cycle.

We consider the following distinct situations:

(I) the branching reaction is $A_j \rightarrow B$ of rate constant k and $k >> k_j$,

(II) the branching reaction is $A_j \rightarrow B$ and $k << k_j$,

(III) the branching reaction is $A_j \rightarrow A_j + B$, or

(IV) the branching reaction is $A_j \rightarrow A_{j+1} + B$ of rate constant k_j.

In the situation (I) the exit reaction is faster and dominates the cycling reaction $A_j \rightarrow A_{j+1}$. According to the rule for auxiliary networks in this case (that we call "broken" cycle) the cycle can be opened (by eliminating the cycling reaction $A_j \rightarrow A_{j+1}$) and the resulting multiscale dynamics is the one of a chain; we recover the previous example and in this case we can safely decide to do nothing.

In the situation (II) the exit reaction is much slower than the cycling reaction. In this case the molecules inside the cycle have rapid transformations and the mass distribution inside the cycle can be considered to reach quasi-stationary distribution. We call this cycle "unbroken".

As discussed in [36,51,52], the relaxation time of a cycle with separated constants is the inverse of the second slowest rate constant $k^{(2)} >> k^{(1)} = k_{lim}$. To understand this, one should consider the two possible paths to equilibrate a cycle, one passing by the slowest step and the quicker one passing by the second slowest step: the quicker short-cuts the first one. Thus, a cycle can be considered quasi-stationary if $k^{(2)} >> 1/\tau_C$. A non-averaged fast cycle is computationally expensive in SSA, if a molecule can perform a huge number of steps along the cycle on the timescale τ_C. The corresponding condition involves the quasi-stationary flux (not the relaxation time) and reads $k^{(1)} = k_{lim} >> 1/\tau_C$.

From a quasi-stationary cycle, the mass is lost stochastically, but slowly by the branching reaction. The intensity of the loss process can be calculated by replacing X_j by its average with respect to the quasi-stationary distribution of the cycle. The average of X_j is $\bar{X}_j = N(t) k_{lim}/k_j$, where $N(t)$ is the total mass inside the cycle $N = \sum_{j=1}^{m} N_j$. We obtain the average intensity $\bar{\lambda} = k\bar{X}_j = N(t) kk_{lim}/k_j$.

In the situations (III) or (IV) the average intensities of the branching reactions are $k\bar{X}_j = N(t) kk_{lim}/k_j$ and $k_j\bar{X}_j = N(t) k_{lim}$, respectively.

All these operations can be presented as a:

Simplification algorithm
Input:

a first order reaction mechanism G with separated kinetic constants.

Output:

a simplified first order reaction mechanism.

while there are fast unbroken cycles

for each cycle **in** G *not containing reactions of the type (I) having a sufficiently intense flux* ($k_{lim} >> 1/\tau_C$) **do**

1: "glue" the cycle into a single node C having the total mass N;

2: replace the exit reaction of the type (II) $A_j \rightarrow B$ of rate constant k by a reaction $C \rightarrow B$ of effective constant $k' = kk_{lim}/k_j$;

3: replace the reaction of the type (III) $A_j \rightarrow A_j + B$ or rate constant k by a reaction $C \rightarrow C + B$ of effective constant $k' = kk_{\text{lim}}/k_j$;

4: replace the reaction of the type (IV) $A_j \rightarrow A_{j+1} + B$ of rate constant k_j by a reaction $C \rightarrow C + B$ of effective constant $k' = k_{\text{lim}}$;

Imposing the "glueing" condition on the cycling flux and not on the cycle relaxation time allows to iterate the cycle averaging algorithm until no more fast unbroken cycles remain. This would not be possible when adopting a relaxation time criterion to perform averaging. Indeed, the relaxation time of a cycle is given by the second slowest constant. This leaves place for a counter-intuitive possibility: one can have slow cycles embedded into rapid cycles. For instance, a fast cycle whose nodes are "glued" cycles can have a slower node as the beginning of its limiting step. Indeed, the internal constants of this node are necessarily more rapid than the limiting step of the large cycle, but can be slower than the second constant which gives the relaxation time of the large cycle. The slow scales are lost by glueing. In order to recover the full multi-scale dynamics, a restoration operation has been considered at the end of the algorithm presented in [36,52]. In this operation, all "glued" cycles are restored without their limiting steps. Thus, slower cycle sub-dynamics can be recovered. Restoration is not needed for glued cycles satisfying the flux condition $k_{lim} \gg \tau_C^{-1}$, because these are automatically faster than the timescale τ_C. Our averaging method works for mechanisms with well separated constants. Generalization for partially separated constants are possible. For instance, one can consider cycles containing reactions with equivalent constants, but such that for each node of the cycle, the constant of the branching reaction is much smaller than the constant of the cycling reaction.

Noisy networks and design rules for hierarchies of cycles

As we have shown in the previous sections, intrinsic noise in biochemical networks is generated at a "microscopic level" by the discrete variables and can be observed at the "mesoscopic level" of the continuous variables either as switching, or as breaking events. Thus, when there are no discrete variables (all species are in large numbers), there is no intrinsic noise. Also, if the switching events are much faster than the deterministic time scale, averaging principles apply and noise is not transmitted to the continuous variables: the deterministic approximation is again recovered. The only way to transmit noise by switching to the mesoscopic level is by intermittency and this needs partic-

ular combinations of slow reactions that change the values of the discrete species and frequent reactions that change the continuous species. Intrinsic noise transmission from micro- to meso-scopic level results from certain (not all) combinations of low numbers, fast and rapid timescales. Some general design rules relating topology to dynamic properties relative to noise can be derived from our approach.

By hierarchies of cycles we understand reaction mechanisms containing cycles connected by branching reactions forming higher level cycles. The nodes of higher level cycles are "glued" cycles from the lower levels. Hierarchies of cycles in discrete variables have interesting properties with respect to noise production. Generally, in cycle hierarchies, effective constants of branching reactions are at least as slow or slower than the limiting step (slowest reaction) of the node (glued cycle) from which they start. Coupling cycles into hierarchies is a way to produce slower and slower reactions from initially rapid reactions and generate thus intermittency.

As a possible design rule, we could state: exit reactions of the type (II) or (III) (but not (IV)) generate intermittency when the exit node is not the beginning of the limiting step in some unbroken fast cycle.

Indeed, unless k_j is the limiting step in the cycle, one has $k_{\text{lim}}/k_j \ll 1$. Then, the average intensity of the exit reaction of the type (II) or (III) is weak and could represent a source of intermittency in the system. This situation should be avoided for less noise in the system, or created when noise is wanted.

An example of how this rule applies to the biochemistry of bacteria will be given below. A systematic investigation of the possibilities of this class of design rules will be presented elsewhere in relation to synthetic biology.

Results and Discussion
Methodology to obtain hybrid simplifications
We demonstrate how hybrid simplifications can be obtained from Markov pure jump models for gene networks. The simplified models preserve the main stochastic properties of the initial complex models and can replace these models in simulations.

The simplification procedure consists of four successive steps:

I Identification of discrete and continuous variables, partition of the reactions.

II Cycle averaging.

III Identification of super-reactions.

IV Construction of the hybrid simplifications.

In order to justify the utility of our approach we show by examples that all types of hybrid processes that we have discussed are represented in gene networks models.

To introduce the examples we employ the following notation. Reactions are numbered by integers. If the i-th reaction is reversible, then R_i^+ is the i-th reaction in the forward direction and R_i^- the i-th reaction in the reverse direction. Irreversible reactions are denoted simply R_i.

The rate constants units are s^{-1} for monomolecular and $s^{-1}(M)^{-1}$ for binary reactions. In order to obtain pseudo-monomolecular rate constants from binary reaction constants with slowly varying substrates X we have used the following formula:

$$k_{mono}[\text{s}^{-1}] = k_{binary}[\text{s}^{-1}(\text{M})^{-1}] \frac{X[\text{molecules}]}{N_A \mathcal{V}[(\text{M})^{-1}]}$$

where N_A is the Avogadro number and $\mathcal{V}$ is the cell volume. For a bacterium, $N_A \mathcal{V} \approx 1(\text{n M})^{-1}$.

Similarly, the reaction rates are calculated as

$$V_{mono}[(\text{reactions})\text{s}^{-1}] = k_{mono}[\text{s}^{-1}]X[\text{molecules}],$$

$$V_{binary}[(\text{reactions})\text{s}^{-1}] = \frac{k_{binary}[\text{s}^{-1}(\text{n M})^{-1}]}{N_A \mathcal{V}[(\text{n M})^{-1}]} X[\text{molecules}]Y[\text{molecules}]$$

We discuss two simple hybrid models, one with switching the second one with breakage, then two more complex models that need averaging. All our simulations were performed using MATLAB 2008 in a Windows XP32 environment with a dual core INTEL 6700, 2.65 GHz processor.

Hybrid model with switching: Cook's model
The simplest model with switching has been introduced by Cook [24] as a model for haploinsufficiency phenomena. This model can be described by the following system of biochemical reactions:

$$G \underset{k_{m1}}{\overset{k_1}{\rightleftharpoons}} G^*, G^* \overset{k_2}{\rightarrow} G^* + P, P \overset{k_3}{\rightarrow}$$
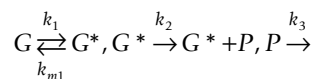
In this model, G, G^* represent inactive, respectively active states of the gene. In the active state, the gene produces some protein P. Let G_0 be the gene copy number, which is a conserved quantity of the dynamics $G + G^* = G_0$. The haploinsufficiency regime corresponds to a small value of G_0. For the simplicity of the argument we consider $G_0 = 1$.

Let $X = (X_D, X_C)$ be the state vector. We notice that, the partitions of the species is the following: the species in large number $X_C = \{P\}$, and the species in small number $X_D = \{G, G^*\}$. This partition of the species defines a partition of reactions. With the above notations we get:

$$\mathcal{R}_D = \{R_1^+, R_1^-\}, \quad \mathcal{R}_C = \{R_3\}, \quad \mathcal{R}_{DC} = \{R_2\}$$

The deterministic timescale is $\tau = (k_3)^{-1}$. The stoichiometric vectors have all lengths of order one, much smaller than N (the number of proteins).

For the parameters values used in [24] we have $k_2/k_3 \gg 1$ and R_2 is a super-reaction of the type $\mathcal{S}_1$. The reaction R_3 satisfies $V_i = N k_3 \gg k_3 = \tau^1$, which means that the law of large numbers can be applied to the continuous variable.

In the first order of the Kramers-Moyal development we obtain a PDP approximation with switching. In the following, we will denote by $x = [P]$ the continuous variable and by $\theta = G^*$ the discrete variable. This model is a piecewise deterministic process with state space $(\theta, x) \in \{0, 1\} \times \mathbb{R}$.

The flow function $\chi(\theta, x)$ is given by

$$\chi(\theta, x) = -k_3 x + k_2 \theta,$$

and the jump intensity λ is defined by

$$\lambda_\theta = k_{m1}(1 - \theta) + k_1 \theta.$$

The resulting PDP is the same as ON-OFF systems studied in operational research [25]. The model has been proposed as abstract model for stochastic protein production in several other places [29,60]. With the PDP description of the system and using the hybrid algorithm introduced in the section Methods, we draw a possible time evolution of this system (see Figure 1a). The same initial condition was used when the Gillespie algorithm was implemented for the model described above. The simulation time to generate a trajectory using a PDP approximation, was 2.6 seconds, while with the Gillespie algorithm the simulation time was 14.5 seconds in the average. The trajectories look qualitatively the same: production intervals are followed by degradation intervals, with no discontinuity in the continuous variable.

In order to quantitatively test the accuracy of the PDP algorithm, we have computed the stationary distribution of the number of proteins using a piecewise deterministic simulation and compared it to the estimated distribution from Gillespie trajectories.
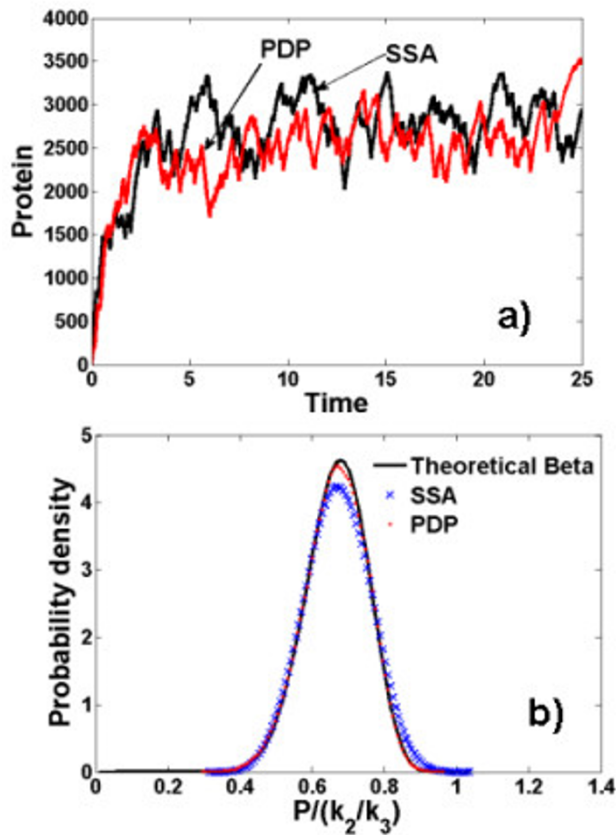
Figure 1
Cook's haploinsufficiency model, an example of PDP with switching. a) Time evolution of the protein concentration using Gillespie SSA method and PDP approximation. b) Estimated Gillespie steady probability distribution vs. estimated steady piecewise-deterministic (PDP) probability distribution. The theoretical beta distribution for the PDP is added.

The numerical methods to estimate stationary distributions are described in the Additional File 1. The theoretical curve for the PDP model is a beta distribution. The variable $x = P/(k_2/k_3)$ follows the Beta distribution $\mathcal{B}(k_1/k_3, k_{m1}/k_3) : p(x) = \frac{1}{B(k_1/k_3, k_{m1}/k_3)} x^{k_1/k_3-1}(1-x)^{k_{m1}/k_3-1}$, B is the Beta function [29]. The result of various comparisons is represented in Figure 1b.

Cook's model operates in the region $k_2 \gg k_{m1}$, which corresponds to a broken cycle in the discrete variables. If the opposite inequality is valid $k_2 \ll k_{m1}$, then cycle averaging is needed. The cycle $G \underset{k_{m1}}{\overset{k_1}{\rightleftharpoons}} G^*$ should be glued to a point $\bar{G}$ of mass $\bar{G} = G_0 = 1$ and the branching reaction R_2

becomes $\bar{G} \overset{k'_2}{\rightarrow} \bar{G} + P$, where $k'_2 = k_2 \frac{k_1}{k_1+k_{m1}}$. The resulting model is a birth and death model with effective birth rate $k'_2 G_0$ and death rate constant k_3. Provided that $k_2/k_3 \gg 1$ (meaning that R_2 remains a super-reaction of the type $\mathcal{S}_1$), the model can be approximated by a completely deterministic process with flow given by the averaged flow function $\chi(x) = -k_3 x + k'_2 G_0$.

Hybrid models with breakage: neuroscience and bacterium operator sites

Neural systems exhibit stochastic behavior. Stein [61,62] proposed a simple Markovian model for the evolution of a neuron membrane potential. In this model, discontinuous random jumps of the potential are followed by exponential decay. We do no discuss here how to obtain this hybrid model from a microscopic pure jump model. Stein's model is a phenomenological representation of very complex electric and biochemical processes. We demonstrate its properties in terms of trajectories and stationary distribution. The subthreshold behavior of the membrane potential of a neuron cell is described in this model by a hybrid Markov process of continuous variable $V(t)$ with jumps of constant intensity λ and constant amplitude in the continuous variable. Although in the original Stein model there are jumps of positive and negative sign, corresponding to excitatory and inhibitory synaptic activations, here we consider only positive sign jumps. If t is the moment of jump, then $V(t^+) - V(t^-) = a$, where $a > 0$ is the amplitude. Between two jumps V decays according to $\frac{dV}{dt} = -\alpha V$ where α is a constant.

The generator of such process is:

$$\mathcal{A}f(x) = -\alpha x f'(x) + \lambda[f(x+a) - f(x)]$$

The steady probability distribution $p(x)$ for such a process satisfies the following delay differential equation:

$$\alpha(xp(x))' + \lambda[p(x-a) - p(x)] = 0 \qquad (25)$$

and $p(x) = 0$, for all $x < 0$.

Analytical solutions are not known for this delay differential equation. In the Additional File 1 we describe the numerical scheme to calculate the steady probability distribution. The comparison between the simulated and calculated distribution is shown in Figure 2c-d. In Figure 2a-b we show trajectories of the system. Stein's model could also apply (even better, because its main defects with

respect to neurons, such as the absence of a refractory period, is not a problem for genes) to gene networks. Generalizations of this model, allowing for continuous distributions of the jumps, could be used as an archetype of intermittent activity of a promoter.

Hybrid models with breakage can result as singular limits of hybrid models with switching as discussed in the Methods section *Hybrid approximations with singular switching*. The typical case is a bacterium operon. This model can be found in many places in literature. A detailed molecular example [63,64] will be studied as our last example (repressed operator site in a bacterium). A simpler version of this model is proposed by [65]. It consists of the following reactions:

$$DNA \xrightarrow{k_1} mRNA \xrightarrow{k_2} Protein, mRNA \xrightarrow{k_3} \varnothing, Protein \xrightarrow{k_4} \varnothing,$$
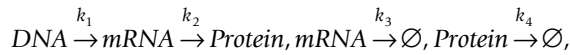
The first reaction is zero order, all the other reactions are first order. The parameters satisfy $k_1 = ak_4$, $k_2 = bk_3$, $k_4 = k_3$. We consider that $b \gg 1$, $\ll 1$.

From the aspect of the trajectories (these show bursting), the authors of [65] hypothetize that a PDP approximation with breakage is the natural simplification and solve the corresponding stationary hybrid Fokker-Planck equation for the protein component. We show here that the hybrid Fokker-Planck equation given in [65] can be found as an application of our approach.

First, we notice that a PDP limit is applicable to the Markov jump process. The species partition is $X_D = mRNA$, $X_C = Protein$. The mRNA component follows a birth and death process with birth intensity k_1 and death intensity $(k_3 + k_2) \, mRNA$. Using the master equation for birth and death processes [42] and considering that $k_1/(k_3 + k_2) = a/(1 + b) \ll 1$, it follows that the probability to have zero or one molecule of mRNA is close to one (the probability to have two or more molecules is negligible).

This justifies that mRNA is a discrete species. The partition of the reactions is $\mathscr{R}_D = \{R_1, R_3\}$, $\mathscr{R}_{DC} = \{R_2\}$, $\mathscr{R}_C = \{R_4\}$. The timescale of the continuous variables is $\tau = (k_4)^{-1}$. $V_2 = k_4 X \gg \tau$, where X is the number of proteins, provided that $X \gg 1$. This condition, allowing application of the PDP approximation, is satisfied because $b \gg 1$ (the significance of b is the average number of proteins produced in a bursting event).

Let us show that we have singular switching, equivalent to breakage. The reaction R_2 is a super-reaction of the type

$\mathscr{S}_3$ because $V_2 = k_2 = b(\tau)^{-1}$ The discrete variable *mRNA* can be considered to remain zero almost all the time, except during the negligible duration of the bursts when *mRNA* = 1. The reaction R_D^+ corresponds to the transition $0 \to 1$, while the reaction R_D^- corresponds to the transition $1 \to 0$ of the discrete component *mRNA*. The intensity of the reaction R_3 satisfies $V_3 = (\tau)^{-1}$. Thus, the mean duration of the burst is τ. According to the section Methods, we are in the case of a PDP with singular switching. The evolution of the unique continuous variable x (protein concentration) is well approximated by the following hybrid generator (obtained after a simple change of the integration variable in Eq.(23)):

$$\mathscr{A}f(x) = -k_4 x f'(x) + k_1 \int_0^\infty [f(x + \tilde{x}) - f(x)] \frac{\mathscr{V}}{b} \exp[-\tilde{x}\frac{\mathscr{V}}{b}] d\tilde{x}$$

which shows that b is the mean number of proteins produced in a burst (mean size of the breakage).

We can notice a continuous distribution of the breakage size (exponential distribution), situation different from the Stein model. The Fokker-Planck equation can be solved in this case. The steady distribution of the continuous variable x is the Gamma distribution $p(x) = \frac{1}{(b/\mathscr{V})^a \Gamma(a)} x^{a-1} e^{-x\mathscr{V}/b}$ [65].

First complex example: λ-phage toggle switch
In this section we revisit a classical example of toggle molecular switch. The model of cro-repressor (*cI*) switch in λ-phage, a temperate bacteriophage of Escherichia coli, was investigated by many authors [66-69].

The life cycle of phages has two alternative pathways, lysogenic when the phage duplicates synchronously with the host genome and lytic when the phage produces large amounts of its own mRNA. The switch between these two pathways is controlled by the level of *cI* protein. The lysogenic pathway corresponds to large levels of *cI*, while the lytic pathway corresponds to low levels of *cI*. A *cI* dimer may bind to any of the three operators: OR1, OR2, and OR3, in this order. By cooperativity, OR1 and OR2 are almost simultaneously occupied by *cI*. The third site OR3 will be occupied only when the *cI* concentration is high enough. A simpler model, in which OR1 is absent, may be considered [68]. Let *cI*, *cI*$_2$ and *D* denote the repressor, repressor dimer, and DNA promoter site. The dynamics of the operator sites is described by the following reaction system:
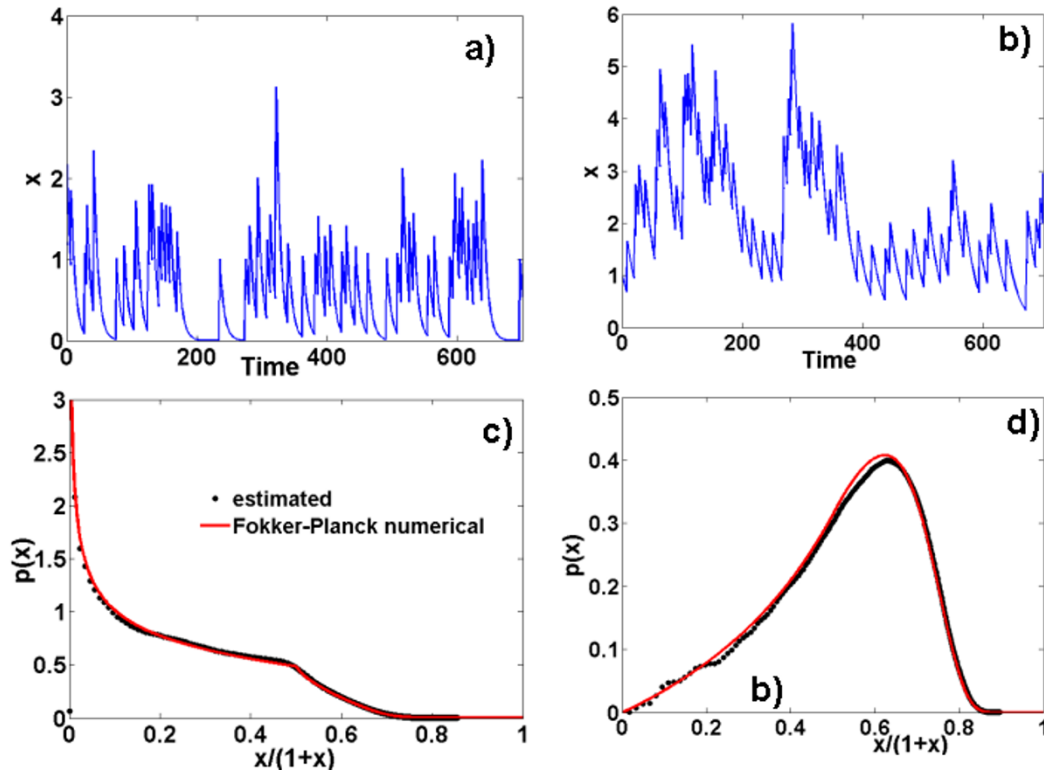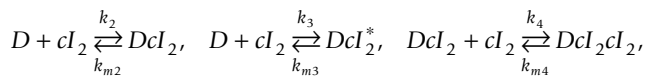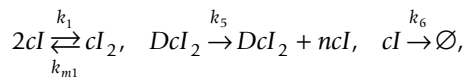
Figure 2
Stein's model with excitatory synaptic activation (also bursting bacterium promoter with constant number of proteins produced per burst), an example of PDP with breakage. Trajectories for a) $a = 1$, $\alpha = 0.15$, $\lambda = 0.1$, b) $a = 1$, $\alpha = 0.05$, $\lambda = 0.1$. Steady probability distributions c) same parameters as a); d) same parameters as b).

$$D + cI_2 \underset{k_{m2}}{\overset{k_2}{\rightleftarrows}} DcI_2, \quad D + cI_2 \underset{k_{m3}}{\overset{k_3}{\rightleftarrows}} DcI_2^*, \quad DcI_2 + cI_2 \underset{k_{m4}}{\overset{k_4}{\rightleftarrows}} DcI_2cI_2,$$

where the DcI_2 and DcI_2^* complexes denote binding to the OR2 or OR3 sites, respectively respectively binding to both sites.

The other reactions concern production, degradation of molecules cI, cI_2:

$$2cI \underset{k_{m1}}{\overset{k_1}{\rightleftarrows}} cI_2, \quad DcI_2 \overset{k_5}{\rightarrow} DcI_2 + ncI, \quad cI \overset{k_6}{\rightarrow} \varnothing,$$

where n is the number of proteins per *mRNA* transcript.

The state vector is $X = (X_D, X_C)$ where $X_D = \{D, DcI_2, DcI_2^*, DcI_2cI_2\}$, $X_C = \{cI, cI_2\}$. Note that the choice of discrete variables is dictated here, like in our first example, by the conservation law $D + DcI_2 + DcI_2^* + DcI_2cI_2 = D_0$ that

restricts the numbers of D, DcI_2, DcI_2^*, DcI_2cI_2 to small values.

The partition of the species induces the following partition of the reactions:

$$\mathscr{R}_D = \varnothing, \quad \mathscr{R}_{DC} = \{R_2^+, R_2^-, R_3^+, R_3^-, R_4^+, R_4^-, R_5\}, \quad \mathscr{R}_C = \{R_1^+, R_1^-, R_6\}.$$

The interesting property of the λ-phage model is its bistability. The naive calculation to find steady states uses quasi-stationarity of the deterministic dynamics. Then, we can use the equilibrium equations for reactions R_i, $i \in [1,4]$, the steady state equation for cI and the conservation law to compute the concentration x of cI at steady-state. $x = 0$ is one of the steady state, namely the lytic attractor. It also corresponds to an absorbing state of the Markov process (the process always stops when it reaches this state). The lytic state can be rendered non-absorbing by including, like in [68], a small basal production rate of cI protein in the model. However, this is not important for the illustration of our approach. If $k_2/k_{m2} = k_3/k_{m3}$ it fol-

lows that steady states (other than the lytic state) must satisfy:

$$\sigma u^4 + 2u^2 - \alpha u + 1 = 0 \qquad (26)$$

where $u = \sqrt{\frac{k_1 k_2}{k_{m1} k_{m2}}} x, \alpha = \frac{n k_5 D_0}{k_6 N_A \mathcal{V}} \sqrt{\frac{k_1 k_2}{k_{m1} k_{m2}}}, \sigma = \frac{k_4 k_{m2}}{k_2 k_{m4}}$.

Let us notice that failure of naive quasi-stationarity calculations was demonstrated for nonlinear subsystems [37]. However, these calculations can be justified here by linear cycle averaging (the promoter sub-mechanism reactions R_i, $i \in [2,4]$ are all pseudo-monomolecular).

Bistability occurs when the quartic equation (26) has real roots, which is the case when σ is small and α is big. More precisely, the model is bistable if $\alpha > \frac{4}{3} \frac{2+\sqrt{1+3\sigma}}{(1+\sqrt{1+3\sigma})^{1/4}}$.

Like in [68] we have used $\sigma = 5$. In order to ensure bistability we have chosen $\alpha = 7$.

The sub-mechanism R_{DC} contains rapid unbroken cycles that need to be averaged before any further approximation of the process. These rapid cycles (see Figure 3a) correspond to rapid binding-unbinding of the dimer cI_2 on the DNA. Three steps lead from the unreduced Model 1 to the averaged Model 2 (see also Figure 3a):

1.1 The cycle DcI_2, $DcI_2 cI_2$ is unbroken. It is glued to the node $\bar{D}cI_2$ whose total mass is equal to the mass of DcI_2 and $DcI_2 cI_2$.

1.2 The limiting step of the cycle is $k_{lim} = k_{m4} \ll k_4 cI_2$.

1.3 The branching reaction $DcI_2 \rightarrow n\ X + DcI_2$ is replaced by $\bar{D}cI_2 \rightarrow n X + \bar{D}cI_2$ of effective constant $k_5' = \frac{k_{m4}}{k_4 cI_2} k_5$. The reaction $DcI_2 \rightarrow D + cI_2$ is replaced by $\bar{D}cI_2 \rightarrow D + cI_2$ with the reaction constant $k_{m2}' = \frac{k_{m4}}{k_4 cI_2} k_{m2}$.

After averaging, two more cycles remain in the resulting Model 2. However, the rates of the remaining reactions $D + DcI_2 \rightarrow DcI_2$ and $D + DcI_2 \rightarrow DcI_2^*$ are equivalent, which does not allow further application of our algorithm. Furthermore, the slow reactions $\bar{D}cI_2 \rightarrow D + cI_2$, and $DcI_2^* \rightarrow D + cI_2$ produce intermittence and should by no means be averaged.

The next approximation is a first order partial Kramers-Moyal expansion. The reactions in R_C and the super-reaction $R_5 \in \mathscr{S}_1$ contribute to the deterministic dynamics of the continous species defined by the following differential equations:

$$\begin{aligned}\frac{dx}{dt} &= -2k_1 x^2 - k_6 x + 2k_{m1} y + k_5 DcI_2 \\ \frac{dy}{dt} &= k_1 x^2 - k_{m1} y\end{aligned} \qquad (27)$$

where x, y are the concentrations of cI and cI_2, respectivelly.

The remaining reactions induce the jump mechanism.

The time evolution towards the lysogenic attractor is represented in Figure 3b for the un-reduced Model 1 and in Figure 3c for the PDP Model 4 (which is obtained by averaging and first order Kramers-Moyal expansion from Model 1). Steady probability distribution close to this attractor is represented in Figure 3d for all models in this study. We can notice the intermittent behavior of the cI, cI_2 components that is well captured in the switching PDP approximations (Models 3 and 4).

Second complex example: Stochastic bursting of a repressed bacterium operon

Under strong repression, protein production from a bacterium operator site undergoes stochastic bursting. A stochastic model for protein production in prokaryotes has been introduced in [63]. The behavior of the operator site under the regulation of a repressor molecule that prevents protein production has been considered in [64]. The corresponding model is represented in the Table 2 and in Figure 4.

In this model, the bacterium is considered to be in exponential growth phase, increasing size and dividing normally. Cell growth is simulated by a linear increase of the volume in time. During replication, the nuclear material doubles (variables D,D.R,DRNAP). At fission, the nuclear material is halved and all other components are divided among daughter cells according to a binomial distribution.

I. Species and reaction partition

Some of the components (D, $D.R$, $D.RN\ AP$, $Tr\ RN\ AP$) are confined to small numbers by the conservation law $D + D.R + D.RN\ AP + Tr\ RN\ AP = const$.

Two more components *RBS* and *RibRBS* have small lifetimes $\approx 2s$ and can not accumulate to significant levels. Thus, the discrete variables are:
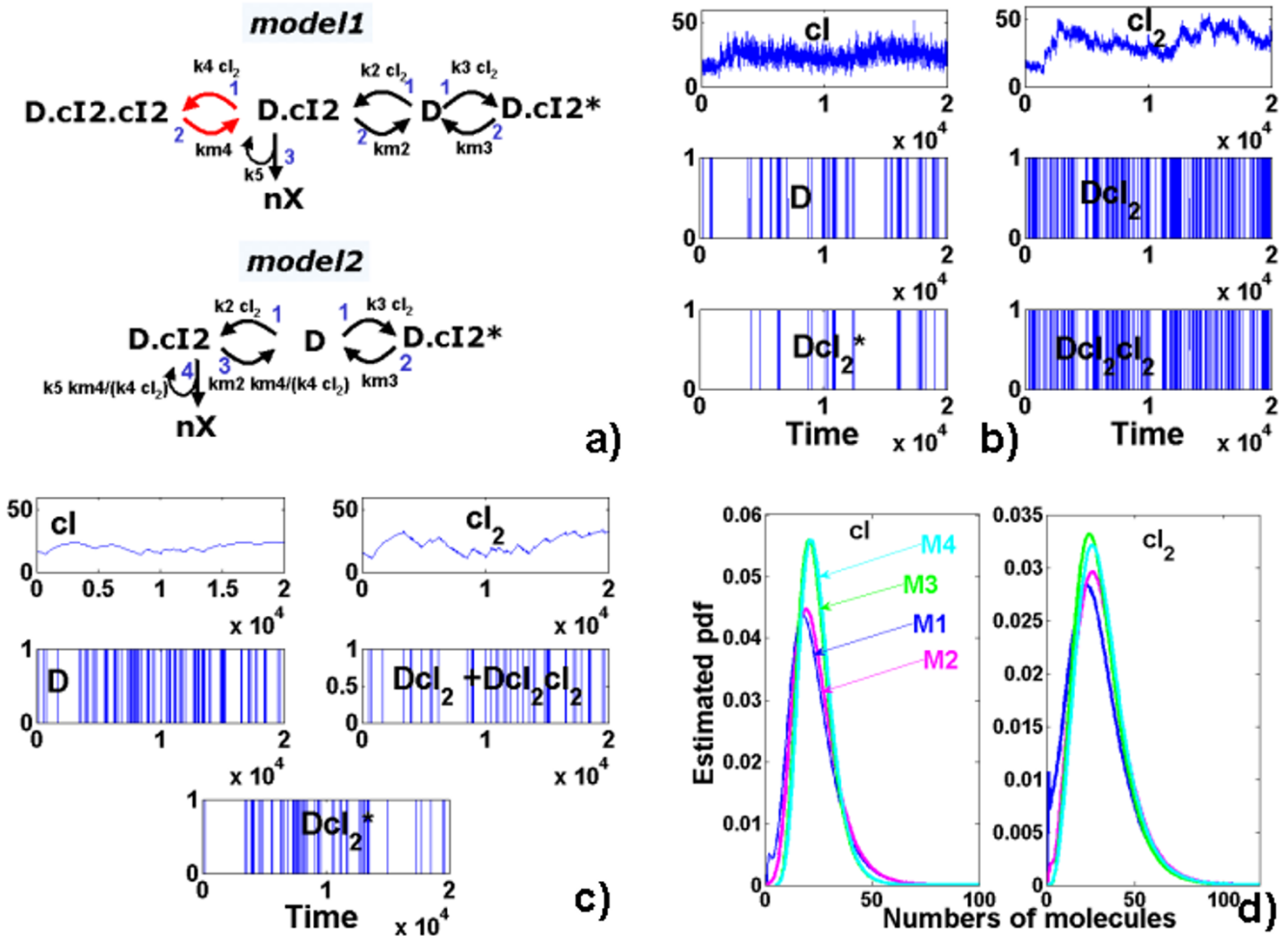
Figure 3
Lambda-phage model, an example of averaged piecewise-deterministic process. a) The cycle to be averaged is in red. Model 1 is the unreduced model, Model 2 is obtained from Model 1 by averaging. The integer labels represent orders of first order rate constants (1 represents the fastest reactions). b) SSA trajectory of the un-averaged model (Model 1). c) PDP trajectory of the averaged PDP (Model 4 which is the first order Kramers-Moyal approximation of Model 2). d) Comparison of estimated steady distribution for trajectories close to the lysogenic attractor (Models 3 and 4 are PDPs, obtained from Models 1 and 2, respectively).

$$X_D = (D, D.R, D.RNAP, TrRNAP, RBS, RibRBS).$$

The remaining components are in large numbers and form the continuous variables:

$$X_C = (ElRib, Protein, FoldedProtein).$$

We notice that Rib, RNAP and R are in relatively large numbers and practically constant, which justifies a first order reaction approximation for the mechanism R_D. The sets of discrete species D, $D.R$, $D.RN\,AP$ and RBS, $RibRBS$ form rapid unbroken cycles and can be averaged.

II. Cycle averaging
The cycle averaging procedure can be applied three times:

1.1 The cycle D, $D.R$ is unbroken. It is glued to the node D^* whose total mass is equal to the mass of D and $D.R$.

1.2 The limiting step of the cycle is $k_{lim} = k_{m1} \ll k_1$.

1.3 The branching reaction $D \to D.RN\,AP$ is replaced by $D^* \to D.RN\,AP$ of effective constant $k'_2 = \dfrac{k_{m1}}{k_1} k_2$.

2.1 The cycle D^*, $D.RN\ AP$ is unbroken. It is glued to the node D^{**} whose total mass is equal to the mass of D and $D.R$ and $D.RN\ AP$.

2.2 We have $k'_2 << k_{m2}$ hence the limiting step of the cycle is k'_2.

2.3 The branching reaction $D.RN\ AP \rightarrow TrRN\ AP$ is replaced by $D^{**} \rightarrow TrRN\ AP$ of effective constant $k'_3 = \frac{km1}{km2} \frac{k2 \times RNAP}{k1 \times R} k3$.

3.1 The cycle RBS, $Rib.RBS$ is unbroken. It is glued to the node RBS^* whose total mass is the one of RBS and of $Rib.RBS$.

3.2 The limiting step is $k_{m6} << k_6\ Rib$.

3.3 The branching reaction $Rib.RBS \rightarrow ElRib + RBS$ is replaced by the reaction $RBS^* \rightarrow ElRib + RBS^*$ of effective constant $k7' = k7$.

3.4 The branching reaction $RBS \rightarrow \varnothing$ is replaced by the reaction $RBS^* \rightarrow \varnothing$ of effective constant $k'_5 = \frac{k5km6}{k6} \frac{N_A V}{Rib}$.

Notice that a loss of accuracy should be expected from the application of the third averaging step. The separation of the branching and cycling reaction rate constants is not that good. Indeed, $k_7/k_{m6} \approx 0.22$ while in theory we need $k_7/k_{m6} << 1$.

Table 2: Set of reactions and parameters for the repressed bacterium operon model, from [64].

| Reaction	Parameters
Repressor Binding	
1) $D + R \underset{km1}{\overset{k1}{\rightleftarrows}} D.R$	$k1 = 10^8 M^{-1}s^{-1}$, $km1 = 1s^{-1}$
2) $D + RNAP \underset{km2}{\overset{k2}{\rightleftarrows}} D.RNAP$	$k2 = 10^8 M^{-1}s^{-1}$, $km2 = 10s^{-1}$
3) $D.RNAP \overset{k3}{\rightarrow} TrRNAP$	$k3 = 0.1s^{-1}$
4) $TrRNAP \overset{k4}{\rightarrow} RBS + D + RNAP$	$k4 = 0.3s^{-1}$
Translation	
5) $RBS \overset{k5}{\rightarrow}$	$k5 = 0.3s^{-1}$
6) $RBS + Rib \underset{km6}{\overset{k6}{\rightleftarrows}} RibRBS$	$k6 = 10^8 M^{-1}s^{-1}$, $km6 = 2.25s^{-1}$
7) $RibRBS \overset{k7}{\rightarrow} ElRib + RBS$	$k7 = 0.5s^{-1}$
8) $ElRib \overset{k8}{\rightarrow} Protein$	$k8 = 0.015s^{-1}$
Protein folding and decay	
9) $Protein \overset{k9}{\rightarrow} FoldedProtein$	$k9 = (ln2/5400)\ s^{-1}$
10) $Protein \overset{k10}{\rightarrow}$	$k10 = 10^{-5}s^{-1}$
11) $FoldedProtein \overset{k11}{\rightarrow}$	$k11 = 10^{-5}s^{-1}$
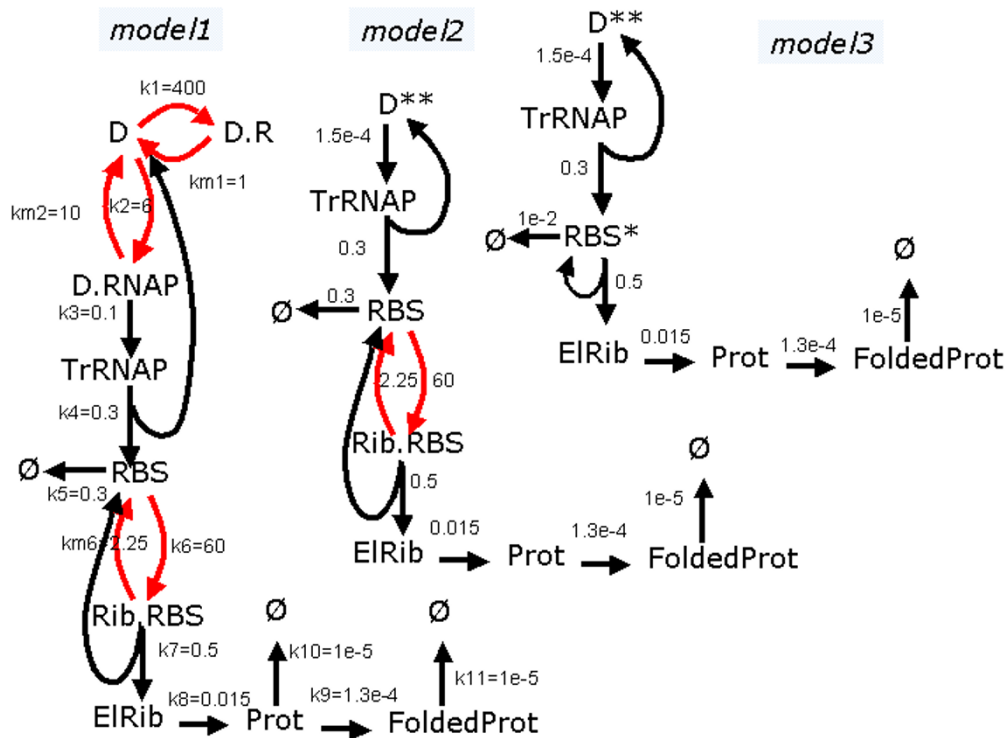
Figure 4
Repressed bacterium operon. The cycles to be averaged are in red. The numbers represent first order rate constants (including concentrations of buffered substrates where it is the case, here for *R* = 2500).

III. Identification of super-reactions
Notice that after cycle averaging, a low intensity reaction $D^{**} \to TRN\ AP$ results, producing intermittency (bursting). The reduced mechanism is represented in Figure 4b. The discrete/continuous partition of the species in the reduced mechanism is inherited from the initial model.

In the reduced mechanism, the reaction $RBS^* \to RBS^* + ElRib$ is very quick, even quicker than the continuous reactions $ElRib \to Prot$, $Prot \to FoldedProt$, $FoldedProt \to \varnothing$, therefore it is a super-reaction of the type $\mathcal{S}_3$.

IV. Hybrid approximation
First order Kramers-Moyal expansion applied to the averaged Markov jump process leads to a PDP with switching.

The continuous variables obey to the following differential equations:

$$\frac{dElRib}{dt} = k'_7 RBS^* - k_8 ElRib$$

$$\frac{dprotein}{dt} = k_8 ElRib - k_9 protein$$

$$\frac{dFoldedprotein}{dt} = k_9 protein - k_{11} Foldedprotein$$

$$(28)$$

The remaining three discrete components (D^{**}, *TrRN AP*, RBS^*) form a Markov jump process. Inside the reaction mechanism there is a rapid chain leading from *TrRN AP* to RBS^* and to *ElRib* production which is fed by a extremely slow reaction producing *TrRN AP*. Thus RBS^* presents unfrequent bursts of activity leading to rapid production of *ElRib*. The increasing part of the burst does not last long, because the discrete component RBS^* rapidly switches back to zero by the reaction $RBS^* \to \varnothing$. Thus, in the continuous variables, after a steep (deterministic) increase of ElRib one observes a slower decrease. We have the case of a PDP with switching, where the steep increase event could be assimilated to a breakage (see Figure 5a). However, the timescales of the decreasing and increasing parts of the peak are not well separated. Thus, the break-

age approximation could be used only to obtain qualitative results.

To summarize, we have considered five models: the unreduced Markov jump model, two averaged reduced jump Markov models (the second model obtained after averaging steps 1 and 2, the third model after application of all three steps) and two PDP models with switching obtained from each of the previous two averaged jump Markov models. The three jump Markov models are simulated using the Gillespie algorithm, and the two hybrid models are simulated with the PDP algorithm.

Of course, the process is Markovian only between two cell cycle events that are under external command. Considering the external command, the process is semi-Markovian [29,70]. We could restore the Markovian framework, by considering a cell cycle variable that has periodic autonomous dynamics and triggers the transitions between the cell cycle stages.

In order to compare the performance of the models (in terms of time complexity) we have represented the total jump intensities for the first three models (exact SSA, and the two averaged models) as functions of time on a trajectory. The model that demands the least computer time is the one with the smallest jump intensity. In Figure 5b, we notice a decrease of several orders of magnitudes of the total intensity from the exact SSA model to the models obtained after the second and third averaging steps.

We can not use the same method to estimate the computation time of the PDP algorithm. Indeed, although it is true generally that the computation time increases with the number of discrete transitions, the number of operations to compute the deterministic parts depends on the deterministic solver. In order to reduce the number of operations per time point we have favored one-step schemes (although this is not important for very large systems, where the main difficulty is represented by the computation of the flow function). We also favored implicit stiff solvers that can function with large time steps, reducing computation time. All the calculations were performed by using ode23s solver of MATLAB (this is a one-step implicit stiff solver using Rosenbrock formula of order 2). A comparison of the times to generate a 20 cycles long trajectories with the different methods is given in Table 3. This table is rather illustrative of the advantages of various approximations. Averaging allows a tremendous reduction of the execution time at least 50 times for weak repression and 10^4 times for strong repression (when stochastic effects are strongest). The Kramers-Moyal expansion leading to PDPs produce an extra 2-fold decrease of the execution time, but this is true only for weak repression and after averaging of all rapid cycles. The

PDP model M4 with incomplete averaging of fast cycles has a very bad performance, that can be even worse than the non-averaged SSA. This can be explained by the strong stiff character of the deterministic dynamics with steep ElRib peaks.

To test the accuracies of various approximations we have computed the steady probability distribution of the protein and of the folded protein using trajectories generated by the five models. The distribution obtained by SSA for the un-reduced model is used as the "exact" reference. The observed errors are the consequences of less good separation between systems constants. For instance, in Figure 4, $k_7 = 0.5$ and $k_{m6} = 2.25$, while in theory we need $k_7 \ll k_{m6}$. However, the approximate models are qualitatively correct. All models render correctly the bursting behavior of the system.

Another advantage of a simplified model is the reduced number of parameters. The full SSA model has 14 parameters. After the first two averaging steps only 9 parameters remain, and after the three averaging steps only 7. The PDPs have the same number of parameters as the corresponding averaged Markov jump processes, namely 9 and 7 parameters. Furthermore, the parameters of the simplified models are monomials of parameters of the unreduced model. This can be used for sensitivity analysis. After identification of the monomials that are critical for a given property the backtracked in uence of the initial parameters is given by the power of the corresponding parameter in the critical monomial. The details of the method can be found in [36].

Conclusion

We have presented, in a unified framework, various hybrid simplifications of stochastic biochemical models. These simplifications are based on partial Kramers-Moyal expansions and on averaging.

The use of simplified models in stochastic studies of cellular processes has several advantages.

The first advantage of simplified models is the reduction of computational time. We have shown that cycle averaging leads to the most drastic reduction of the computation time. This averaging algorithm, that can be used independently of the Kramers-Moyal expansion, represents a general preconditioning method for both stochastic and deterministic simulations. In stochastic studies, it reduces the number of discrete events to be simulated. In deterministic studies, it produces less stiff systems. The preconditioned models can be the starting point for other reduction methods.
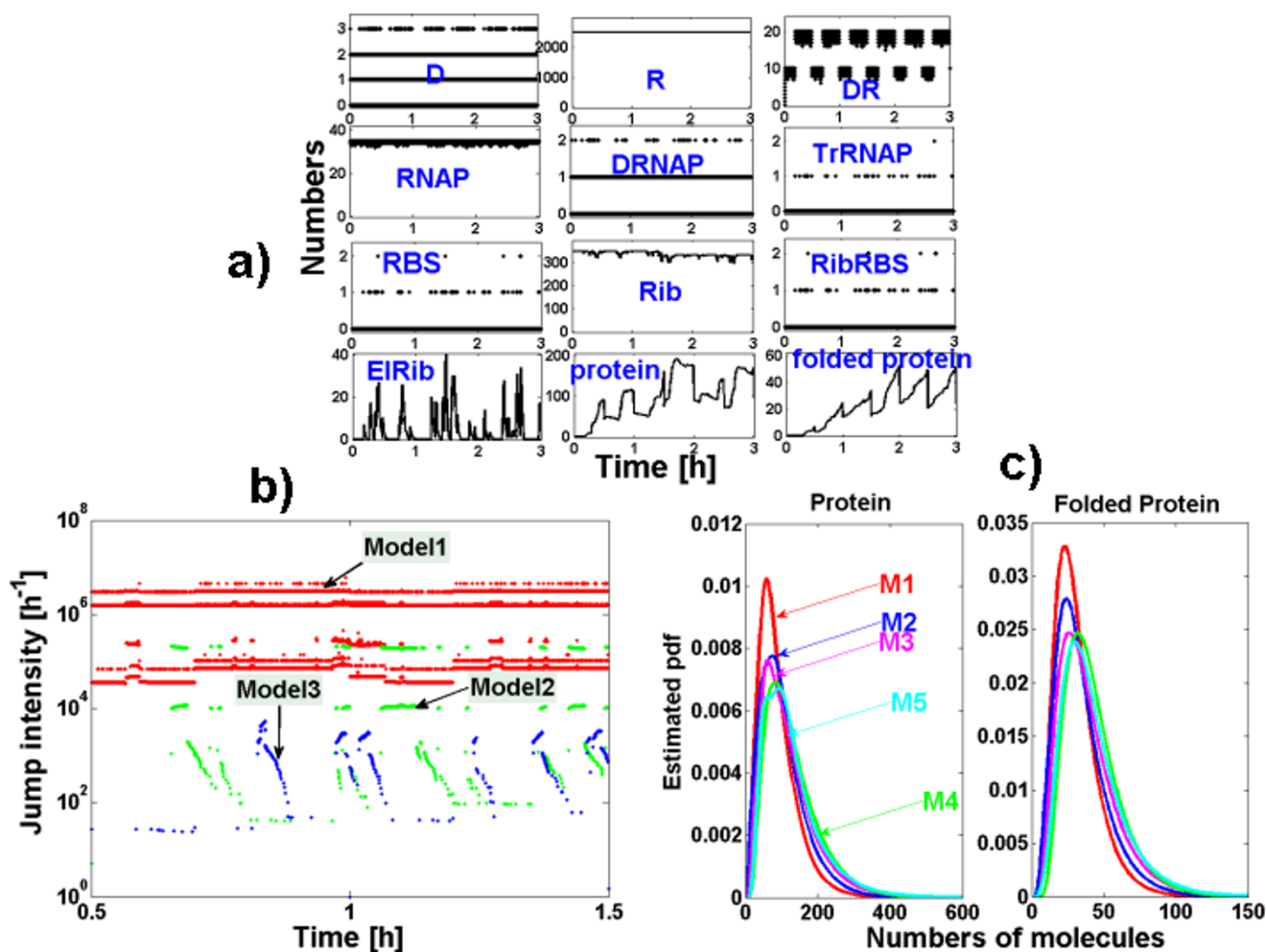
Figure 5
Repressed bacterium operon. a) Trajectories for *R* = 2500. b) Jump intensities for three SSA versions of the repressed bacterium operon. Model 1 (without averaging), Model 2 (averaging of *D*, *D.R*, *D.RN AP*), Model 3 (averaging of *D*, *D.R*, *D.RN AP*, *RBS*, *Rib.RBS*). c) Steady probability distributions for the protein and folded protein obtained with the five versions of the model, for *R* = 2500. *M*3 and *M*4 are PDP obtained from *M*2 and *M*3, respectively.

Table 3: Execution times for the repressed bacterium operon model.

R	M1 - SSA	M2 - averaging + SSA	M3 - averaging + SSA	M4 - PDP ode23s	M5 - PDP ode23s
50	$7.4 \ 10^3$	$2.9 \ 10^3$	$1.4 \ 10^2$	$1.1 \ 10^4$	$6.7 \ 10^1$
100	$6.8 \ 10^3$	$6.9 \ 10^2$	$3.5 \ 10^1$	$3.7 \ 10^3$	$3.6 \ 10^1$
200	$5.4 \ 10^3$	$1.7 \ 10^2$	$1.3 \ 10^1$	$1.6 \ 10^3$	$1.9 \ 10^1$
300	$4.9 \ 10^3$	$8.1 \ 10^1$	7.6	$8.0 \ 10^2$	$1.4 \ 10^1$
2500	$3.7 \ 10^3$	2.3	$5.2 \ 10^{-1}$	$1.0 \ 10^2$	4.1
5000	$3.6 \ 10^3$	$8 \ 10^{-1}$	$1.3 \ 10^{-1}$	$4.7 \ 10^1$	2.9
10000	$3.6 \ 10^3$	$1 \ 10^{-1}$	$2.1 \ 10^{-2}$	$3.1 \ 10^1$	2.0

The time, expressed in [s] correspond to generating a 20 division cycles long trajectory ($3.6 \ 10^4$ s biological time) for models M1-M5. The values represent log-averages for 10 launches. Model M4 is obtained from M2 and model M5 is obtained from M3, via the Kramers-Moyal expansion.

Mathematically, our simplified models are weak approximations of the fully detailed jump Markov processes. This means that all statistical properties of the trajectories of the full model including steady distributions, transition times, etc. should be rendered without significant loss of accuracy by the simplified models. Of course, after the reduction procedure, some variables and reactions may disappear and some resulting parameters are synthetic. Because the correspondence with the original variables and parameters is known, it is always possible to go back to the details of the full model. In particular, the identification of the critical parameters of the reduced model allows to backtrack the critical parameters of the full model. The Kramers-Moyal expansion, leading to hybrid simplifications, produced only a moderate decrease of the computation time. This limitation was partly due to our choice of low to medium size models with rather small numbers of molecules and with simple dynamics of the continuous variables. More obvious computational advantage can be obtained for more complex models. Particularly, this could be the case for models with oscillating dynamics of the continuous variables, such as molecular clocks.

The second advantage of simplified models lies in the understanding that these bring with respect to the stochastic properties of the system. Averaging and hybrid simplifications unravel the origin of noise in multiscale biochemical systems. Noise is generated at the microscopic level of discrete variables and transferred to the mesoscopic level of the the continuous variables. In this transfer, the topology of the reaction network plays a role, but also other properties are important such as the hierarchy of timescales of the system. Our simplification algorithm explains how and when hierarchies of cycles lead to intermittence and noise in gene networks. In many cases, important properties such as the stationary distribution, noise amplitude, waiting times between successive bursts can be analytically calculated for the hybrid simplifications.

We have discussed several types of hybrid approximations: piecewise deterministic and hybrid diffusions with switching, with breakage, and singular switching that can be assimilated to breakage. Taken together, these results offer a rather complete panorama of various intermittence phenomena in biochemical systems.

Authors' contributions
OR proposed the methodology to obtain hybrid simplifications. AC and OR chose the examples and implemented the algorithms. AC, AD and OR contributed to the mathematical aspects of the methods part. All authors drafted, read and approved the final manuscript.

Additional material

Additional file 1
Additional material. This file contains mathematical proofs of some results concerning averaging and singular switching. It also contains a description of the methods to estimate steady probability density function for PDPs.
Click here for file
[http://www.biomedcentral.com/content/supplementary/1752-0509-3-89-S1.pdf]

References
1. Kramers H: **Brownian motion in a field of force and the diffusion model of chemical reactions.** *Physica A* 1940, **7**:284-304.
2. Moyal J: **Stochastic Processes and Statistical Physics.** *J R Stat Soc London* 1949, **Ser.B 11**:150-210.
3. Risken H: *The Fokker-Planck equation: Methods of Solution and Applications* Berlin: Springer; 1989.
4. Ozbudak E, Thattai M, Kurtser I, Grossman A, van Oudenaarden A: **Regulation of noise in the expression of a single gene.** *Nature Genet* 2002, **31**:69-73.
5. Swain P, Elowitz M, Siggia E: **Intrinsic and extrinsic contributions to stochasticity in gene expression.** *Proc Natl Acad Sci USA* 2002, **99**:12795-12800.
6. Kaufmann BB, Yang Q, Mettetal JT, van Oudenaarden A: **Heritable Stochastic Switching Revealed by Single-Cell Genealogy.** *Plos Biology* 2007, **5**:1973-1980.
7. Yu J, Xiao J, Ren X, Lao K, Xie XS: **Probing Gene Expression in Live Cells, One Protein Molecule at a Time.** *Science* 2006, **311**:1600-1603.
8. Delbrück M: **Statistical Fluctuations in Autocatalytic Reactions.** *J Chem Phys* 1940, **8**:120-124.
9. Cai L, Friedman N, Xie X: **Stochastic protein expression in individual cells at the single molecule level.** *Nature* 2006, **440(7082)**:358-362.
10. Kaern M, Elston TA, Blake WJ, Collins JJ: **Stochasticity in gene expression: from theories to phenotypes.** *Nature Rev Genet* 2005, **6**:451-464.
11. Krishna S, Banerjee B, Ramakrishnan T, Shivashankar G: **Stochastic simulations of the origins and implications of long-tailed distributions in gene expression.** *PNAS* 2005, **102**:4771-4776.
12. Paulsson J: **Summing up the noise in gene networks.** *Nature* 2004, **427**:415-418.
13. Warren P, Tanase-Nicola S, Wolde P: **Exact results for noise power spectra in linear biochemical reaction networks.** *J Chem Phys* 2006, **125(14)**:144904.
14. Kierzek A, Zaim J, Zielenkiewicz P: **The Effect of Transcription and Translation Initiation Frequencies on the Stochastic Fluctuations in Prokaryotic Gene Expression.** *J Biol Chem* 2001, **276**:8165-8172.
15. Gillespie DT: *J Comput Phys* 1976, **22**:403.
16. Tian T, Burrage K: **Binomial leap methods for simulating stochastic chemical kinetics.** *The Journal of Chemical Physics* 2004, **121**:10356.
17. Gillespie D, Petzold L: **Improved leap-size selection for accelerated stochastic simulation.** *The Journal of Chemical Physics* 2003, **119**:8229.
18. Haseltine EL, Rawlings JB: **Approximate simulation of coupled fast and slow reactions for stochastic chemical kinetics.** *J Chem Phys* 2002, **117**:6959-6969.

19. Ball K, Kurtz TG, Popovic L, Rempala G: **Asymptotic analysis of multiscale approximations to reaction networks.** *Ann Appl Probab* 2006, **16:**1925-1961.
20. Alfonsi A, Cances E, Turinici G, Di Ventura B, Huisinga W: **Adaptive simulation of hybrid stochastic and deterministic models for biochemical systems.** *ESAIM Proceedings* 2005, **14:**1-13.
21. Alfonsi A, Cancès E, Turinici G, Di Ventura B, Huisinga W: **Exact simulation of hybrid stochastic and deterministic models for biochemical systems.** *Research Report RR-5435, INRIA* 2004 [http://hal.inria.fr/inria-00070572/en/].
22. Stein R, Gossen E, Jones K: **Neuronal variability: noise or part of the signal?** *Nature Reviews Neuroscience* 2005, **6(5):**389.
23. Rudiger S, Shuai J, Huisinga W, Nagaiah C, Warnecke G, Parker I, Falcke M: **Hybrid Stochastic and Deterministic Simulations of Calcium Blips.** *Biophysical Journal* 2007, **93(6):**1847.
24. Cook DL, Gerber AN, Tapscott SJ: **Modeling stochastic gene expression: Implications for haploinsufficiency.** *Proc Natl Acad Sci USA* 1998, **95:**15641-15646.
25. Boxma O, Kaspi H, Kella O, Perry D: **On/Off Storage Systems with State-Dependent Input, Output and Switching Rates.** *Probability in the Engineering and Informational Sciences* 2005, **19:**1-14.
26. Ghosh M, Bagchi A: **Modeling stochastic hybrid systems.** *System Modeling and Optimization* 2005, **166:**269-280.
27. Pola G, Bujorianu M, Lygeros J, Di Benedetto M: **Stochastic hybrid models: An overview.** *Proceedings IFAC Conference on Analysis and Design of Hybrid Systems* 2003.
28. Bujorianu M, Lygeros J: **General stochastic hybrid systems: Modelling and optimal control.** *Proc 43th Conference in Decision and Control* 2004.
29. Radulescu O, Muller A, Crudu A: **Théorèmes limites pour des processus de Markov à sauts. Synthèse des resultats et applications en biologie moleculaire.** *Technique et Science Informatique* 2007, **26:**443-469.
30. Zeiser S, Franz U, Wittich O, Liebscher V: **Simulation of genetic networks modelled by piecewise deterministic Markov processes.** *Systems Biology, IET* 2008, **2(3):**113-135.
31. Gillespie DT: **The Chemical Langevin equation.** *J Chem Phys* 2000, **113:**297-306.
32. Bogoliubov NN, Mitropolski YA: *Asymptotic Methods in the Theory of Nonlinear Oscillations* New York: Gordon and Breach; 1961.
33. Givon D, Kupferman R, Stuart A: **Extracting macroscopic dynamics: model problems and algorithms.** *Nonlinearity* 2004, **17:**R55-R127.
34. Acharya A, Sawant A: **On a computational approach for the approximate dynamics of averaged variables in nonlinear ODE systems: Toward the derivation of constitutive laws of the rate type.** *J Mech Phys Sol* 2006, **54:**2183-2213.
35. Yin G, Zhang Q, Badowski G: **Singularly Perturbed Markov Chains: Convergence and Aggregation.** *Journal of Multivariate Analysis* 2000, **72(2):**208-229.
36. Radulescu O, Gorban AN, Zinovyev A, Lilienbaum A: **Robust simplifications of multiscale biochemical networks.** *BMC Systems Biology* 2008, **2:**86.
37. Mastny E, Haseltine E, Rawlings J: **Two classes of quasi-steady-state model reductions for stochastic kinetics.** *The Journal of Chemical Physics* 2007, **127:**094106.
38. Van Kampen N: *Stochastic processes in physics and chemistry* third edition. Amsterdam: North Holland; 2007.
39. Crudu A, Debussche A, Muller A, Radulescu O: **Hybrid weak limits and averaging for multiscale stochastic gene networks.** in press.
40. Davis M: *Markov Models and Optimization* London: Chapman and Hall; 1993.
41. Ethier SN, Kurtz TG: *Markov Processes* New York: John Wiley & Sons; 1986.
42. Barucha-Reid A: *Elements of the Theory of Markov Processes and their Applications* New York: McGraw-Hill Book Co; 1960.
43. Ikeda N, Watanabe S: *Stochastic differential equations and diffusion processes* Amsterdam: North-Holland.
44. Gillespie DT: **Exact stochastic simulation of coupled chemical reactions.** *The Journal of Physical Chemistry* 1977, **81:**2340-2361.
45. Gorban AN, Karlin IV: *Invariant manifolds for physical and chemical kinetics, Lect Notes Phys 660* Berlin, Heidelberg: Springer; 2005.
46. Allain M: **Approximation par un processus de diffusion, des oscillations, autour d'une valeur moyenne, d'un processus de Markov de saut pur.** *C R Acad Sc Paris* 1976, **t.282:**891-894.

47. Kurtz TG: **Limit theorems for sequences of jump Markov processes approximating ordinary differential processes.** *J Appl Prob* 1971, **8:**344-356.
48. Surovtsova I, Sahle S, Pahle J, Kummer U: **Approaches to complexity reduction in a systems biology research environment (SYCAMORE).** *Proceedings of the 37th conference on Winter simulation, Winter Simulation Conference* 2006:1683-1689.
49. Salis H, Sotiropoulos V, Kaznessis Y: **Multiscale Hy3S: hybrid stochastic simulation for supercomputers.** *BMC Bioinformatics* 2006, **7:**93.
50. Griffith M, Courtney T, Peccoud J, Sanders W: **Dynamic partitioning for hybrid simulation of the bistable HIV-1 transactivation network.** *Bioinformatics* 2006, **22(22):**2782.
51. Gorban AN, Radulescu O: **Dynamical robustness of biological networks with hierarchical distribution of time scales.** *IET Systems Biology* 2007, **1:**238-246.
52. Gorban AN, Radulescu O: **Dynamic and static limitation in multiscale reaction networks, revisited.** In *Advances in Chemical Engineering: Mathematics and Chemical Engineering and Kinetics Volume 34.* Edited by: Marin G, West D, Yablonsky G. Academic Press; 2008:103-173.
53. Arnold V: *Supplementary chapters to the theory of ordinary differential equations* Moscow: MIR; 1978.
54. Sanders J, Verhulst F: *Averaging methods in nonlinear dynamical systems* New York: Springer; 1985.
55. Artstein Z: **Averaging of time-varying differential equations revisited.** *Journal of Differential Equations* 2007, **243(2):**146-167.
56. Freidlin M: *Markov processes and differential equations: asymptotic problems* Basel: Birkhauser; 1996.
57. Yin G, Zhang Q, Yang H, Yin K: **Discrete-time dynamic systems arising from singularly perturbed Markov chains.** *Nonlinear Analysis of Theory Methods and Applications* 2001, **47:**4763-4774.
58. Auger P, de la Para RB, Poggiale JC, Sanchez E, Huu TN: **Aggregation of variables and applications to population dynamics.** In *Structured Population Models in Biology and Epidemiology, LNM 1936, Mathematical Biosciences Subseries* Edited by: Magal P, Ruan S. Berlin: Springer; 2008:209-263.
59. Radulescu O, Gorban A: **Limitation and averaging for deterministic and stochastic biochemical reaction networks.** *International Workshop Model Reduction in Reacting Flow, Notre Dame, unpublished proceedings* 2009 [http://cam.nd.edu/upcoming-conferences/spring2009/talk%20_abstracts/radulescu_abstract.pdf].
60. Karmarkar R, Bose I: **Graded and binary responses in stochastic gene expressions.** *Phys Biol* 2004, **1:**197-204.
61. Stein R: **Some models of neuronal variability.** *Biophysical Journal* 1967, **7:**37-68.
62. Tuckwell H: *Stochastic processes in the neurosciences* Philadelphia: Society for Industrial and Applied Mathematics; 1989.
63. Kierzek A, Zaim J, Zielenkiewicz P: **The Effect of Transcription and Translation Initiation Frequencies on the Stochastic Fluctuations in Prokaryotic Gene Expression.** *Journal of Biological Chemistry* 2001, **276(11):**8165-8172.
64. Krishna S, Banerjee B, Ramakrishnan T, Shivashankar G: **Stochastic simulations of the origins and implications of long-tailed distributions in gene expression.** *Proceedings of the National Academy of Sciences* 2005, **102(13):**4771-4776.
65. Friedman N, Cai L, Xie X: **Linking stochastic dynamics to population distribution: an analytical framework of gene expression.** *Physical review letters* 2006, **97(16):**168302.
66. Reinitz J, Vaisnys J: **Theoretical and experimental analysis of the phage lambda genetic switch implies missing levels of cooperativity.** *J Theor Biol* 1990, **145(3):**295-318.
67. Arkin A, Ross J, McAdams HH: **Stochastic Kinetic Analysis of Developmental Pathway Bifurcation in Phage λ-infected *Escherichia Coli* Cells.** *Genetics* 1998, **149:**1633-1648.
68. Hasty J, Pradines J, Dolnik M, Collins J: **Noise-based switches and amplifiers for gene expression.** *Proc Natl Acad Sci USA* 2000, **97:**2075-2080.
69. Tian T, Burrage K: **Bistability and switching in the lysis/lysigeny genetic regulatory network of bacteriophage λ.** *J Theor bio* 2004, **227:**229-237.
70. Korolyuk V, Swishchuk A: *Semi-Markov Random Evolutions* Dordrecht: Kluwer; 1995.
71. Gorban AN, Radulescu O: **Dynamical robustness of biological networks with hierarchical distribution of time scales.** *IET Systems Biology* 2007, **1:**238-246.