



# Symmetry-based model reduction for approximate stochastic analysis

Kirill Batmanov, Celine Kuttler, François Lemaire, Cédric Lhoussaine,  
Cristian Versari

## ► To cite this version:

Kirill Batmanov, Celine Kuttler, François Lemaire, Cédric Lhoussaine, Cristian Versari. Symmetry-based model reduction for approximate stochastic analysis. Computational Methods in Systems Biology 2012 (CMSB 2012), Oct 2012, London, United Kingdom. pp.49-68. hal-00713386

**HAL Id: hal-00713386**

**<https://hal.science/hal-00713386>**

Submitted on 25 Jul 2012

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Symmetry-based model reduction for approximate stochastic analysis

Kirill Batmanov<sup>1\*</sup>, Celine Kuttler<sup>1</sup>, Francois Lemaire<sup>1</sup>, Cedric Lhoussaine<sup>1</sup>,  
and Cristian Versari<sup>1,2</sup>

<sup>1</sup> LIFL (CNRS UMR 8022), University of Lille 1, France

<sup>2</sup> INRIA, Lille, France

**Abstract.** For models of cell-to-cell communication, with many reactions and species per cell, the computational cost of stochastic simulation soon becomes intractable. Deterministic methods, while computationally more efficient, may fail to contribute reliable approximations for those models. In this paper, we suggest a reduction for models of cell-to-cell communication, based on symmetries of the underlying reaction network. To carry out a stochastic analysis that otherwise comes at an excessive computational cost, we apply a moment closure (MC) approach. We illustrate with a community effect, that allows synchronization of a group of cells in animal development. Comparing the results of stochastic simulation with deterministic and MC approximation, we show the benefits of our approach. The reduction presented here is potentially applicable to a broad range of highly regular systems.

**Keywords:** model reduction, stochastic analysis, moment closure, model symmetry, cell-to-cell communication, community effect

## 1 Introduction

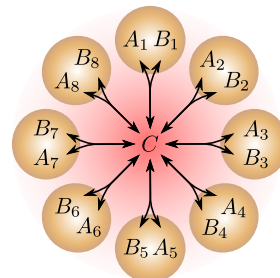
The dynamics of biochemical reaction systems are traditionally formalized as systems of ordinary differential equations (ODEs), whose variables represent concentrations of molecular species in a well-mixed solution. This assumes that the inherent stochastic fluctuations are negligible. However, this assumption is invalid for certain systems, such as gene regulatory networks [7]. Those systems must be analyzed stochastically, accounting for randomness of biomolecular interactions.

Exact solutions for the dynamics of most non-linear chemical systems are practically impossible to obtain. Various methods for *approximate stochastic analysis* have been suggested: Monte Carlo sampling of probability density functions of species' counts over time, known as Gillespie's algorithm [11]; approximations of this sampling [4,19]; explicit treatment of fluctuations with stochastic differential equations [12]; consideration of subspace of system states with highest probability mass [6,24]; aggregation of states [17].

---

\* [kirill.batmanov@lifl.fr](mailto:kirill.batmanov@lifl.fr)

**Fig. 1.** Symmetry in cell-to-cell communication:  $n$  cells with equal intracellular reaction network, involving molecules  $A$  and  $B$ , interact through the exchange of the extracellular molecule  $C$ .



*Moment closure* (MC) is a promising method for approximate analysis of the behavior of stochastic systems. It allows efficient calculation of approximate dynamics of *moments* of random variables associated with the system under study. MCs were successfully used in the fields of ecology [34], demographics [14], epidemiology [15] and statistical physics [21]. Traditionally they were derived manually. Recently, methods for automatic MC derivation were proposed [10,18,33] for biochemical reaction systems.

Because MC provides a system of ordinary differential equations, it allows computing the approximate dynamics more efficiently than any of the previously cited approaches to stochastic approximation. In particular, it relatively quickly yields solutions for different parameter values, such as reaction rates and initial conditions. This property has been exploited for efficient parameter estimation [22], and can possibly be used for other tasks, such as real-time control [32]. A disadvantage of the MC method is that the number of generated ODEs quickly grows with the system size, making it potentially difficult to scale to larger systems. *Reduction techniques* are needed to keep the analysis tractable.

In this work, we exploit *symmetries in cell signaling* to perform model reduction. Consider a pool of identical cells communicating over a short distance through the exchange of molecules, which are released by one cell, then diffuse and make contact with another cell. The reaction set

$$r_i : \quad A_i + B_i \rightleftharpoons C \quad , \text{for } i \in \{1 \dots n\} \quad (1)$$

describes such a system with  $n$  cells. Its symmetry is illustrated in Figure 1. A reaction between a pair of  $A$  and  $B$ , within the  $i^{\text{th}}$  cell, results in a  $C$ , which is expelled to the extracellular medium. Note that the extracellular  $C$  lacks a positional index, unlike the other molecules.  $C$  can migrate back from the extracellular medium to *any* of the  $n$  cells. The symmetry of this minimal system clearly appears, with  $C$  as the center, around which the  $n$  equal cells gather, and through which they communicate. Our model reduction strategy uses a notion of symmetry based on *invariance under certain changes of the chemical reaction network*. Intuitively, we observe that the global dynamics of the system remains invariant as we swap cell indices, because all cells are equal.

Symmetries in cell-to-cell communication are widespread, going far beyond the illustrative example of Figure 1. In bipartite bacterial communities, inter-species exchange of metabolites can enable important metabolic functions, that

are not reached by either of the isolated species [30]. One can distinguish between different types of metabolic interaction, the mechanisms of which remain under debate [35]. However, symmetries in the reaction network for cell-to-cell communication are a common feature to most of them. A last prominent example is the one considered by A. Turing in [31]: a ring of identical cells with rotational symmetries.

*The community effect* is thought to be a widespread phenomenon in animal development [2,5,16]. It allows a cell population within an embryo to forge a common identity, that is, to express a common set of genes. This synchronization is based on cell-to-cell communication, in which cells produce and exchange a diffusible molecule, resembling  $C$  in (1). Only when the cell population exceeds a critical number  $n_c$ , the common gene expression is maintained over extended periods of time. With fewer cells, after an initial induction, gene expression soon ceases.

A recent model for the community effect in *Xenopus laevis* [28], detailing on intracellular cascade of gene expression, mediated to intercellular communication, requires 17 reactions per cell. The community effect threshold  $n_c$  is about a hundred cells. Analysis of this model requires stochastic simulation over a wide range of parameters, cell numbers, and over extended periods of time, in order to yield realistic results, and namely to determine the critical number. The computational cost for the stochastic simulation becomes intractable, due to the multiplication of numbers of reactions per cell and numbers of cells. Deterministic approximations, on the other hand, do not provide precise predictions. The model reduction technique in combination with MC, as presented in this paper, provides a more convincing stochastic approximation of this model. Comparison of the solution of the truncated moment ODEs with stochastic simulations shows that MC is significantly closer to the stochastic dynamics than the deterministic solution.

*Related work.* Various symmetries have been exploited previously to facilitate finding the solutions of the biochemical models and to infer their properties. In [25], results from group theory are applied to the analytical solution of a simple model of gene expression, demonstrating how interesting properties of the model follow from continuous symmetries. In [13], symmetrically connected cell networks are considered, and some properties of them are shown in the deterministic regime. In [3], a model reduction technique is presented, which is based on particular kinds of symmetries expressed in Kappa language. That model reduction is applicable either to the deterministic approximation or, in fewer cases, to the stochastic semantics; the stochastic version of that reduction is not applicable to the community effect model. Here we present a different method, which provides a way to reduce higher order approximations (MC), under a different set of assumptions.

*Paper outline.* Section 2 reviews MC, Section 3 shows symmetry-based reduction, Section 4 applies our approach to a community effect model, and Section 5 concludes.

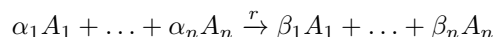
## 2 Approximate Stochastic Analysis

In this section we assume that the chemical reactions follow the *mass action* law, the first discovered and widely used kinetic law. Therefore the kinetics of a reaction can be described by a single parameter  $r$ , called *rate constant*, as well as stoichiometric coefficients.

For a system of  $n$  chemical species  $A_1, \dots, A_n$ , define tuples

$$(r, \boldsymbol{\alpha}, \boldsymbol{\beta}) \in \text{Reacts} = \mathbb{R}_{>0} \times \mathbb{N}^n \times \mathbb{N}^n$$

such that  $(r, \boldsymbol{\alpha}, \boldsymbol{\beta})$  represents the reaction



In the following, we index chemical species with  $i$  and reactions with  $j$ . A system of  $k$  chemical reactions is a set  $\mathcal{R} = \{U_1, \dots, U_k\}$  of  $k$  tuples  $(r, \boldsymbol{\alpha}, \boldsymbol{\beta})$ . For this set, we define the reactant stoichiometric matrix  $R$ , the product stoichiometric matrix  $P$ , and the vector of rate constants  $\mathbf{r}$  such that:

$$\begin{aligned} \forall U_j \in \mathcal{R} : \text{denote } (r, \boldsymbol{\alpha}, \boldsymbol{\beta}) = U_j : \\ r_j = r, \forall i : R_{ij} = \alpha_i, P_{ij} = \beta_i \end{aligned}$$

so that a chemical system is fully determined by the tuple  $(\mathbf{r}, R, P)$ .

*Example 1.* Given the 3 molecular species  $A$ ,  $B$  and  $C$ , the tuple

$$\left( (0.1 \ 0.3), \begin{pmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 0 & 1 \\ 0 & 1 \\ 1 & 0 \end{pmatrix} \right) \text{ represents the reactions } \begin{cases} A + B \xrightarrow{0.1} C \\ C \xrightarrow{0.3} A + B \end{cases}$$

### 2.1 Deterministic dynamics: the coarser approximation

The computational analysis of chemical reaction systems often assumes deterministic behavior. Following the mass action law, at any given time  $t$  a system's dynamics is driven by the *concentrations* of the reacting chemical species. For example, given some initial condition, the dynamics of Example 1 would be described by a set of ordinary differential equations

$$\frac{d[A]}{dt} = \frac{d[B]}{dt} = -0.1 [A][B] + 0.3 [C] \quad \frac{d[C]}{dt} = 0.1 [A][B] - 0.3 [C] \quad (2)$$

where  $[X]$  denotes the concentration of the species  $X$ . Despite the fact that concentrations of species can only assume discrete values, the deterministic approach is often justified by arguing that the high number of molecules usually

present in a solution make these discrete values so close that continuous domains for concentrations constitute a safe approximation. The corresponding mathematical argument is that the limit behavior of the system of Example 1 is exactly described by (2) when the volume of the system *tends to infinity* while the concentrations remain the same.

Formally, given a system of chemical reactions  $(\mathbf{r}, R, P)$ , its state space under the deterministic assumption is the space of the concentrations of the chemical species  $A_1, \dots, A_n$  in the system. Concentration of every species is assumed to be a non-negative real number. Its dynamics is defined, for a given initial state  $\boldsymbol{\nu}_0$ , by the function of time

$$\mathbf{x}(t) : \mathbb{R}_{\geq 0} \rightarrow \mathbb{R}_{\geq 0}^n, \quad \mathbf{x} \text{ is the solution of the ODE } \begin{cases} \dot{\mathbf{x}} = (P - R)\boldsymbol{\lambda} \\ \mathbf{x}(0) = \boldsymbol{\nu}_0 \end{cases}$$

where  $\boldsymbol{\lambda}$  is the vector of reaction rate laws in state  $\mathbf{x}(t)$ , that is, for reaction number  $j \in \{1, \dots, k\}$ , the rate law is  $\lambda_j = r_j \prod_i x_i^{R_{ij}}$ .

*Example 2.* For the system of Example 1, we have

$$\boldsymbol{\lambda} = \begin{pmatrix} 0.1x_1x_2 \\ 0.3x_3 \end{pmatrix} \text{ and } P - R = \begin{pmatrix} -1 & 1 \\ -1 & 1 \\ 1 & -1 \end{pmatrix}$$

leading to the system of ODEs

$$\begin{cases} \dot{x}_1 = 0.3x_3 - 0.1x_1x_2 \\ \dot{x}_2 = 0.3x_3 - 0.1x_1x_2 \\ \dot{x}_3 = 0.1x_1x_2 - 0.3x_3 \end{cases} \quad (3)$$

## 2.2 Stochastic Semantics

The deterministic assumption is often invalid at the cellular scale, where the number of molecules per species and cell can be low – one for genes, or a few for mRNA. In this case, it is preferable to formalize the stochastic behavior of the system in terms of a continuous time Markov chain (CTMC), whose states represent the different possible configurations of the system: each state is determined by the number of molecules per chemical species. Formally, we can define a CTMC as a collection  $\{\mathbf{x}(t) \mid t \in \mathbb{R}_{\geq 0}\}$  of  $n$  time-dependent *random variables* with state space in  $\mathbb{N}^n$ .

The system's evolution is then interpreted stochastically: at a given time  $t$ , a probability is assigned to each state and its variation in time is governed by a differential equation. The set of all such equations, one per state of the system, constitutes the *chemical master equation (CME)*, whose solution gives the complete information about the system's kinetics at any time. If  $\pi_{\boldsymbol{\nu}}(t) = Pr(\mathbf{x}(t) = \boldsymbol{\nu})$  is the probability of being in the state  $\boldsymbol{\nu}$  at time  $t$ , the CME becomes [33]:

$$\dot{\pi}_{\boldsymbol{\nu}} = \sum_{(r, \boldsymbol{\alpha}, \boldsymbol{\beta}) \in \mathcal{R}} r \binom{\boldsymbol{\nu} + \boldsymbol{\alpha} - \boldsymbol{\beta}}{\boldsymbol{\alpha}} \pi_{\boldsymbol{\nu} - \boldsymbol{\alpha} + \boldsymbol{\beta}} - r \binom{\boldsymbol{\nu}}{\boldsymbol{\alpha}} \pi_{\boldsymbol{\nu}} \quad (4)$$

where  $\binom{\mathbf{a}}{\mathbf{b}}$  denotes the product  $\prod_i \binom{a_i}{b_i}$ .

The number of states (and of corresponding differential equations in the master equation) scales exponentially with the number of species and possible number of molecules per species. When some reaction creates unbounded numbers of new molecules, it even becomes infinite (but countable). In this case, while it is still possible to find solutions of the CME for some particular systems, the general, automatic numeric solution of the CME becomes intractable. As a consequence, one can only obtain *partial or approximated information* on the CME’s solution.

### 2.3 Moments and moments calculation

A radically different perspective on a system is, instead of computing state probabilities, to directly consider the time evolution of the *moments* of its variables. Given a vector  $\mathbf{m} \in \mathbb{N}_{\geq 0}^n$ , the mixed moment  $\mu^{(\mathbf{m})}$  about zero (i.e. *uncentred*) of a subset of variables defined by non-zero elements of  $\mathbf{m}$  is

$$\mu^{(\mathbf{m})} : \begin{cases} \mathbb{R}_{\geq 0} \rightarrow \mathbb{R}_{\geq 0} \\ t \mapsto E[\mathbf{x}^{\mathbf{m}}(t)] \end{cases}$$

where  $E$  denotes the expectation, and  $\mathbf{x}^{\mathbf{m}}$  denotes a product  $\prod_i x_i^{m_i}$ . Therefore,  $\mu^{(1,0,\dots,0)}$  denotes, for instance, the dynamics of the expectation of the random variable  $x_1$ . The *order* of moment  $\mu^{(\mathbf{m})}$  is the sum of the indices  $\sum_i m_i$ .

This change of perspective arises when general characterizations of a model matter more than the probability of each single state. For example, one may only need to extract the average concentrations of species in time and their stochastic noise. Those correspond to the first two (central) moments: mean and variance.

The moment-based analysis of a chemical system requires a preliminary step, which is to replace the state-centric description of the dynamics given by the CME with a new description directly focused on the evolution of the value of moments in time. In practice, this corresponds to building a new set of differential equations, one for each moment, and can be performed in several ways. For example, when the rate functions associated with chemical reactions are polynomial, moment equations may be calculated using the moment-generating function [10], or equivalently adopting a generator operator [18] or by a probability generating function [33]. Extensions to more general cases may require more sophisticated methods, e.g. when rational rate functions are considered [23].

Although the model reduction presented in this paper is valid for models with kinetic laws of any kind, here we follow the method of moment generation presented in [33], which is reasonably simple and covers those kinetic laws needed for Section 4 (the mass-action family). According to this method, moment equations can be calculated thanks to the probability-generating function

$$\phi(\mathbf{z}, t) = E[\mathbf{z}^{\mathbf{x}(t)}] = \sum_{\nu} \pi_{\nu}(t) \mathbf{z}^{\nu} \quad (5)$$

where  $\mathbf{z}$  is a formal parameter consisting of a vector of variables  $(z_1, \dots, z_n)$ . The introduction of  $\mathbf{z}$  allows the calculation of moments of any order by applying properly the operation of partial differentiation to  $\phi$  with respect to the variables  $z_1, \dots, z_n$ , and then by setting their value to 1. For example, the expectation of  $x_1$  of Example 1 can be calculated by differentiating  $\phi$  once with respect to  $z_1$ , and then by setting  $\mathbf{z} = \mathbf{1} = (1, 1, 1)$ :

$$\begin{aligned}\phi_{z_1}(\mathbf{z}, t)|_{\mathbf{z}=\mathbf{1}} &= \left( \sum_{\nu} \nu_1 \pi_{\nu}(t) \mathbf{z}^{(\nu - (1, 0, 0))} \right) \Big|_{\mathbf{z}=\mathbf{1}} \\ &= \sum_{\nu} \nu_1 \pi_{\nu}(t) = E[x_1(t)] = \mu^{(1, 0, 0)}\end{aligned}$$

where  $\phi_{z_1}$  denotes the partial derivative  $\frac{\partial \phi}{\partial z_1}$ . The same procedure applies to calculate the expectation of  $x_2$  (respectively  $x_3$ ), where one has to differentiate with respect to  $z_2$  (respectively  $z_3$ ). Second order factorial moments can be calculated by differentiating twice with respect to the corresponding variables, from which the general (uncentred) moments can be obtained, for example:

$$\left( \frac{\partial^k \phi}{\partial z_1^k} \right) \Big|_{\mathbf{z}=\mathbf{1}} = E[x_1(t)(x_1(t) - 1) \dots (x_1(t) - k + 1)] = \mu_F^{(k, 0, 0)} = \sum_{i=1}^k s(k, i) \mu^{(i, 0, 0)}$$

where  $\mu_F$  is a factorial moment and  $s(k, i) = (-1)^{k-i} \binom{k}{i}$  is the Stirling's number of the first kind. This procedure generalizes to any order, so that univariate and multivariate higher order factorial moments are calculated by differentiating the proper number of times with respect to each variable in  $\mathbf{z}$ . Joint factorial moments can be expressed in terms of general moments as follows:

$$\frac{\partial^{\mathbf{m}} \phi}{\partial \mathbf{z}^{\mathbf{m}} \Big|_{\mathbf{z}=\mathbf{1}}} = \frac{\partial^{m_1}}{\partial z_1^{m_1}} \dots \frac{\partial^{m_n}}{\partial z_n^{m_n}} \Big|_{\mathbf{z}=\mathbf{1}} = \mu_F^{(\mathbf{m})} = \sum_{i_1=1}^{m_1} \dots \sum_{i_n=1}^{m_n} \mu^{(\mathbf{m})} \prod_{j=1}^n s(m_j, i_j) \quad (6)$$

From (4) and (5) it is possible to derive (see [8], (5.60) modified for Kurtz-type combinatorial mass action model used here) the following partial differential equation:

$$\phi_t(\mathbf{z}, t) = H\phi(\mathbf{z}, t) \quad (7)$$

where  $H$  is the so-called Hamiltonian operator, obtained by the stoichiometric coefficients and the reaction rates of the chemical system:

$$H = \sum_{(r, \alpha, \beta) \in \mathcal{R}} \frac{r}{\alpha!} (\mathbf{z}^{\beta} - \mathbf{z}^{\alpha}) \left( \frac{\partial}{\partial \mathbf{z}} \right)^{\alpha} \quad (8)$$

where  $\alpha!$  denotes the product  $\prod_i \alpha_i!$ .

The procedure for calculating moment equations summarizes as follows:

1. Determine the Hamiltonian operator for the reaction system by (8).



2. Calculate the moment equation for any desired moment  $\mu^{(\mathbf{m})} = \mu^{(m_1, \dots, m_n)}$  by applying the partial derivative  $\frac{\partial^{\mathbf{m}}}{\partial \mathbf{z}^{\mathbf{m}}}$  to both sides of equation (7) and setting  $\mathbf{z}$  to 1.
3. Convert higher order factorial moments in the equation to general moments using (6) (the first order factorial moments are equal to the general moments).

*Example 3.* The Hamiltonian for Example 1 is

$$H = 0.1 (z_3 - z_1 z_2) \frac{\partial^2}{\partial z_1 \partial z_2} + 0.3 (z_1 z_2 - z_3) \frac{\partial}{\partial z_3}$$

To get the differential equation for  $\mu^{(1,0,0)}$ , we first apply to (7) partial differentiation with respect to  $z_1$  and then set  $\mathbf{z} = \mathbf{1}$ :

$$\phi_{tz_1}(\mathbf{z}, t)|_{\mathbf{z}=\mathbf{1}} = \phi_{z_1 t}(\mathbf{z}, t)|_{\mathbf{z}=\mathbf{1}} = \dot{\mu}^{(1,0,0)}$$

The same procedure is applied to the r.h.s.:

$$\begin{aligned} \left( \frac{\partial}{\partial z_1} H \phi(\mathbf{z}, t) \right) |_{\mathbf{z}=\mathbf{1}} &= \left( \frac{\partial}{\partial z_1} \left( 0.1 (z_3 - z_1 z_2) \phi_{z_1 z_2}(\mathbf{z}, t) \right) \right) |_{\mathbf{z}=\mathbf{1}} + \\ &\quad \left( \frac{\partial}{\partial z_1} \left( 0.3 (z_1 z_2 - z_3) \phi_{z_3}(\mathbf{z}, t) \right) \right) |_{\mathbf{z}=\mathbf{1}} \\ &= \left( -0.1 z_2 \phi_{z_1 z_2}(\mathbf{z}, t) + 0.1 (z_3 - z_1 z_2) \phi_{z_1^2 z_2}(\mathbf{z}, t) \right) |_{\mathbf{z}=\mathbf{1}} + \\ &\quad \left( 0.3 z_2 \phi_{z_3}(\mathbf{z}, t) + 0.3 (z_1 z_2 - z_3) \phi_{z_1 z_3}(\mathbf{z}, t) \right) |_{\mathbf{z}=\mathbf{1}} \end{aligned}$$

By applying the property of the probability-generating function, we obtain:

$$\dot{\mu}^{(1,0,0)} = \left( \frac{\partial}{\partial z_1} H \phi(\mathbf{z}, t) \right) |_{\mathbf{z}=\mathbf{1}} = -0.1 \mu^{(1,1,0)} + 0.3 \mu^{(0,0,1)}$$

Similarly, one finds that

$$\dot{\mu}^{(0,1,0)} = -0.1 \mu^{(1,1,0)} + 0.3 \mu^{(0,0,1)} \quad (9)$$

$$\dot{\mu}^{(0,0,1)} = +0.1 \mu^{(1,1,0)} - 0.3 \mu^{(0,0,1)} \quad (10)$$

$$\begin{aligned} \dot{\mu}^{(1,1,0)} &= -0.1 \mu^{(2,1,0)} + 0.3 \mu^{(1,0,1)} - 0.1 \mu^{(1,2,0)} + \\ &\quad 0.3 \mu^{(0,1,1)} + 0.1 \mu^{(1,1,0)} + 0.3 \mu^{(0,0,1)} \end{aligned} \quad (11)$$

It is important to note that the moment equations are always linear, like those of the CME (4).

The implementation of analysis tools based on moments usually relies on libraries for symbolic computation that help to automate the calculation of moment equations explained above.

## 2.4 Moment closure

Switching from a state-based description to a moment-based one seemingly happens without any particular gain (or loss). Indeed, given a system with a finite number  $n$  of states, the number of (independent) equations in the CME is  $n - 1$ . Its characterization in terms of moment equations gives a different set of ODEs, but with the same number  $n - 1$  of (independent) equations. If the CME is defined by an infinite (countable) number of equations, the same holds for the corresponding system of moment equations. Mathematically, both descriptions contain the same information: switching back and forth between them is fully reversible. Practically, the high number of equations renders both systems equally intractable. However, information is differently distributed across the two ODE systems, such that different approximation techniques can be applied.

Let us point out a dependency in the structure of equation systems associated to moments: an  $m^{\text{th}}$  order moment generally depends on moments of order at most  $m + h$ , where  $h$  is a constant whose value depends on the stoichiometry matrix. In Example 3 we have  $h = 1$ , so that each moment of order  $m$  depends – besides lower and equal order moments – on some moment of order  $m + 1$ : for example, the expectation values of the number of molecules per chemical species  $\mu^{(1,0,0)}, \mu^{(0,1,0)}, \mu^{(0,0,1)}$  depend on the moment  $\mu^{(1,1,0)}$  of order two, which in turn depends on some moments of order three, and so on. In cases when we were only interested in the mean number per chemical species, we could confine our attention to first order moments. If we also wanted information on the stochastic noise, we could also consider second order moments (and so on: the higher the order considered, the more complete the information about the probability distribution of chemical species). The problem is then how to break the previously described *infinite cascade* of dependencies.

*Moment closure* denotes a wide set of techniques allowing to effectively break infinite cascades. The closure usually follows from some assumption on the probability distribution of chemical species: the closure of *order*  $m$ , for example, allows rewriting  $m^{\text{th}}$  order equations, such that they no longer depend on higher order moments. Breaking the cascade comes at the cost of introducing an approximation error: the more faithful is the assumption allowing the closure, the closer to the precise solution is the resulting set of equations.

*Example 4.* Assume the correlation between species  $A$  and  $B$  in Example 1, given time, is negligible. Then the covariance of  $x_1$  and  $x_2$  is zero at any time  $t$ . Thus,

$$\begin{aligned} 0 = \text{Cov}(x_1, x_2) &= E[(x_1 - \mu^{(1,0,0)})(x_2 - \mu^{(0,1,0)})] \\ &= E[x_1 x_2] - \mu^{(1,0,0)} \mu^{(0,1,0)} \end{aligned}$$

and from this it follows that

$$\mu^{(1,1,0)} = E[x_1 x_2] = \mu^{(1,0,0)} \mu^{(0,1,0)} \quad (12)$$

By applying (12) in (9)-(10), one obtains

$$\begin{aligned}\dot{\mu}^{(1,0,0)} &= -0.1 \mu^{(1,0,0)} \mu^{(0,1,0)} + 0.3 \mu^{(0,0,1)} \\ \dot{\mu}^{(0,1,0)} &= -0.1 \mu^{(1,0,0)} \mu^{(0,1,0)} + 0.3 \mu^{(0,0,1)} \\ \dot{\mu}^{(0,0,1)} &= +0.1 \mu^{(1,0,0)} \mu^{(0,1,0)} - 0.3 \mu^{(0,0,1)}\end{aligned}$$

so that equation (11) is no longer considered. Remarkably, the above set of closed moment equations corresponds exactly to the ODE system (3). This simple result supports a different interpretation of the deterministic approximation introduced with equation (2). Under this interpretation, the continuity of the domain of species concentrations is fully justified without introducing any limit behavior, because the “concentration” is instead thought as the expected value of the number of molecules of chemical species. Moreover, the set of ODEs describing the (approximate) evolution of expectations is not derived by the application of a limit involving the number of molecules: it follows as a direct consequence of the assumption of zero correlation between the species participating in higher-order reactions. Although the concentration limit and the zero correlation assumptions are inherently related, they lead to the same result through different mathematical procedures.

More interesting applications of moment closure are those where the deterministic approximation fails to capture the real behavior of the system. Here, higher order closures may be applied in order to get better quantitative approximations, as well as information about the stochasticity of the system under analysis. Important results in this direction are presented in [20] for the class of *zero central moment closures*, where the closure of order  $m$  is obtained by setting to zero all  $(m + 1)^{\text{th}}$  central moments. For this class, the approximation error has been proven to decrease as the order of the closure increases. Similar results have been shown in [29] for the first orders of another class of closures, obtained by a procedure called derivative matching.

Remarkably, the first in the class of zero central moment closures is the deterministic approximation applied in Example 4, therefore it is usually regarded as the coarsest among moment closure approximations.

*Normal closure.* The second closure of this class, one of the first to be applied and still widely used, is the so called *normal closure*, consistent with the assumption that the counts of all chemical species, at any time point, is jointly normally distributed. This assumption is obviously wrong: first, the support of a Gaussian distribution is continuous, while the probability distributions associated with chemical systems are discrete (in fact their support is given by the set of reachable states). Moreover, the support of normal distributions also includes negative values: in the context of chemical systems, this would correspond to allowing states with a negative number of molecules, which is clearly impossible. However, this assumption is one of the easiest to apply and works very well in many practical cases, including the study of the community effect model presented in Section 4.

Formally, the normal closure is obtained by setting to zero each central moment of order three. Given three random variables  $x_1, x_2, x_3$ , the closure follows by the equation

$$\begin{aligned} 0 &= E[(x_1 - E[x_1])(x_2 - E[x_2])(x_3 - E[x_3])] \\ &= E[x_1 x_2 x_3] + 2E[x_1]E[x_2]E[x_3] \\ &\quad - E[x_3]E[x_1 x_2] - E[x_2]E[x_1 x_3] - E[x_1]E[x_2 x_3] \end{aligned}$$

from which we get

$$\begin{aligned} E[x_1 x_2 x_3] &= E[x_3]E[x_1 x_2] + E[x_2]E[x_1 x_3] + \\ &\quad E[x_1]E[x_2 x_3] - 2E[x_1]E[x_2]E[x_3] \end{aligned} \quad (13)$$

*Example 5.* In order to apply (13) to (11), we must calculate (13) for the moments  $\mu^{(2,1,0)}$  and  $\mu^{(1,2,0)}$ :

$$\begin{aligned} \mu^{(1,2,0)} &= E[x_1 x_2^2] = 2 E[x_2]E[x_1 x_2] + E[x_1]E[x_2^2] - 2E[x_1]E[x_2]^2 \\ &= 2 \mu^{(0,1,0)} \mu^{(1,1,0)} + \mu^{(1,0,0)} \mu^{(0,2,0)} - 2 \mu^{(1,0,0)} \mu^{(0,1,0)^2} \\ \mu^{(2,1,0)} &= E[x_1^2 x_2] = 2 \mu^{(1,0,0)} \mu^{(1,1,0)} + \mu^{(0,1,0)} \mu^{(2,0,0)} - 2 \mu^{(0,1,0)} \mu^{(1,0,0)^2} \end{aligned}$$

By substituting in (11) we get

$$\begin{aligned} \dot{\mu}^{(1,1,0)} &= 0.3 \mu^{(1,0,1)} + 0.3 \mu^{(0,1,1)} + 0.1 \mu^{(1,1,0)} + 0.3 \mu^{(0,0,1)} \\ &\quad + 0.1 (2 \mu^{(0,1,0)} \mu^{(1,0,0)^2} - 2 \mu^{(1,0,0)} \mu^{(1,1,0)} - \mu^{(0,1,0)} \mu^{(2,0,0)}) \\ &\quad + 0.1 (2 \mu^{(1,0,0)} \mu^{(0,1,0)^2} - 2 \mu^{(0,1,0)} \mu^{(1,1,0)} - \mu^{(1,0,0)} \mu^{(0,2,0)}) \end{aligned}$$

so that  $\dot{\mu}^{(1,1,0)}$  no longer depends on third order moments. In order to eliminate their dependencies on third order moments, the same steps apply to the equations of any further moment of order two – including those of  $\mu^{(2,0,0)}, \mu^{(0,2,0)}, \mu^{(1,0,1)}$  and  $\mu^{(0,1,1)}$  in the above equation.

The resulting system of ODEs, typically non-linear after the closure, can be solved numerically. The initial values for the moments are usually given under the zero-variance assumption, that is  $\mu^{(\mathbf{m})}(0) = \mathbf{x}^{\mathbf{m}}(0)$ .

### 3 Model reduction based on symmetries

Moment closure of order  $m$  for a system with  $n$  species generates  $O(n^m)$  ODEs, because moments for all combinations of  $n$  species may be included. The equation system may thus become difficult to handle. Attempts were made to simplify models based on various properties, e.g. conservation laws and bounds on numbers of species [33]. Here we present a model reduction method based on *symmetries in the reaction set*, which can reduce the model dramatically in some cases.

### 3.1 Reduction by example

We demonstrate the idea by a simple example. Consider the set of  $p = 2n + 1$  chemical species  $A_1, \dots, A_n, B_1, \dots, B_n, C$  associated with state random variables  $\mathbf{x}(t) = (A_1(t), \dots, A_n(t), B_1(t), \dots, B_n(t), C(t))$  and the following set of reactions

$$\mathcal{R} = \{A_i + B_i \xrightleftharpoons[\kappa_2]{\kappa_1} C \mid i \in \{1, \dots, n\}\}$$

We assume that for any fixed time  $t$ ,  $\mathcal{R}$  and an initial state  $\mathbf{x}(0) = \boldsymbol{\nu}_0 \in \mathbb{N}^p$  defines a probability distribution over  $\mathbf{x}(t)$  with probability mass function  $\pi_{\boldsymbol{\nu}}(t)$ . For instance, considering combinatorial mass action kinetics,  $\pi_{\boldsymbol{\nu}}(t)$  will be the solution of the CME (4).

We want to identify moment equalities from simple *symmetries* in the reaction system. Since a moment is fully defined by the *marginal distributions* of the variables composing it, we actually identify equal marginal distributions from symmetries. The marginal distribution of the variables  $A_1$  and  $B_1$  is the probability distribution of this set of variables, ignoring the others. By symmetry of the reaction set, we mean that the reaction set remains *invariant under permutation of the chemical species*. An obvious permutation of this kind for  $\mathcal{R}$  is swapping  $A_1$  with  $A_2$  and  $B_1$  with  $B_2$ . In that case,  $\mathcal{R}$  remains unchanged. Suppose we further assume that the initial state is invariant with respect to the same permutation. That is, initial numbers of  $A_1$  and  $A_2$  are the same, as well as those of  $B_1$  and  $B_2$ . Because we consider probability distributions that are completely defined by the reactions  $\mathcal{R}$  and the initial state, the stochastic dynamics of the variable set  $\{A_1, B_1\}$  cannot be distinguished from that of  $\{A_2, B_2\}$ . As proved below, this means that their marginal distributions are equal

$$Pr(A_1(t) = a, B_1(t) = b) = Pr(A_2(t) = a, B_2(t) = b), \forall t \in \mathbb{R}_{\geq 0}, a, b \in \mathbb{N}$$

Permuting of  $C$  with itself, we get

$$\begin{aligned} Pr(A_1(t) = a, B_1(t) = b, C(t) = c) &= Pr(A_2(t) = a, B_2(t) = b, C(t) = c), \\ \forall t \in \mathbb{R}_{\geq 0}, a, b, c \in \mathbb{N} \end{aligned} \tag{14}$$

Importantly, this entails the moment equalities  $E[A_1^i B_1^j C^k] = E[A_2^i B_2^j C^k]$  for any  $i, j, k \geq 0$ . As another example of symmetry, assuming  $n \geq 4$  one can swap  $A_1$  with  $A_3$ ,  $A_2$  with  $A_4$ ,  $B_1$  with  $B_3$ ,  $B_2$  with  $B_4$ ,  $C$  with itself. Again, also assuming invariance of the initial state by this permutation, we have

$$\begin{aligned} Pr(A_1(t) = a_1, A_2(t) = a_2, B_1(t) = b_1, B_2(t) = b_2, C(t) = c) &= \\ Pr(A_3(t) = a_1, A_4(t) = a_2, B_3(t) = b_1, B_4(t) = b_2, C(t) = c), \\ \forall t \in \mathbb{R}_{\geq 0}, a_1, a_2, b_1, b_2, c \in \mathbb{N} \end{aligned} \tag{15}$$

It is straightforward to use equalities of the form (14)-(15) to reduce a set of moment equations of the system. For example, the system considered above

generates, among others, the following moment equations for second order moments:

$$\begin{aligned} \frac{dE[A_i C]}{dt} = & \kappa_1 \left( \sum_{j=1}^n E[A_i A_j B_j] \right) - \kappa_1 (E[A_i B_i] + E[A_i B_i C]) - \\ & \kappa_2 (E[C] - E[C^2] + n \cdot E[A_i C]) \quad i = 1 \dots n \end{aligned} \quad (16)$$

Using relations as (14), we can infer the following moment equalities:  $E[A_i C] = E[A_1 C]$ ,  $E[A_i B_i] = E[A_1 B_1]$ ,  $E[A_i B_i C] = E[A_1 B_1 C]$ ,  $E[A_i^2 B_i] = E[A_1^2 B_1]$ , and using equalities as (15) we have  $E[A_i A_j B_j] = E[A_1 A_2 B_2]$ ,  $i = 1 \dots n, j = 1 \dots n, i \neq j$ . Therefore we can equivalently rewrite all  $n$  equations in (16) into one:

$$\begin{aligned} \frac{dE[A_1 C]}{dt} = & \kappa_1 ((n-1)E[A_1 A_2 B_2] + E[A_1^2 B_1]) - \\ & \kappa_1 (E[A_1 B_1] + E[A_1 B_1 C]) - \\ & \kappa_2 (E[C] - E[C^2] + n \cdot E[A_1 C]) \end{aligned} \quad (17)$$

We can't exchange moments for  $A_i$  and  $B_i$  because they may have different initial conditions in general. Using this approach, the system of moment equations up to order two is reduced from  $2n^2 + 5n + 2$  to 11 ODEs for any  $n \geq 2$ . The rest of the equations are redundant and can be safely excluded. The transformation is exact, and we can recover the dynamics of the original system from the reduced one. In order to compute the moment dynamics, it is necessary to perform a closure of the reduced system as described in Section 2.4.

### 3.2 Formal reduction

We now formally define the previous notions. We however won't make use of marginal distributions, since equivalence of the full joint probability distribution entails equivalence of its marginal distributions.

We consider permutations  $\sigma$  over the set of species indices  $\{1, \dots, n\}$ . Permutations of vectors and reaction sets are defined as

$$\begin{aligned} \mathbf{a}_\sigma &= (a_{\sigma(1)}, \dots, a_{\sigma(n)}) \\ \mathcal{R}_\sigma &= \{(\kappa, \boldsymbol{\alpha}_\sigma, \boldsymbol{\beta}_\sigma) \mid (\kappa, \boldsymbol{\alpha}, \boldsymbol{\beta}) \in \mathcal{R}\} \end{aligned}$$

We say that a vector  $\mathbf{a}$ , resp. a reaction set  $\mathcal{R}$ , is  $\sigma$ -invariant, iff  $\mathbf{a} = \mathbf{a}_\sigma$ , resp.  $\mathcal{R} = \mathcal{R}_\sigma$ . We denote  $\mathcal{P}$  the function that, for a given set  $\mathcal{R}$  of reactions, initial state  $\boldsymbol{\nu}_0$  and time  $t$ , gives a probability distribution over the counts of the species with probability mass function  $\pi_\boldsymbol{\nu}(t)$ . Somehow  $\mathcal{P}$  provides the stochastic semantics of the reactions. For example,  $\mathcal{P}$  could be the stochastic semantics of the reactions (Section 2.2), or an approximation of it. We make the following assumption about  $\mathcal{P}$ .

**Assumption 1** Let  $\pi_\boldsymbol{\nu}(t) = \mathcal{P}(\mathcal{R}, \boldsymbol{\nu}_0, t)$  and  $\pi'_\boldsymbol{\nu}(t) = \mathcal{P}(\mathcal{R}_\sigma, \boldsymbol{\nu}_{0\sigma}, t)$ , for some reaction set  $\mathcal{R}$ , initial state  $\boldsymbol{\nu}_0$ , time  $t$ , and permutation  $\sigma$  of the species indices. For any  $\boldsymbol{\nu} \in \mathbb{N}^n$ , we have  $\pi_\boldsymbol{\nu}(t) = \pi'_\boldsymbol{\nu}(t)$ .

This assumption relates permutations at the level of reactions to permutation at the level of its stochastic semantics. It just states that the stochastic dynamics of a species  $A$  provided by  $\mathcal{P}$  does not depend on its position in the state vector. Saying it differently, we assume that the stochastic semantics is insensitive to species renaming, provided that this renaming doesn't create name conflicts. This is a reasonable assumption that is, for instance, satisfied by the master equation.

**Theorem 1.** *Let  $\mathcal{R}$  be a set of  $k$  reactions of  $n$  species,  $\boldsymbol{\nu}_0 \in \mathbb{N}^n$  be an initial state, and  $\sigma$  be a permutation over  $\{1, \dots, n\}$ . Let  $\pi_{\boldsymbol{\nu}}(t) = \mathcal{P}(\mathcal{R}, \boldsymbol{\nu}_0, t)$ , if  $\mathcal{R}$  and  $\boldsymbol{\nu}_0$  are  $\sigma$ -invariant, then, for any  $\boldsymbol{\nu} \in \mathbb{N}^n$ ,  $\pi_{\boldsymbol{\nu}}(t) = \pi_{\boldsymbol{\nu}_\sigma}(t)$ .*

*Proof.* This theorem is a straightforward consequence of the above assumption. Indeed, let  $\pi_{\boldsymbol{\nu}}(t) = \mathcal{P}(\mathcal{R}, \boldsymbol{\nu}_0, t)$  and  $\pi'_{\boldsymbol{\nu}}(t) = \mathcal{P}(\mathcal{R}_\sigma, \boldsymbol{\nu}_{0\sigma}, t)$ , since  $\mathcal{R} = \mathcal{R}_\sigma$  and  $\boldsymbol{\nu}_0 = \boldsymbol{\nu}_{0\sigma}$ , we have  $\pi_{\boldsymbol{\nu}}(t) = \pi'_{\boldsymbol{\nu}}(t)$ . By the Assumption 1 it follows that  $\pi_{\boldsymbol{\nu}}(t) = \pi'_{\boldsymbol{\nu}_\sigma}(t) = \pi_{\boldsymbol{\nu}_\sigma}(t)$ .  $\square$

**Corollary 1.** *Let  $\mathcal{R}$  be a set of  $k$  reactions of  $n$  species,  $\boldsymbol{\nu}_0 \in \mathbb{N}^n$  be an initial state and  $\sigma$  a permutation of the species indices. If  $\mathcal{R}$  and  $\boldsymbol{\nu}_0$  are  $\sigma$ -invariant, then  $\mu^{(\mathbf{m})} = \mu^{(\mathbf{m}_\sigma)}$ .*

*Proof.* At any time  $t$  we have

$$\begin{aligned} \mu^{(\mathbf{m})}(t) &= E[\mathbf{x}^{\mathbf{m}}(t)] \\ &= \sum_{\boldsymbol{\nu}} \boldsymbol{\nu}^{\mathbf{m}} \pi_{\boldsymbol{\nu}}(t) \\ &= \sum_{\boldsymbol{\nu}} \boldsymbol{\nu}^{\mathbf{m}_\sigma} \pi_{\boldsymbol{\nu}}(t) \text{ by commutativity of multiplication} \\ &= \sum_{\boldsymbol{\nu}} \boldsymbol{\nu}^{\mathbf{m}_\sigma} \pi_{\boldsymbol{\nu}_\sigma}(t) \text{ by Theorem 1} \\ &= E[\mathbf{x}^{\mathbf{m}_\sigma}(t)] = \mu^{(\mathbf{m}_\sigma)}(t) \end{aligned}$$

$\square$

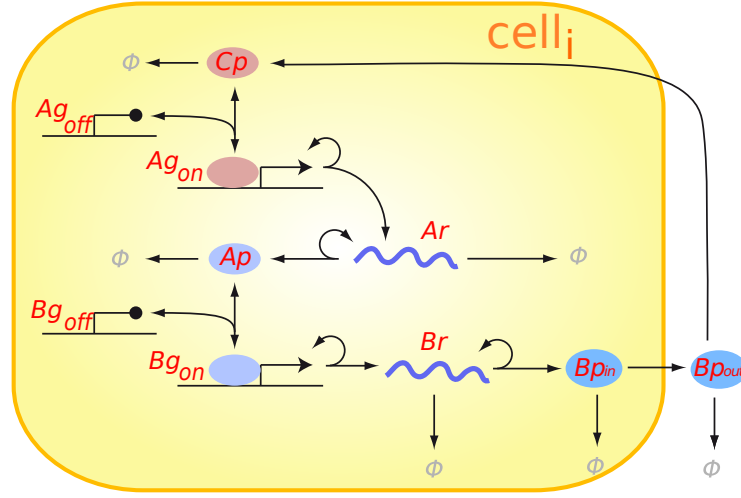
We denote by  $\Sigma(\mathcal{R}, \boldsymbol{\nu}_0)$  the set of permutations  $\sigma$  such that  $\mathcal{R}$  and  $\boldsymbol{\nu}_0$  are  $\sigma$ -invariant. By Corollary 1, this set defines equivalence classes  $[\mu^{(\mathbf{m})}]_\Sigma$  of moments, i.e. the set of moments  $\mu^{(\mathbf{m}')}$  such that  $\mu^{(\mathbf{m}')} = \mu^{(\mathbf{m}_\sigma)}$  for some  $\sigma \in \Sigma = \Sigma(\mathcal{R}, \boldsymbol{\nu}_0)$ . As usual, we also write  $[\mu^{(\mathbf{m})}]_\Sigma$  for the representative moment of this set that is, for instance, the smallest of those moments for the lexicographical order on  $\mathbb{N}^n$ . We denote  $\boldsymbol{\mu}_k$  a vector of all  $M$  moments up to order  $k$ . Let

$$\mathcal{M}(\mathcal{R}, k) = \{\dot{\mu}^{(\mathbf{m})} = L \cdot \boldsymbol{\mu}_h \mid \text{order}(\mathbf{m}) \leq k, L \in \mathbb{R}^M\}$$

be a set of moment equations obtained by some moment generation method, with moments up to order  $k$  (recall that moment equations are always linear).  $h$  is the maximum order of the moments in the equations, it can be greater than  $k$  for systems of moment equations with an unclosed cascade of dependencies.

The *reduced set of moment equations* is defined by

$$\mathcal{M}_{red}(\mathcal{R}, \boldsymbol{\nu}_0, k) = \{\rho(\dot{\mu}^{(\mathbf{m})}) = L \cdot \rho(\boldsymbol{\mu}_h) \mid (\dot{\mu}^{(\mathbf{m})} = L \cdot \boldsymbol{\mu}_h) \in \mathcal{M}(\mathcal{R}, k)\}$$



**Fig. 2.** Model of a community effect in Xenopus [28]. The extracellular molecule  $Bp_{out}$  mediates communication between  $n$  cells with identical intracellular reaction network.

where  $\rho$  is the substitution of moments for their representative

$$\rho = \{\mu^{(\mathbf{m})} \text{ is substituted by } [\mu^{(\mathbf{m})}]_{\Sigma} \mid \text{order}(\mathbf{m}) \leq k \text{ and } \Sigma = \Sigma(\mathcal{R}, \nu_0)\}$$

This transformation just excludes from  $\mathcal{M}(\mathcal{R}, k)$  repeated equations for the variables which are provably equal, and therefore is exact.

## 4 Application to a community effect

We applied the analytical tools described in this paper to a model of a community effect in Xenopus [28]. We derived a second order MC using normal approximation for simplicity. In order to do that efficiently, we reduced the moment equations as described in Section 3. Comparison of the solution of the truncated moment ODEs with stochastic simulations shows that the MC is significantly closer to the stochastic dynamics than the deterministic solution.

### 4.1 Community effect model

The model of a community effect in Xenopus [28] is summarized in Fig. 2. It features the species  $Bp_{out}$  for communication between  $n$  cells, each having the same intracellular network: within a cell,  $Bp_{out}$  triggers a cascade of two genes, and results in more  $Bp_{out}$  for cell-to-cell communication.

The intracellular details are: In the receiving cell, a signaling mechanism transforms  $Bp_{out}$  into the  $C_p$  protein, which binds to the first gene, and activates



its transcription into mRNA  $Ar$ , which in turn translates into the  $Ap$  protein.  $Ap$  activates the second gene, yielding mRNA  $Br$ , which translates into the  $Bp_{in}$  protein. The model distinguishes passive modes of genes ( $Ag_{off}$ ,  $Bg_{off}$ ) from active ( $Ag_{on}$ ,  $Bg_{on}$ ) – where genes are bound by their respective activator proteins, and constantly produce mRNA.

Leaving its original cell, the protein  $Bp_{in}$  becomes  $Bp_{out}$  when joining the common pool for cell-to-cell communication. From there, it can reach any cell in the system, and activate the gene cascade there. This closes the positive feedback loop of the community effect. Because diffusion is assumed infinitely fast,  $Bp_{out}$  equally likely reaches any of the  $n$  cells.

Finally, all species except the genes can degrade, yielding the pseudo species  $\phi$ . The complete model for  $n$  cells, with  $17n + 1$  reactions and the associated rates, is provided in a supplementary file.

Studies of the deterministic approximation of this model have shown that its behavior changes if the number of cells,  $n$ , exceeds a threshold  $n_c$ , that we call the *critical number*. If  $n > n_c$ , all cells continuously express their genes, otherwise all activity ceases after a short time.

Since gene and mRNA concentrations are always low in this system, stochastic effects may play a significant role in its dynamics. Indeed, stochastic simulations indicate that  $n_c = 97$  derived from the deterministic approximation in [28] may be imprecise: at  $n = 100$  we observed that in all 1000 simulations the gene expression stopped early. We studied the stochastic behavior of this system with the MC method.

## 4.2 Reduction of the community effect model

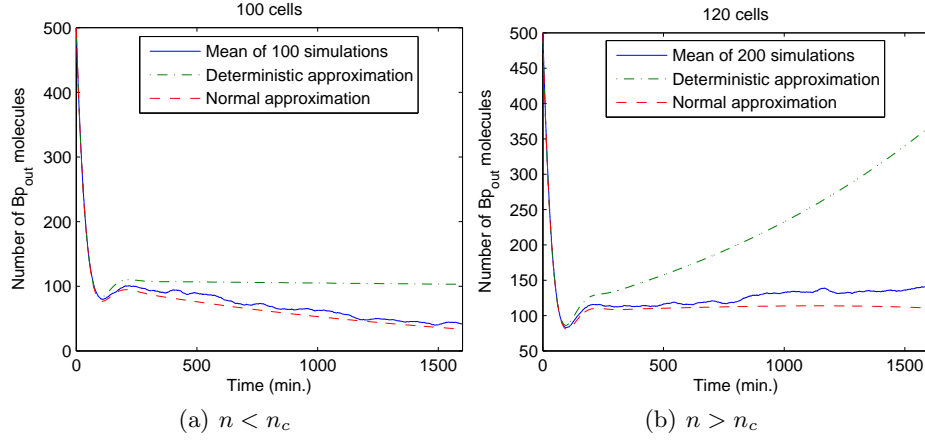
The community effect model's structure resembles a star, just as Figure 1 on page 2 does. It is easy to see that the community effect model exhibits the symmetries required by Theorem 1, which allows to reduce its moment equations for any number of cells to a system of constant size, similar to the example in Section 3.1.

The model contains 9 species per cell, and the procedure described in Section 2.3 generates  $40.5n^2 + 22.5n + 2$  moment equations<sup>3</sup> up to order two for  $n$  cells. For 120 cells, which is near this system's true  $n_c$ , it would generate 585902 equations. This by far exceeds the processing capabilities of the software we used. The reduced model contains 146 equations for any  $n$ . The deterministic approximation, which is equivalent to the first order MC, can also be reduced using the same method. This kind of reduction, among others, has been done in [28], where the deterministic approximation consisted of only 8 ODEs for any  $n$ . Our Maple<sup>4</sup> implementation of the MC method and the reduction for this model are available online<sup>5</sup>.

<sup>3</sup> This is always an integer.

<sup>4</sup> <http://www.maplesoft.com/>

<sup>5</sup> <http://www.lifl.fr/~batmanov/cmsb2012-files/>



**Fig. 3.** Traces of  $Bp_{out}$  over time for systems of sizes below and above  $n_c$ , computed using stochastic simulations, deterministic approximation and second order MC using normal approximation.

### 4.3 Comparison of the approximations

Figure 3 plots the  $Bp_{out}$  dynamics, computed by three different approximations. First is a *mean of many stochastic simulations*. As the number of simulations increases, this converges to the true mean - however it is noisy and the simulations take very long time. The simulations were done using COPASI software [?]. Second is a (usual) *deterministic approximation* of the system, with one ODE per species, which is the fastest in terms of computation. However, it tends to diverge from the stochastic estimates. This indicates the presence of significant stochastic effects in the model. Third is a *reduced MC of order two*, using the normal approximation for truncation. The normal approximation is not the best choice for chemical reaction systems generally, but it is simple to implement and it gives good results in this case.

Due to the complexity of the resulting system of ODEs in the MC, we couldn't derive an analytical solution for  $n_c$ . By examining the numerical solutions for different values of  $n$ , we found that the MC gives  $n_c = 117$ , the same as derived from statistical analysis of stochastic simulations.

The deterministic approximation, on the other hand, predicts  $n_c = 97$ , and therefore miscalculates the qualitative behavior of the system for a range of  $n$ . In addition, the deterministic estimate of  $Bp_{out}$  strongly diverges from the stochastic one, especially when the cell number  $n$  is close to  $n_c$ .

## 5 Conclusion

The moment closure method reviewed here is a flexible tool for approximate stochastic analysis. It allows manipulations of moment equations similar to those

that can be done with deterministic ODEs, but including, approximately, the stochastic effects.

Model reduction is one kind of such manipulations. It aims to eliminate redundant variables from the system of ODEs, making it easier to solve. For MC, which tend to generate a large number of ODEs, reductions are especially important. We have described a model reduction method based on symmetries, which in case of MC is more complicated than what is used with deterministic approximation.

Currently, the only way to exploit such symmetries while performing stochastic analysis of a system is through MC: reduced models are not amenable to Gillespie simulation. If the corresponding species in the cells are "lumped" together, in the same way as in the deterministic approximation, the results diverge quickly from the non-reduced system. Also note that, as the order of the closure grows, the symmetric reduction can eliminate a smaller fraction of moments, suggesting that for the limit case of the exact solution the gain from the reduction will be negligible. We believe that, for the community effect model, the approach presented here provides the only tractable analysis.

Symmetry-based reduction is potentially applicable to many highly regular systems. For example, in [3] a model reduction method for deterministic approximation is applied to a system that contains a protein with symmetric activation sites. That system is also symmetric in the sense described here, w.r.t. exchanging activation states for different sites. Thus, its MC could be reduced with our method as well. Another example is a discrete ring of identical cells, considered in [31], which is symmetric under rotation of all cells. Many higher order moments could be eliminated with our method using this property.

The approach presented here can be extended in a number of ways. Checking and finding the required symmetric properties of a reaction set can be automated rather easily. The symmetries considered here are just the automorphisms of the reaction graph with the additional constraints that the initial conditions of the corresponding species must be equal and the rates of the corresponding reactions must be equal. The problem of finding all automorphisms belongs to the NP class of complexity, however for real systems the requirement of having the same rates and initial conditions restricts the number of possible symmetries. Verification of a specified symmetry can be done in polynomial time.

One can also directly derive a reduced MC from a rule-based representation, without expanding it to the full system. This becomes interesting if the expanded system's size is huge, but the system is highly symmetric and can be described by a manageable set of moment equations. It resembles what is done in [3].

However, the current method is not applicable to spatial systems with borders. By borders, we mean the outermost cells in a one dimensional row of cells, or in a two-dimensional grid, those cells that frame the grid. In such system, a distinct distance from the border(s) uniquely identifies each cell. For example, one-dimensional spatial models of the community effect [1] can be reduced in half by central symmetry. But, for a second order MC, the quadratic dependency of number of equations on the system size remains. To deal with this,

[1] constructed systems of partial differential equations (PDEs), as limit cases when the number of cells tends to infinity and simultaneously their size goes to zero. While this allows efficient treatment in the deterministic regime, stochastic analysis is still required to run many long simulations.

Moment closures for spatial models have been previously derived in ecology [9] and statistical physics [21], and it may be possible to infer them automatically for chemical reaction systems as well.

**Acknowledgements.** We wish to thank Michel Petitot for discussions, and the Agence Nationale de Recherche for funding us (ANR BioSpace 2009-12, ANR Iceberg 2012-17).

## References

1. K Batmanov, C Kuttler, C Lhoussaine, and Y Saka. Self-organized patterning by diffusible factors: roles of a community effect. *Fundamenta Informaticae*, 2012.
2. H Bolouri and EH Davidson. The gene regulatory network basis of the "community effect," and analysis of a sea urchin embryo example. *Dev Biol*, 340(2):170–8, 2010.
3. F Camporesi and J Feret. Formal reduction of rule-based models. In *Math Foundations Programming Semantics*, volume 276C of *ENTCS*, pages 31–61, 2011.
4. Y. Cao, D.T. Gillespie, and L.R. Petzold. Efficient step size selection for the tau-leaping simulation method. *The Journal of chemical physics*, 124:044109, 2006.
5. Eric H. Davidson. *The Regulatory Genome: Gene Regulatory Networks In Development And Evolution*. Academic Press, 2006.
6. F Didier, TA Henzinger, M Mateescu, and V Wolf. Approximation of event probabilities in noisy cellular processes. *TCS*, 412(21):2128 – 2141, 2011.
7. Michael B. Elowitz, Arnold J. Levine, Eric D. Siggia, and Peter S. Swain. Stochastic gene expression in a single cell. *Science*, 297:1183–1186, 2002.
8. P. Erdi and J. Toth. *Mathematical Models of Chemical Reactions - Theory & Applications of Deterministic & Stochastic Models*. Nonlinear Science: Theory and Applications. John Wiley & Sons, 1992.
9. A Gandhi, S Levin, and S Orszag. Moment expansions in spatial ecological models and moment closure through gaussian approximation. *Bull Math Biol*, 62:595–632, 2000. 10.1006/bulm.1999.0119.
10. CS Gillespie. Moment-closure approximations for mass-action models. *IET Sys Biol*, 3(1):52, 2009.
11. DT Gillespie. A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. *J Comp Physics*, 22:403–434, 1976.
12. DT Gillespie. Chemical Langevin equation. *J Chem Physics*, 113:297–306, 2000.
13. M. Golubitsky, M. Pivato, and I. Stewart. Interior symmetry and local bifurcation in coupled cell networks. *Dynamical Systems*, 19(4):389–407, December 2004.
14. Leo A. Goodman. Population growth of the sexes. *Biometrics*, 9(2):212–225, 1953.
15. B T Grenfell, K Wilson, V S Isham, H E Boyd, and K Dietz. Modelling patterns of parasite aggregation in natural populations. *Parasitology*, 111, January 1995.
16. J. B. Gurdon. A community effect in animal development. *Nature*, 336(6201):772–4, 1988.
17. M. Hegland. Approximating the solution of the chemical master equation by aggregation. In *14th Computational Techniques and Applications Conference*, volume 50 of *ANZIAM J.*, pages C371–C384, 2008.

18. JP Hespanha and A Singh. Stochastic models for chemically reacting systems using polynomial stochastic hybrid systems. *J. Robust Control*, 15:669–689, 2005.
19. Stefan Hoops, Sven Sahle, Ralph Gauges, Christine Lee, Jurgen Pahle, Natalia Simus, Mudita Singhal, Liang Xu, Pedro Mendes, and Ursula Kummer. Copasi – a complex pathway simulator. *Bioinformatics*, 22(24):3067–3074, 2006.
20. Thomas R. Kiehl, Robert M. Mattheyses, and Melvin K. Simmons. Hybrid simulation of cellular behavior. *Bioinformatics*, 20(3):316–322, 2004.
21. C.H. Lee, K.H. Kim, and P. Kim. A moment closure method for stochastic reaction networks. *The Journal of Chemical Physics*, 130:134107, 2009.
22. C. Levermore. Moment closure hierarchies for kinetic theories. *Journal of Statistical Physics*, 83:1021–1065, 1996. 10.1007/BF02179552.
23. P Milner, C Gillespie, and D Wilkinson. Moment closure based parameter inference of stochastic kinetic models. *Statistics and Computing*, pages 1–9, 2012.
24. P Milner, CS Gillespie, and DJ Wilkinson. Moment closure approximations for stochastic kinetic models with rational rate laws. *Mathematical biosciences*, 231(2):99–104, June 2011.
25. B Munsky and M Khammash. The finite state projection algorithm for the solution of the chemical master equation. *Journal chemical physics*, 124(4):044104, 2006.
26. A Ramos, G Innocentini, F Forger, and J Hornos. Symmetry in biology: from genetic code to stochastic gene regulation. *IET systems biology*, 4(5):311–29, September 2010.
27. Y Saka, C Lhoussaine, C Kuttler, E Ullner, and M Thiel. Theoretical basis of the community effect in development. *BMC Systems Biology*, 5:54, March 2011.
28. Abhyudai Singh and João Pedro Hespanha. Lognormal moment closures for biochemical reactions. In *Proc. of the 45th Conf. on Decision and Contr.*, Dec. 2006.
29. Zarath M Summers, Heather E Fogarty, Ching Leang, Ashley E Franks, Nikhil S Malvankar, and Derek R Lovley. Direct exchange of electrons within aggregates of an evolved syntrophic coculture of anaerobic bacteria. *Science (New York, N.Y.)*, 330(6009):1413–5, December 2010.
30. A. M. Turing. The chemical basis of morphogenesis. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 237(641):37–72, 1952.
31. J Uhlendorf, P Hersen, and G Batt. Towards real-time control of gene expression : in silico analysis. *IFAC*, 18:14844–14850, 2011.
32. S Vidal, Petitot, F Boulier, F Lemaire, and C Kuttler. Models of stochastic gene expression and Weyl algebra. *Algebraic Biology, LNCS*, June 2010.
33. P. Whittle. On the use of the normal approximation in the treatment of stochastic processes. *Journal Royal Statistical Society. Series B*, 19(2):268–281, 1957.
34. Karsten Zengler and Bernhard O. Palsson. A road map for the development of community systems (CoSy) biology. *Nature Reviews Microbiology*, 10(5):366–372, March 2012.