Approximation and inference methods for stochastic biochemical kinetics—a tutorial review

View the article online for updates and enhancements.

You may also like

Corrigendum

Corrigendum: Approximation and inference methods for stochastic biochemical kinetics—a tutorial review (2017 *J. Phys. A: Math. Theor.* 50 093001)

David Schnoerr[1,2,3]**, Guido Sanguinetti**[2,3] **and Ramon Grima**[1,3]

[1] School of Biological Sciences, University of Edinburgh, Edinburgh EH9 3JH, United Kingdom
[2] School of Informatics, University of Edinburgh, Edinburgh EH8 9AB, United Kingdom
[3] SynthSys, University of Edinburgh, Edinburgh EH9 3JD, United Kingdom

E-mail: ramon.grima@ed.ac.uk

CrossMark

We found mistakes in equations: (57), (58) and (62). Equation (57) should read

$$\mathcal{L}^{(1)} = \Omega^{-1/2}\left(\frac{1}{2}J^{(0)q}_{i,j}\partial_i\partial_j\epsilon_q - \frac{1}{2}J^{(0)q,r}_i\partial_i\epsilon_q\epsilon_r - \frac{1}{6}D^{(0)}_{i,j,k}\partial_i\partial_j\partial_k - D^{(1)}_i\partial_i\right),$$

while (58) should read

$$\mathcal{L}^{(2)} = \Omega^{-1}\left(-\frac{1}{6}J^{(0)q}_{i,j,k}\partial_i\partial_j\partial_k\epsilon_q + \frac{1}{4}J^{(0)q,r}_{i,j}\partial_i\partial_j\epsilon_q\epsilon_r - \frac{1}{6}J^{(0)q,r,s}_i\partial_i\epsilon_q\epsilon_r\epsilon_s\right.$$
$$\left. - J^{(1)q}_i\partial_i\epsilon_q + \frac{1}{2}D^{(1)}_{i,j}\partial_i\partial_j + \frac{1}{24}D^{(0)}_{i,j,k,l}\partial_i\partial_j\partial_k\partial_l\right).$$

Finally, equation (62) should read

$$\partial_t\boldsymbol{\Sigma} = \boldsymbol{J}^{(0)}\boldsymbol{\Sigma} + \boldsymbol{\Sigma}\boldsymbol{J}^{(0)^T} + \Omega\boldsymbol{D}^{(0)}.$$

Topical Review

Approximation and inference methods for stochastic biochemical kinetics—a tutorial review

David Schnoerr[1,2,3], **Guido Sanguinetti**[2,3]
and Ramon Grima[1,3]

[1] School of Biological Sciences, University of Edinburgh, Edinburgh EH9 3JH, UK
[2] School of Informatics, University of Edinburgh, Edinburgh EH8 9AB, UK
[3] SynthSys, University of Edinburgh, Edinburgh EH9 3JD, UK

E-mail: ramon.grima@ed.ac.uk

Abstract

Stochastic fluctuations of molecule numbers are ubiquitous in biological systems. Important examples include gene expression and enzymatic processes in living cells. Such systems are typically modelled as chemical reaction networks whose dynamics are governed by the chemical master equation. Despite its simple structure, no analytic solutions to the chemical master equation are known for most systems. Moreover, stochastic simulations are computationally expensive, making systematic analysis and statistical inference a challenging task. Consequently, significant effort has been spent in recent decades on the development of efficient approximation and inference methods. This article gives an introduction to basic modelling concepts as well as an overview of state of the art methods. First, we motivate and introduce deterministic and stochastic methods for modelling chemical networks, and give an overview of simulation and exact solution methods. Next, we discuss several approximation methods, including the chemical Langevin equation, the system size expansion, moment closure approximations, time-scale separation approximations and hybrid methods. We discuss their various properties and review recent advances and remaining challenges for these methods. We present a comparison of several of these methods by means of a numerical case study and highlight some of their respective advantages and disadvantages. Finally, we discuss the problem of inference from experimental data in the Bayesian framework and review recent methods developed the literature. In summary, this review gives a

self-contained introduction to modelling, approximations and inference methods for stochastic chemical kinetics.

(Some figures may appear in colour only in the online journal)

1. Introduction

Understanding the functioning of living cells and biological organisms at the system level has gained increasing attention in recent years and defines a key research programme for the next decades. Experimental techniques are developing at breathtaking speed producing a wealth of data at finer and finer resolutions. However, such experimental data does not by itself reveal the function of such biological systems. The underlying processes typically involve large numbers of interacting components giving rise to highly complex behaviour [1]. Moreover, experimental data are generally corrupted by measurement noise and incomplete, thus posing the mathematical and statistical challenge to infer the relevant biological information from such measurements.

We focus here on mathematical and statistical modelling of chemical reaction networks in biological systems in which *random fluctuations* of molecule numbers play an important role. Recent experiments have shown this to be the case in many biological processes, gene expression being a prominent example [2]. Such random fluctuations or *stochastic effects* in gene expression have been found to lead to dramatically differing behaviours of genetically identical cells. From a modelling point of view, stochastic systems are considerably harder to analyse than their deterministic counterparts.

The *rate equations* give a macroscopic, deterministic description of the dynamics of chemical reaction networks. They consist of a set of ordinary differential equations governing the dynamics of the mean concentrations of the species in a system, thereby ignoring stochastic fluctuations. The rate equations have been successfully applied to various problems and have the advantage of being relatively straightforward to analyse. They typically give an accurate description of systems with large numbers of molecules, which is generally the case in *in vitro* experiments. However, the rate equations do no longer provide a valid description in cases where the effects of stochastic fluctuations become significant. This is typically the case when some species in a system occur at low molecule numbers, a common feature of chemical networks in cells. In this case, the *chemical master equation* (CME) constitutes the accepted probabilistic description of the resulting *stochastic process* [3]. In this framework, the state of a system is given by the molecule numbers of different species, and the CME governs the single-time probability of being in a certain state at a given time.

Despite its simple structure, no analytic solutions to the CME are known for all but the simplest systems. It is however possible to simulate exact sample paths of the underlying stochastic process, by means of the *stochastic simulation algorithm* [4, 5]. The latter allows us to draw exact samples from the process. However, since the stochastic simulation algorithm simulates each and every chemical reaction event in a system, it is computationally expensive and quickly becomes infeasible for larger systems. Accordingly, significant effort has been spent in recent decades on the development of approximation methods of the CME, and a large variety of different methods has emerged. Similarly, due to the growing amount of

experimental data becoming available, a considerable amount of work has been devoted to the development of statistical inference methods for such data, i.e. methods that allow to calibrate a model to observational data.

There exists a significant amount of literature on modelling of stochastic chemical kinetics. However, most reviews are mainly concerned with simulation based methods [6–12]. Others are rather technical and require a high level of pre-existing and/or mathematical knowledge [13–17]. None of these references gives a thorough introduction into modelling of stochastic chemical kinetics in the biological context that is accessible for non-experts, neither do they give an overview or comparison of state of the art approximation and inference methods.

With this article, we aim at filling these gaps in several ways:

1. We give a self-contained introduction to deterministic and stochastic modelling techniques for chemical reaction networks. First, we introduce and give a historic motivation to the deterministic rate equations. Next, we derive and give a detailed discussion of the CME. We review exact solution methods for the CME as well as simulation methods.
2. We give a detailed derivation and discussion of the following approximation methods of the CME: the chemical Langevin equation and its associated chemical Fokker–Planck equation; the system size expansion; moment closure approximations. Moreover, we give an introduction and overview of other types of approximations, including time-scale separation based methods and hybrid approximations.
3. We perform a numerical case study comparing the various approximation methods mentioned before.
4. We give an introduction to inference methods in the Bayesian framework and review existing methods from the literature.

The presentation is written to be accessible for non-experts that are new to the field of stochastic modelling.

Even though this review is motivated by stochastic effects in systems biology, it is important to stress that many systems in other scientific fields are frequently modelled by means of Master Equations. The methods discussed in this article can therefore readily be applied to such systems. Examples include ecology [18–20], epidemiology [21–26], social sciences [27–29] and neuroscience [30–36].

This article is structured as follows. We start by discussing the importance of stochasticity in biological systems in section 2. We describe the underlying mechanisms giving rise to stochastic fluctuations in cells (section 2.1) and discuss key experimental studies that measured such fluctuations (section 2.2). Next in section 3 we discuss deterministic and stochastic modelling methods for chemical kinetics. We start by introducing the concept of chemical reaction networks and deterministic descriptions in terms of macroscopic rate equations in section 3.1. Next, we introduce stochastic modelling techniques in terms of the CME and stochastic simulation algorithms in sections 3.2 and 3.3, respectively. We discuss analytic solution methods for certain classes of reaction systems in section 3.4. Section 4 is devoted to approximation methods of the CME. We give detailed introductions and discussions of the chemical Langevin equation (section 4.1), the system size expansion (section 4.2) and moment closure approximations (section 4.3). Next, we discuss how approximate moment values obtained from these methods can be used to construct probability distributions by means of the maximum entropy principle in section 4.4. In section 4.5 we review existing software packages implementing the approximation methods discussed in sections 4.1–4.3. We give an introduction to other approximation methods in section 4.6, including the finite state projection algorithm, time-scale separation based approximations and hybrid methods.

In section 5 we perform a numerical case study and compare the chemical Langevin equation, the system size expansion and moment closure approximations. Section 6 deals with the problem of inference for CME type systems from observational data. We introduce the Bayesian approach to this problem and review existing methods from the literature. Finally, we summarise and conclude in section 7.

2. Stochasticity in biological systems

Stochastic effects play an important role in many chemical reaction networks in living cells. Examples are enzymatic/catalytic processes, transduction of external signals to the interior of cells or the process of gene expression, to name just a few. Here we discuss different sources of stochasticity and illustrate their emergence in biochemical networks using gene expression as an example in section 2.1. Subsequently in section 2.2 we explain how the different sources of stochasticity can be measured and distinguished experimentally and highlight the importance of stochasticity for living cells.

2.1. Emergence of stochasticity

The term 'gene expression' denotes the process of synthesis of functional gene products such as proteins. The mechanism is illustrated in figure 1 for prokaryotic cells, namely cells lacking in-membrane-bound organelles (e.g. nucleus, mitochondria), such as bacteria. The process includes two main steps: *transcription* during which mRNA molecules are produced, and *translation* during which protein molecules are synthesised [37]. The transcription process involves the enzyme RNA polymerase. For the mechanism to initiate, an RNA polymerase enzyme must bind to the initial sequence of a gene. It then slides along the gene and produces an mRNA molecule that reproduces the DNA code of the gene. The RNA polymerase molecules move around randomly in the cell, a process which can be approximately described as Brownian motion [15]. This means that the RNA polymerase binding to the gene is a stochastic event that happens randomly in time. As it turns out, not only the binding of the RNA polymerase to the gene, but also the sliding along the gene happens stochastically. Therefore, the production of mRNA molecules is a stochastic process.

The production of protein molecules from mRNA during translation is conducted by ribosomes, which are RNA and protein complexes. The ribosomes and mRNA diffuse in the cell and hence meet randomly before translation can occur. Translation is thus also a stochastic process. Similarly, the degradation of mRNA molecules and proteins is conducted by certain enzymes and hence happens stochastically.

The transcriptional regulation of gene expression is frequently modulated by regulatory proteins named *transcription factors*. Transcription factors are gene specific and bind to the *promoter region*, which is located upstream of the gene's encoding region. Upon binding, transcription factors tune the affinity of the RNA polymerase molecules for the promoter, thereby modulating the rate at which transcription initiation events occur and hence the overall transcription rate. A transcription factor can either increase or decrease the binding rate of RNA polymerase and thus either enhance or suppress gene expression.

In eukaryotes, i.e. cells possessing in-membrane-bound organelles such as a nucleus, gene expression happens in a similar but somewhat more complicated way involving more steps. For instance, genes are generally located in the nucleus and transcription hence happens in the nucleus. Processes such as mRNA maturation and diffusion of mRNA to the cytoplasm need to happen before the mature mRNA can be translated into protein.
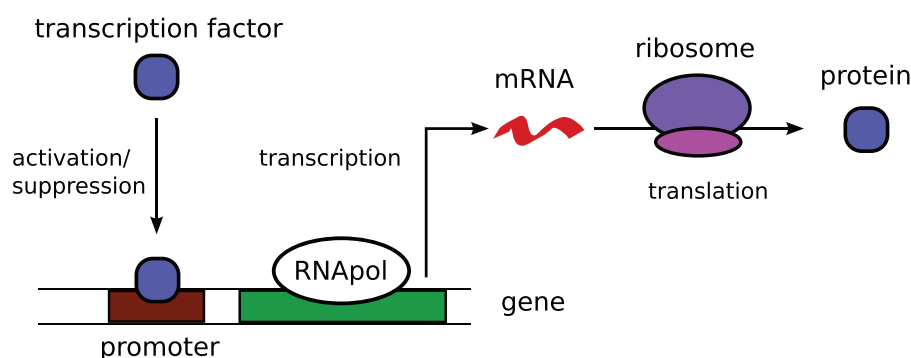
Figure 1. Illustration of gene expression in prokaryotic cells. Transcription is conducted by a RNA polymerase enzyme (RNApol) that binds to the gene and produces an mRNA molecule from the gene's DNA. The mRNA is then translated into proteins by ribosomes. Transcription factors may bind to the promoter of the gene and thereby influence the recruitment of RNA polymerase and hence the transcriptional rate of the gene. Gene expression in eukaryotes happens similarly but is compartmentalised in the nucleus (transcription) and the cytosol (translation).
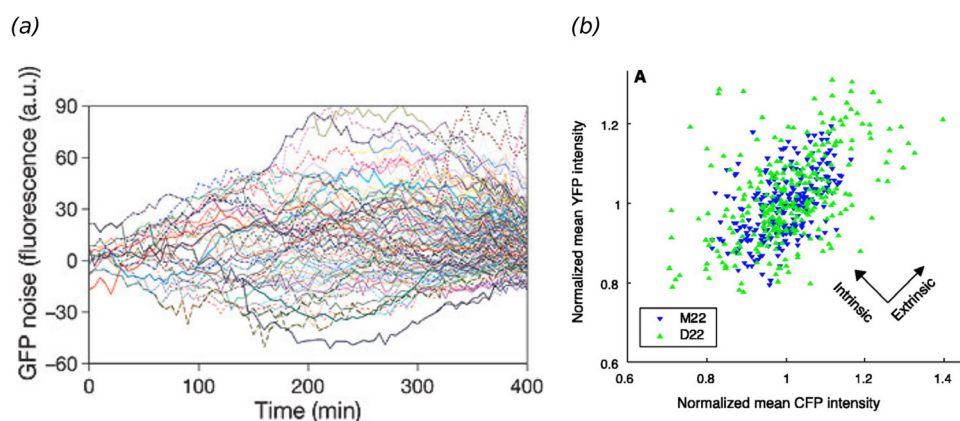
(a)   *(b)*

Figure 2. Measurements of intrinsic and extrinsic noise. (a) Fluorescent time series acquired on *E. coli* cells expressing a green fluorescent protein (GFP) encoding gene. The time trajectories correspond to the difference of the measured intensities in individual cells and the population average. We observe that the protein numbers fluctuate strongly over time [39]. From [2]. Reprinted with permission from AAAS. (b) Fluorescence intensities for the dual reporter technique measured in [2] for the two *E. coli* strains M22 and D22. Permission kindly granted by the publisher. Each triangle corresponds to the measured fluorescence intensity in one cell. Variations along and perpendicular to the diagonal correspond to extrinsic and intrinsic noise, respectively. We observe that in this system both intrinsic and extrinsic noise contribute significantly to the overall fluctuations. Reprinted by permission from Macmillan Publishers Ltd: *Nature* [39], copyright 2006.

The inherent stochasticity of chemical processes leads to fluctuations of molecule numbers in time. As an example figure 2(a) shows fluorescence intensity measurements of a fluorescent protein in individual *Escherichia coli* cells, indicating strong fluctuations of protein numbers in time. Temporal fluctuations that are due to the stochasticity of chemical processes is what we call *intrinsic noise* [2]. Differences in molecule numbers of a certain species, say proteins, between different cells can originate from this type of fluctuations. However, such differences

can also stem from other effects, such as physiological differences between cells or differing environmental conditions. For example, the numbers of RNA polymerase or ribosomes may differ between different cells, or different cells may be exposed to varying nutrient concentrations due to environmental fluctuations. Such differences that are not due to the stochasticity of chemical reactions are referred to as *extrinsic noise* [38].

2.2. Experimental evidence

As explained above, stochastic fluctuations are inherent to biochemical processes such as gene expression. The question hence arises: what effects do these fluctuations have on the functioning of cells and which processes dominate the emergence of these fluctuations? To answer such questions it is crucial to be able to empirically distinguish and quantify intrinsic and extrinsic noise.

One of the first experiments that aimed at explicitly separating the effects of intrinsic and extrinsic noise on gene expression was conducted by Elowitz *et al* in 2002 [2] by means of the *dual reporter technique*. In this study, the authors integrated two distinguishable but identically regulated fluorescent reporter genes into the chromosome of *E. coli* cells, one expressing cyan fluorescent protein (CFP) and one yellow fluorescent protein (YFP). The two genes were controlled by identical promoters, which means that they experienced the same external effects, i.e. the same extrinsic noise. Assuming that the two promoters are independent, fluctuations of the two proteins originating from the stochasticity of chemical reactions, i.e. intrinsic noise, should be completely uncorrelated. On the other hand, extrinsic effects should influence both expression systems equally, and the corresponding fluctuations should therefore be strongly correlated. This can be visualised as in figure 2(b) which is taken from [2]. The figure shows the measured intensities of individual cells for two different strains, with the YFP intensity on the *y*-axis and the CFP on the *x*-axis. The width of the cloud along the diagonal corresponds to correlated fluctuations and hence extrinsic noise, while the width orthogonal to the diagonal corresponds to uncorrelated fluctuations, i.e. intrinsic noise. The figure indicates that for this particular system both intrinsic and extrinsic noise significantly contribute to the overall noise.

Many other experimental studies have been conducted investigating the origins and role of intrinsic and extrinsic noise in cells (for example [39–44]). For an overview of experimental techniques see [45, 46]. In [43], for example, the authors measured the expression levels of 43 genes in *E. coli* under various different environmental conditions. They observed that the variance of protein numbers scales roughly like the mean protein numbers and that for intermediate abundance proteins the intrinsic fluctuations are comparable or larger than extrinsic fluctuations. In another large scale study on *E. coli* in [44], the authors found that fluctuations are dominated by extrinsic noise at high expression levels.

The reported fluctuations can have significant influence on the functional behaviour of cells [1]. A particularly important example are *stochastic cell fate decisions* [47], where genetically identical cells under the same environmental conditions stochastically differentiate into functionally different states. Such differing cell fate decisions of genetically identical cells are believed to be beneficial for a population of cells experiencing fluctuating environments [47].

Most of the mentioned experimental studies measure certain statistics of gene expression systems, such as the equilibrium mean and variance of protein numbers. While experimental methods such as the *dual reporter technique* allow to estimate the magnitude of intrinsic and extrinsic noise, they do not explain where the stochasticity originates from and what

the underlying processes are. In order to answer such questions, mathematical methods are needed that allow researchers to model the dynamics and stochasticity of such biochemical reaction systems.

3. Stochastic chemical kinetics

After having given a biological motivation for the importance of stochasticity in chemical reaction networks in the previous section, we consider next the question of how such systems can be described mathematically. First, we introduce the concept of chemical reaction networks in section 3.1 and describe the classical mathematical description of such systems in terms of deterministic rate equations. Next, we discuss the stochastic description of such systems in section 3.2. We derive the chemical master equation and discuss its validity and solutions. Next, we discuss the stochastic simulation algorithm in section 3.3. Finally, we review methods for exact analytic solutions of the CME for certain classes of systems in section 3.4 and give an overview of available analytic solutions of specific systems that do not fall under any of these categories.

3.1. Chemical reaction networks and deterministic rate equations

Biological processes such as gene expression generally consist of complicated mechanisms involving several different types of molecules and physical operations. For a mathematical description of certain processes, one typically does not model all these mechanisms explicitly, but rather replaces them by an effective single chemical reaction event. In the context of gene expression, for example, transcription or translation may be modelled as single chemical reactions. A finite set of chemical species that interact via a set of such chemical reactions constitutes what we call a *chemical reaction network*. Given a set of chemical species X_i, $i = 1, ..., N$, we define R chemical reactions by the notation

$$\sum_{i=1}^{N} s_{ir} X_i \xrightarrow{k_r} \sum_{i=1}^{N} s'_{ir} X_i, \quad r = 1, ..., R, \tag{1}$$

where the *stoichiometric coefficients* s_{ir} and s'_{ir} are non-negative integer numbers denoting numbers of reactant and product molecules, respectively. We say that the rth reaction is 'of order m' if $\sum_{i=1}^{N} s_{ir} = m$, i.e. if it involves m reactant molecules. We further call a reaction 'unimolecular' if $m = 1$, 'bimolecular' if $m = 2$ and a system 'linear' if $m \leqslant 1$ for all reactions occurring in the system. We further call a system 'open' if it contains a chemical process that generates molecules, and 'closed' otherwise. The quantity k_r in equation (1) is called the *reaction rate constant* of the rth reaction.

Classically, the dynamics of a chemical reaction system as in equation (1) has been modelled by *the law of mass action* which was developed by Guldberg and Waage in its first version in the 1860s in the context of macroscopic *in vitro* experiments, i.e. macroscopic amounts of chemical molecules in solutions [48–51]. The law of mass action states that the rate of a reaction is proportional to the product of the concentrations of reactant molecules. Specifically, if we define $\phi = (\phi_1, ..., \phi_N)$, where ϕ_i denotes the concentration of species X_i, the rate g_r of a reaction as in equation (1) is given by

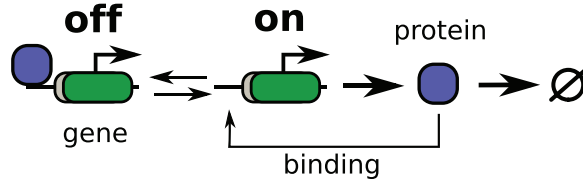$$g_r(\phi) = k_r \prod_{i=1}^{N} \phi_i^{s_{ir}}. \tag{2}$$

Figure 3. Illustration of a gene expression system with negative feedback loop. When the gene is in the 'on' state it produces proteins of a certain type. The protein can decay or bind to the gene's promoter. In the bound state the gene is in the 'off' state, i.e. no protein is produced. The corresponding reactions are given in equation (5). The protein hence suppresses its own production.

We call g_r the *macroscopic rate function* of the rth reaction. Let us further define the *stoichiometric matrix* **S** as

$$S_{ir} = s'_{ir} - s_{ir}, \quad i = 1, \dots, N, \quad r = 1, \dots, R. \tag{3}$$

The entry S_{ir} corresponds to the net change of X_i molecules when the rth reaction occurs. Consequently, $S_{ir}g_r(\phi)$ is the rate at which the concentration ϕ_i of species X_i changes due to the rth reaction. Summing up the contributions from all reactions in a given system one obtains [52]

$$\frac{\mathrm{d}}{\mathrm{d}t}\phi_i = \sum_{r=1}^{R} S_{ir}g_r(\phi), \quad i = 1, \dots, N. \tag{4}$$

This set of ordinary differential equations is called the *rate equations*. The law of mass action and the associated rate equations assume continuous concentrations, which means that they ignore the discreteness of molecule numbers. Moreover, they constitute a deterministic method, i.e. for a fixed initial condition the state of the system is exactly determined for all times. However, in the previous section we have seen that chemical reactions occur stochastically leading to fluctuations of molecule numbers in time. The reason why the rate equations give an accurate description in experiments such as the ones of Guldberg and Waage is that they studied chemical systems in solutions which typically contain a large number of substrate molecules. For large molecule numbers, it has been found experimentally that the relative fluctuations, i.e. the standard deviation divided by the mean value of molecule concentrations, scales like the inverse square root of the mean concentration and hence becomes small for systems with large molecule numbers [43]. One can therefore expect the rate equations to be an accurate description whenever a system has large molecule numbers for all species.

Example. As an example, consider the gene system in figure 3. This system can be viewed as a simplified version of the gene expression process described in figure 1: we replace the process of transcription (gene produces an mRNA molecule) and translation (mRNA molecule produces a protein) by a single effective reaction in which the gene directly produces the protein.

The example in figure 3 depicts negative autoregulation. The latter is one of the most common regulatory mechanisms: the gene's product downregulates its own expression. In the bound state G^{off}, the gene does not produce any protein. The protein thus suppresses its own production, which means the system is an example of a *negative feedback loop*. The corresponding reactions in the notation of equation (1) read

$$G^{\mathrm{on}} \xrightarrow{k_1} G^{\mathrm{on}} + P, \quad G^{\mathrm{on}} + P \underset{k_3}{\overset{k_2}{\rightleftharpoons}} G^{\mathrm{off}}, \quad P \xrightarrow{k_4} \varnothing, \tag{5}$$

where we call the gene in the bound and unbound state G^{on} and G^{off}, respectively and the protein P. In our nomenclature, $G^{\mathrm{on}} + P \to G^{\mathrm{off}}$ is a second-order or bimolecular reaction, while the other three reactions are first order or linear. By '$P \to \varnothing$' we indicate that P leaves the system under consideration. This could for example mean that P becomes converted into different types of chemical species or degraded in its constitutive elements that are not included in our model.

Let us order reactions according to the rate constants in equation (5), and species as G^{on}, P and G^{off}, respectively. Consider the stoichiometric coefficients s_{ir} and s'_{ir} defined in equation (1), which correspond to the number of reactant and product molecules, respectively, of species X_i in the rth reaction. For the reaction $G^{\mathrm{on}} \to G^{\mathrm{on}} + P$, for example, we have $s_{11} = 1, s_{21} = 0$ and $s_{31} = 0$, since there is one G^{on} on the lhs of the reaction but no P or G^{off}, and $s'_{11} = 1, s'_{21} = 1$ and $s'_{31} = 0$ since there is one G^{on}, one P and no G^{off} molecules on the rhs. Proceeding similarly for the other reactions in equation (5), we find

$$\mathbf{s} = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix}, \qquad \mathbf{s'} = \begin{pmatrix} 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}. \tag{6}$$

Accordingly, the stoichiometric matrix $\mathbf{S} = \mathbf{s'} - \mathbf{s}$ reads

$$\mathbf{S} = \begin{pmatrix} 0 & -1 & 1 & 0 \\ 1 & -1 & 1 & -1 \\ 0 & 1 & -1 & 0 \end{pmatrix}. \tag{7}$$

Let $\boldsymbol{\phi} = (\phi_1, \phi_2, \phi_3)$, where ϕ_1, ϕ_2 and ϕ_3 denote the concentrations of G^{on}, P and G^{off}, respectively. The macroscopic rate vector $\mathbf{g}^{(0)}(\boldsymbol{\phi}) = (g_1^{(0)}(\boldsymbol{\phi}), ..., g_r(\boldsymbol{\phi}))^T$, with the g_r defined in equation (2), is obtained by using $\mathbf{s}$ in equation (6) and reads

$$\mathbf{g}(\boldsymbol{\phi}) = (k_1\phi_1, k_2\phi_1\phi_2, k_3\phi_3, k_4\phi_2)^T. \tag{8}$$

Using equations (4), (7) and (8) it is easy to write down the corresponding rate equations. However, note that the system has a conservation law in particle numbers which we can use to find a simplified description by reducing the number of variables: the total number of genes in the 'on' state and genes in the 'off' state is constant, i.e. $\phi_1 + \phi_3 = \mathrm{const.} \equiv g_0$. We can thus reduce the system to a two species system by using $\phi_3 = g_0 - \phi_1$. The matrices $\mathbf{s}$ and $\mathbf{s'}$ for the reduced system are obtained from equation (6) by dropping the last row,

$$\mathbf{s} = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 \end{pmatrix}, \qquad \mathbf{s'} = \begin{pmatrix} 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \end{pmatrix}, \tag{9}$$

and the stoichiometric matrix and propensity vector of the reduced system read accordingly

$$\mathbf{S} = \begin{pmatrix} 0 & -1 & 1 & 0 \\ 1 & -1 & 1 & -1 \end{pmatrix}, \quad \mathbf{g}(\boldsymbol{\phi}) = (k_1\phi_1, k_2\phi_1\phi_2, k_3(g_0 - \phi_1), k_4\phi_2)^T. \tag{10}$$

Using equation (4) we hence obtain the rate equations

$$\begin{aligned} \frac{\partial}{\partial t}\phi_1 &= -k_2\phi_1\phi_2 + k_3(g_0 - \phi_1), \\ \frac{\partial}{\partial t}\phi_2 &= k_1\phi_1 - k_2\phi_1\phi_2 + k_3(g_0 - \phi_1) - k_4\phi_2. \end{aligned} \tag{11}$$

The gene system in figure 3 will be used to showcase several different methods in this article, and we will throughout use the reduced system.

3.2. Stochastic methods

The rate equations discussed in the previous section constitute a deterministic description of chemical reaction systems in terms of continuous molecule concentrations. As we have seen in section 2, however, fluctuations become important in many biological processes such as gene expression. The reason for this is that often some species occur at low molecule counts. A more general, stochastic description would keep the discrete nature of the X_i molecules in equation (1) intact. One general approach is to explicitly model the spatial positions of molecules and to model their movement as Brownian motion, with chemical reactions happening stochastically under certain rules. Exact analytic results for such systems are generally not known. Simulations, on the other hand, have to keep track of every single particle and quickly become computationally unfeasible. However, under certain conditions, which we discuss in the following section, simplified descriptions can be employed making spatial models and the simulation of single particles unnecessary.

3.2.1. The chemical master equation. We now consider a chemical reaction system as in equation (1) in a closed compartment of volume Ω. We seek a stochastic description of the system under well-mixed and dilute conditions. By 'well-mixed' we mean that the diffusion of particles in the compartment constitutes the fastest time scale of the system, in the sense that the expected distance travelled by each particle between successive reactive collisions is much larger then the length scale of the compartment. This implies that the spatial positions of molecules can be ignored and the dynamics of the system only depends on the total molecule numbers. By 'dilute' we mean that the combined volume of all the considered molecules is much smaller than the total volume, which means that the molecules can be considered as point particles.

If these two conditions are met, it can be shown [3] that the state of the system at any time is fully determined by the state vector $\mathbf{n} = (n_1, \ldots, n_N)$, where n_i is the molecule number of species X_i in the compartment. In particular, the spatial locations and diffusion of molecules does not have to be modelled, and the system corresponds to a *continuous-time Markov jump process*. It can further be shown that the probability for the rth reaction to happen in an infinitesimal time step $\mathrm{d}t$ is given by $f_r(\mathbf{n})\mathrm{d}t$ where $f_r(\mathbf{n})$ is the *propensity function* of the rth reaction and proportional to the number of combinations of reactant molecules in $\mathbf{n} = (n_1, \ldots, n_N)$. Consider for example a bimolecular reaction of the form $A + B \rightarrow \varnothing$. The number of pairs with one A and one B molecule is $n_A n_B$, where n_A and n_B are the molecule numbers of A and B, respectively. The corresponding propensity function is hence given by $k_r n_A n_B / \Omega$. The scaling with the volume Ω stems from the fact that the probability for two molecules to collide is proportional to $1/\Omega$. Generalising these arguments to reactions as in equation (1) leads to

$$f_r(\mathbf{n}) = k_r \Omega \prod_{i=1}^{N} \frac{n_i!}{(n_i - s_{ir})!\, \Omega^{s_{ir}}}. \tag{12}$$

Propensity functions of this form are called *mass-action kinetics type* [14]. Although equation (2) is often, due to historical reasons, stated to be the law of mass action, from a microscopic perspective it is more accurate to state equation (12) as the law of mass action. Equation (2) can be viewed as the macroscopic version of (12) obtained in the limit of large molecule numbers and small fluctuations. Specifically, for reactions of up to order two equation (12) becomes

$$
\begin{aligned}
\text{zeroth order} \quad & \varnothing \rightarrow & f_r(\mathbf{n}) &= k_r \Omega, \\
\text{first order} \quad & A \rightarrow & f_r(\mathbf{n}) &= k_r n_A, \\
\text{second order} \quad & A + B \rightarrow & f_r(\mathbf{n}) &= \frac{k_r}{\Omega} n_A n_B, \\
\text{second order} \quad & A + A \rightarrow & f_r(\mathbf{n}) &= \frac{k_r}{\Omega} n_A (n_A - 1).
\end{aligned}
\tag{13}
$$

Let us now consider how we can mathematically describe the dynamics of such a system. To this end, consider the probability distribution $P(\mathbf{n}, t | \mathbf{n}_0, t_0)$ for the system to be in state $\mathbf{n}$ at time t given that it was in state $\mathbf{n}_0$ at time t_0. We will use the shorthand $P(\mathbf{n}, t) = P(\mathbf{n}, t | \mathbf{n}_0, t_0)$ in the following and implicitly assume conditioning on an initial state. The probability $P(\mathbf{n}, t + \mathrm{d}t)$ after an infinitesimal time step $\mathrm{d}t$ is given by $P(\mathbf{n}, t)$ plus the probability to transition into state $\mathbf{n}$ from a different state $\mathbf{n}^*$ minus the probability to leave state $\mathbf{n}$, which leads us to

$$P(\mathbf{n}, t + \mathrm{d}t) = P(\mathbf{n}, t) + \mathrm{d}t \left(\sum_{r=1}^{R} f_r(\mathbf{n} - \mathbf{S}_r) P(\mathbf{n} - \mathbf{S}_r, t) - \sum_{r=1}^{R} f_r(\mathbf{n}) P(\mathbf{n}, t) \right), \tag{14}$$

where we used the fact that probability of the rth reaction to happen in an infinitesimal time interval $\mathrm{d}t$ is given by $f_r(\mathbf{n})\mathrm{d}t$ and where $\mathbf{S}_r$ is the rth column of the stoichiometric matrix $\mathbf{S}$. Subtracting $P(\mathbf{n}, t)$, dividing by $\mathrm{d}t$ and taking the limit $\mathrm{d}t \to 0$ gives the *chemical master equation* (CME) [3],

$$\partial_t P(\mathbf{n}, t) = \sum_{r=1}^{R} f_r(\mathbf{n} - \mathbf{S}_r) P(\mathbf{n} - \mathbf{S}_r, t) - \sum_{r=1}^{R} f_r(\mathbf{n}) P(\mathbf{n}, t). \tag{15}$$

Since $\mathbf{n}$ is a discrete-valued vector, equation (15) is a coupled system of linear ordinary differential equations. Note that this system is infinite whenever $\mathbf{n}$ is unbounded. Despite its simple structure, there are generally no analytic solutions known to the CME. We call a distribution $P(\mathbf{n}, t)$ a *steady-state solution* of the CME if it solves the CME and fulfils $\partial_t P(\mathbf{n}, t) = 0$.

In the context of chemical reactions, one of the first applications of the CME was done by Delbrück in 1940 for describing the dynamics of an autocatalytic reaction [53]. It has later been applied to linear reactions by Bartholomay [54] and McQuarrie [55], and bimolecular reactions by Ishida [56] and McQuarrie *et al* [57]. Gillespie derived the CME from molecular physics of a dilute and well-mixed gas in 1992 [3], which he extended to liquids in 2009 [58].

Example. Consider again the gene system in figure 3 with reactions given in equation (5). The corresponding stoichiometric matrix $\mathbf{S}$ is given in equation (10). Let n_1 and n_2 be the molecule numbers of the gene in the unbound state G^{on} and the protein P, respectively. Let us assume that there is only one gene in the system, which implies that the number of bound genes G^{off} is $1 - n_1$. Using equation (12) we find the propensity vector

$$\mathbf{f}(\mathbf{n}) = (k_1 n_1, \frac{k_2}{\Omega} n_1 n_2, k_3(1 - n_1), k_4 n_2)^T. \tag{16}$$

The corresponding CME becomes (see equation (15))

$$\partial_t P(n_1, n_2, t) = k_1 n_1 P(n_1, n_2 - 1, t) + \frac{k_2}{\Omega}(n_1 + 1)(n_2 + 1) P(n_1 + 1, n_2 + 1, t)$$

$$+ k_3(2 - n_1) P(n_1 - 1, n_2 - 1, t) + k_4(n_2 + 1) P(n_1, n_2 + 1, t)$$

$$- (k_1 n_1 + \frac{k_2}{\Omega} n_1 n_2 + k_3(1 - n_1) + k_4 n_2) P(n_1, n_2, t). \tag{17}$$

Despite having a relatively simple system here with effectively only two species, no time-dependent solution for its CME in equation (17) is known to our knowledge. A solution in steady state has been derived in [59], but for most other systems not even a steady state solution is available. Therefore, one generally needs to rely on stochastic simulations or approximations to study the behaviour of such systems.

3.2.2. Moment equations.　Suppose we are not interested in the whole distribution solution of the CME but only in its first few moments, say the mean and variance. Starting from the CME in equation (15) one can derive time evolution equations for the moments of its solution as follows. To obtain the time evolution equation for the moment $\langle n_i \dots n_l \rangle$ we multiply equation (15) by $n_i \dots n_l$ and sum over all molecule numbers, leading to

$$\partial_t \langle n_i \dots n_l \rangle = \sum_{r=1}^{R} \langle (n_i + S_{ir}) \dots (n_l + S_{lr}) f_r(\mathbf{n}) \rangle - \sum_{r=1}^{R} \langle n_i \dots n_l f_r(\mathbf{n}) \rangle. \tag{18}$$

Here, $\langle \cdot \rangle$ denotes the expectation with respect to the solution $P(\mathbf{n}, t)$ of the CME in equation (15). For moments of up to order two equation (18) becomes

$$\partial_t \langle n_i \rangle = \sum_{r=1}^{R} S_{ir} \langle f_r(\mathbf{n}) \rangle, \tag{19}$$

$$\partial_t \langle n_i n_j \rangle = \sum_{r=1}^{R} [S_{jr} \langle n_i f_r(\mathbf{n}) \rangle + S_{ir} \langle f_r(\mathbf{n}) n_j \rangle + S_{ir} S_{jr} \langle f_r(\mathbf{n}) \rangle]. \tag{20}$$

We see that if all $f_r(\mathbf{n})$ are zeroth or first-order polynomials in $\mathbf{n}$, i.e. the system is linear without any bimolecular or higher order reactions, the equation of a moment of order m depends only on moments of order m or lower, i.e. the equations are not coupled to higher order equations. The equations up to a certain order hence constitute a finite set of linear ordinary differential equations which can be readily solved numerically or by matrix exponentiation as will be described in the context of the CME in section 3.4. Note that no approximation has been made here which means that the equations describe the exact moments of the CME's solution. This means that *for linear systems, the exact moments up to a finite order of the process can be obtained by solving a finite set of linear ordinary differential equations*.

　　If the systems is non-linear, i.e. contains bimolecular or higher order reactions, the equation of a certain moment depends on higher order moments. This means that the moment equations of different orders are coupled to each other, leading to an infinite hierarchy of coupled equations. This can obviously not be solved directly but gives the basis of approximation methods such as *moment closure approximations* which we will discuss in section 4.3.

Example. Let us again consider the gene system in figure 3 with reactions given in equation (5). The corresponding stoichiometric matrix and propensity vector are given in equation (10). Using these in equations (19) and (20) one obtains

$$\partial_t y_1 = -\frac{k_2}{\Omega} y_{1,2} + k_3(1 - y_1), \tag{21}$$

$$\partial_t y_2 = k_1 y_1 - \frac{k_2}{\Omega} y_{1,2} + k_3(1 - y_1) - k_4 y_2, \tag{22}$$

$$\partial_t y_{1,1} = \frac{k_2}{\Omega}(-2y_{1,1,2} + y_{1,2}) + k_3(-2y_{1,1} + y_1 + 1), \tag{23}$$

$$\partial_t y_{1,2} = k_1 y_{1,1} + \frac{k_2}{\Omega}(-y_{1,1,2} - y_{1,2,2} + y_{1,2}) + k_3(-y_{1,1} - y_{1,2} + y_2 + 1) - k_4 y_{1,2}, \tag{24}$$

$$\partial_t y_{2,2} = k_1(2y_{1,2} + y_1) + \frac{k_2}{\Omega}(-2y_{1,2,2} + y_{1,2}) + k_3(-2y_{1,1} - y_1 + 2y_2 + 1) + k_4(-2y_{2,2} + y_2), \qquad (25)$$

where we introduced the shorthand

$$y_{i_1,\ldots,i_k} = \langle n_{i_1} \ldots n_{i_k} \rangle. \qquad (26)$$

Note that the equations for the first order moments in equations (21) and (22) depend on the second moment $y_{1,2}$, and that the equations for the second moments (23)–(25) depend on the third moments $y_{1,1,2}$ and $y_{1,2,2}$. Similarly, it is easy to see that moment equations of any order depend on higher order moments, which means that we have an infinite system of coupled equations. Note that all terms in equations (21)–(25) depending on higher order moments are proportional to the rate constant k_2 of the bimolecular reaction in equation (5), illustrating that the moment equations decouple in the absence of bimolecular (and higher order) reactions. This could be achieved here by setting $k_2 = 0$ for which the moment equations would decouple and could thus be solved numerically.

3.2.3. Non-mass-action propensity functions. So far we only considered propensity functions of mass-action type defined in equation (12). However, in the literature propensity functions that are not of mass-action kinetics type are frequently used, such as Michaelis–Menten or Hill functions to model the dependence of the production rate of a product on the substrate concentration in an enzymatic catalysis or the dependence of the expression level of a gene on its transcription factor. Such propensity functions typically arise in reduced models where an effective reaction replaces several microscopic reactions. For a gene that becomes regulated by a transcription factor, for example, the binding of the transcription factor to the promoter of the gene is not modelled explicitly, but the effect of its concentration included in the modified propensity function of the expression reaction. Such non-mass-action type reactions should thus normally be seen as an effective reaction replacing a set of mass-action type reactions.

A possible reduced model of the gene expression system in figure 3, for example, eliminates the gene from the system and combines the binding, unbinding and protein production reactions into one effective reaction $\varnothing \to P$ with a Michaelis–Menten type propensity function $k_1 k_3 n_P / (k_2 n_P + k_3)$, which means that the reaction producing protein now depends on the protein number n_P. Such reductions often emerge from separations of time-scales in a system and become exact in certain limits, see section 4.6 for more details.

Independently of the origin of a non-mass-action propensity function, as long as it is interpreted as the firing rate of the corresponding reaction, the CME description remains valid. Unless explicitly stated otherwise, we will assume mass-action kinetics in this article.

3.3. Stochastic simulations

As mentioned above, there are generally no analytic solutions known to the CME. However, it is possible to directly simulate the underlying process. The stochastic simulation algorithm (SSA) is a popular Monte Carlo method that allows one to simulate exact sample paths of the stochastic process described by the CME.

The stochastic simulation algorithm was first proposed in the context of chemical kinetics by Gillespie [4, 5], and several variants have been proposed in the literature, see [9–11] for reviews. The basic idea is to simulate reaction events explicitly in time and to update the time and state vector accordingly. The stochastic process described by the CME is a continuous-time Markov jump process, which has the important property that waiting times, i.e. time

intervals between successive reaction events, are exponentially distributed [15]. Since it is easy to sample from exponential distributions, it is straightforward to simulate the occurrences of chemical reactions.

One example is the so-called *direct method* [4], which samples the time step τ for the next reaction to happen and subsequently which of the different reactions occurs. Specifically, let $p(\tau|\mathbf{n}, t)$ be the probability for the next reaction to happen in an infinitesimal time interval $\mathrm{d}t$ around $\tau + t$, given that the state of the system is $\mathbf{n}$ at time t, and $p(r|\mathbf{n}, t)$ the probability that the next reaction is a reaction of type r. Using that $f_r(\mathbf{n})\mathrm{d}t$ is the probability for the rth reaction to happen in $\mathrm{d}t$, it can be shown that [4]

$$p(\tau|\mathbf{n}, t) = \lambda \exp(-\tau\lambda), \quad \lambda = \sum_{r=1}^{R} f_r(\mathbf{n}), \tag{27}$$

$$p(r|\mathbf{n}, t) = \frac{f_r(\mathbf{n})}{\lambda}. \tag{28}$$

Samples from equations (27) and (28) can be respectively obtained as

$$\tau = -\ln(u_1)/\lambda, \tag{29}$$

$$r = \text{smallest integer satisfying } \sum_{i=1}^{r} f_r(\mathbf{n}) > u_2\lambda, \tag{30}$$

where u_1 and u_2 are uniform random numbers between 0 and 1. The direct SSA method iteratively updates the state vector and time of the process by first sampling the time point of the next reaction event according to equation (29) and subsequently sampling which reaction happens according to equation (30).

Unfortunately, the applicability of the SSA is severely limited due to its computational cost. Since each and every reaction is simulated explicitly, the SSA becomes computationally expensive even for systems with few species. This is particularly the case if the molecule numbers have large fluctuations or if many reactions happen per unit time. In the first case a large number of samples have to be simulated to obtain statistically accurate results, whereas in the second case single simulations become expensive since the time between reaction events becomes small.

3.3.1. Extrinsic noise. So far we have only discussed stochastic simulations of systems with propensity functions that depend solely on the system's molecule numbers and the system's volume. This implies that the propensity functions are constant between reactions events, leading to inter-reaction intervals being exponentially distributed. However, as already mentioned in section 3.2.1, it is frequently of interest to consider stochastically varying propensity functions. This could for example represent the effect of a randomly fluctuating environment on a cell.

To model the effect of extrinsic noise, one typically includes a stochastic variable in one or several propensity functions, whose dynamics may for example be governed by a stochastic differential equation. This means that the corresponding propensities become explicit (stochastic) functions of time. While the CME in equation (15) is still valid in this case, stochastic simulation algorithms discussed above are not since the inter-reaction times are now not exponentially distributed anymore.

The simplest approach is to assume the propensity functions to be constant between reaction events, which allows one to use the standard simulation algorithm [4, 60]. However,

this is of course an approximation and can lead to inaccuracies if the extrinsic process (and hence the corresponding propensity functions) fluctuates strongly between consecutive reaction times. The most straightforward way to simulate such processes exactly is to integrate the propensity functions step-wise over time until a certain target value is reached [61–63]. Apart from numerical integration errors this class of algorithms is exact. However, due to the numerical integration it becomes computationally highly expensive. More recently an alternative exact method has been proposed [64]. This method introduces an additional reaction channel in such a way that the total propensity (i.e. λ in equation (27)) is constant between successive reactions. This in turn allows one to use a standard SSA algorithm on the augmented system. It is hence generally more efficient than the aforementioned integral methods.

3.4. Exact results

As pointed out before, for most systems no analytic solutions of the CME are known. However, for certain classes of systems analytic solutions do exist. For some classes the general time-dependent case can be solved while others can be solved only in steady state. Moreover, analytic solutions have been derived for various simple example systems that do not fall under any of these classes. We give here first an overview of general classes of systems that can be solved analytically, and subsequently list some additional systems for which analytic solutions are known.

3.4.1. Finite state space. Suppose the state space of **n** is finite with M elements. Let us associate with each state a probability $p_i(t), i = 1, \ldots, M$. In section 3.2 we have noted that in this case the CME in equation (15) is a finite system of coupled ordinary differential equations. If we write $\mathbf{p}(t) = (p_1(t), \ldots, p_M(t))^T$, the CME can be written in the matrix form as

$$\partial_t \mathbf{p}(t) = \mathbb{E}\mathbf{p}(t), \tag{31}$$

where $\mathbb{E}$ is a $M \times M$ matrix whose elements can be easily derived for a given system using equation (15). The solution of equation (31) is simply given by

$$\mathbf{p}(t) = \exp(\mathbb{E}t)\mathbf{p}(0). \tag{32}$$

Therefore, for systems with finite state space a solution of the CME for all times can be in principle computed using equation (32). However, even for systems with finite state space, the dimension of equation (32) is often quite large in practice, making matrix exponentiation computationally expensive. Efficient numerical methods for matrix exponentiation have been developed in recent years [65, 66], but for many chemical reaction systems of practical interest it remains an intractable task. For systems with large or infinite state space equation (32) is hence not of direct practical use. It forms the basis for certain approximation methods, however, see section 4.6.

3.4.2. Linear systems. In section 3.2.2 we found that the moment equations decouple from higher order equations for linear reaction systems, i.e. systems without bimolecular or higher order reactions. The moment equations up to a finite order can hence be directly solved numerically for such systems. For non-linear systems, in contrast, this is not possible since the moment equations couple to higher orders. Similarly, exact solutions for the whole distribution solution of the CME are easier to obtain for linear systems. Often this can be done by means of the *generating function method* which transforms the CME into a partial differential

equation [15]. The form of the latter becomes particularly simple for linear systems and can often be solved analytically.

In 1966 Darvey and Staff used the generating function approach to derive a multinomial solution for $P(\mathbf{n}, t)$ for multinomial initial conditions and for systems with $\sum_{i=1}^{N} s_{ir} = \sum_{i=1}^{N} s'_{ir} = 1$ for all reactions r, i.e. exactly one reactant and exactly one product molecule for all reactions [67]. Note that these are closed systems. Later Gardiner used the so-called 'Poisson representation' to show that for systems with $\sum_{i=1}^{N} s_{ir}, \sum_{i=1}^{N} s'_{ir} \leqslant 1$ for all reactions r and Poisson product initial conditions, the solution of the CME is a Poisson product for all times [68]. The auto- and cross-correlation functions for such systems have more recently been computed in [69]. These results have been generalised to systems with $\sum_{i=1}^{N} s_{ir}, \sum_{i=1}^{N} s'_{ir} \leqslant 1$ for all reactions r, with arbitrary initial conditions in [70]. In this case the solution can be written as a convolution of multinomial and product Poisson distributions. In all these cases, the solutions can be constructed from the solution of the deterministic rate equations discussed in section 3.1 and can hence be obtained efficiently by (numerically) solving a finite set of ordinary differential equations.

We would like to stress that the methods described here do not apply to all linear reactions, but only to those with less than two product molecules ($\sum_{i=1}^{N} s'_{ir} \leqslant 1$). This excludes reactions of the type $A \rightarrow A + B$ which are common in biology, such as translation of a protein B from its mRNA, A.

3.4.3. Non-linear systems. As discussed in the previous section, for linear systems where each reaction has at most one product molecule, an analytic solution of the CME can be derived. In contrast, for non-linear systems or linear systems including reactions with two product molecules no analytic solutions are known in general. However, for certain subclasses of systems it is possible to derive steady-state solutions, i.e. solutions satisfying $\partial_t P(\mathbf{n}, t) = 0$. One example are *reversible systems* that obey *detailed balance*, which is also called *thermodynamic equilibrium*. By 'reversible' we mean that each reaction possesses a corresponding reversing reaction. Note that the steady-state condition $\partial_t P(\mathbf{n}, t) = 0$ in the CME in equation (15) merely means that the probability flow out of a state (second term in the CME) equals the probability flow into a state (first term in the CME). These probabilities consist of sums over all reactions. Detailed balance is a stronger condition, as it requires the flow of each single reaction in each state to be balanced by its corresponding reversible reaction. Specifically, let r and r' be the indices of two reactions that revert each other and let $P(\mathbf{n})$ be a steady-state solution. The detailed balance condition reads [15]

$$f_r(\mathbf{n})P(\mathbf{n}) = f_{r'}(\mathbf{n} + \mathbf{S}_r)P(\mathbf{n} + \mathbf{S}_r), \tag{33}$$

which needs to be fulfilled for all reactions and all $\mathbf{n}$. Note that $\mathbf{S}_r = -\mathbf{S}_{r'}$. The left side of equation (33) is the flow out of state $\mathbf{n}$ due to the rth reaction, while the right side is the flow into state $\mathbf{n}$ due to the backward reaction r'. An analogue detailed balance condition can be formulated for the deterministic rate equations. It can be shown that for a reversible reaction system with mass-action kinetics, the stochastic system is in detailed balance if and only if the deterministic system is in detailed balance [71]. If the detailed balance condition is fulfilled, the solution of the CME is a product of Poisson distributions times a function accounting for conservation laws in molecule numbers, where the mean values of the Poisson distributions are simply given by the detailed balance solutions of the rate equations [15, 72].

As an example, consider a closed, bimolecular system with reactions $A + A \underset{k_2}{\overset{k_1}{\rightleftharpoons}} B$. Let n_A and n_B be the molecule numbers of A and B respectively. The quantity $n_A + 2n_B = C$ is conserved. The corresponding steady-state solution of the CME in detailed balance reads

$$P(n_A, n_B) = \frac{e^{-\alpha_1}\alpha_1^{n_A}}{n_A!}\frac{e^{-\alpha_2}\alpha_2^{n_B}}{n_B!}\delta(n_A + 2n_B - C), \tag{34}$$

where $\alpha_1 = \Omega\,\phi_1^0$ and $\alpha_2 = \Omega\,\phi_2^0$, ϕ_1^0 and ϕ_2^0 are the steady-state solutions of the rate equations of A and B, respectively, Ω is the system volume, and δ is the delta function accounting for the conservation law.

These results have more recently been generalised to *weakly reversible* reaction systems [73]. A system is called weakly reversible when for each reaction the change in state space can be reversed by a chain of other reactions. For example, a system with the reactions $X_1 \rightarrow X_2 \rightarrow X_3 \rightarrow X_1$ is not reversible but it is weakly reversible. Weak reversibility is a generalisation of reversibility since each reversible system is also weakly reversible. Correspondingly, the concept of *complex balance* has been employed which is a generalisation of detailed balance [74]. It can be shown that if the rate equations of a system with mass-action kinetics are complex balanced, the CME possesses a steady-state solution given by a product of Poisson distributions times a function accounting for conservation laws [73]. The mean values of the Poisson distributions are given by the complex-balance solutions of the rate equations. Moreover, it can be shown that for any mass-action kinetics reaction system with deficiency zero (see [75] for a definition), the rate equations have a complex-balance solution if and only if the system is weakly reversible [75].

3.4.4. Exact results for special cases. Most reaction systems of practical interest are neither linear nor do they satisfy the detailed/complex balance conditions. However, the CME has been solved for several specific systems that do not fall under any of these categories. In particular for systems with only a single reaction it is often straightforward to obtain a time-dependent solution of the CME. Some examples are given below. Most of these solutions are obtained either by means of the aforementioned generating function method or the Poisson representation. While the former transforms the CME into a partial differential equation, the latter maps the discrete process underlying the CME to a continuous diffusion process governed by a Langevin equation defined in a complex-valued space [68]. We would like to stress that both approaches are exact. However the Poisson representation has not been used much in the literature, probably because of its relative complexity compared to other methods and hence we do not cover it in detail here. We refer the reader to the following references for specific applications [76–82]. For more details on the Poisson representation and the generating function approach we refer the reader to [15].

Single reaction systems.
Time-dependent solutions have been derived for

- $A \rightarrow A + A$ [53]
- $A + B \rightarrow C$ [56]
- $A + A \rightarrow C$ [56]
- $A + B \leftrightarrow C$ [83]

Gene regulatory networks.

- $G^{\mathrm{on}} \rightarrow G^{\mathrm{on}} + P, G^{\mathrm{on}} + P \leftrightarrow G^{\mathrm{off}}, P \rightarrow \varnothing, G^{\mathrm{off}} \rightarrow G^{\mathrm{off}} + P, G^{\mathrm{off}} \rightarrow G^{\mathrm{on}}$: a steady-state solution has been derived if there is one gene in the system [59]. Note that if we ignore the last two reactions this system corresponds to the gene system in figure 3.

- A system similar to the previous one but with geometrically distributed bursts of protein production has been solved in steady state in [84].

Enzyme systems.

- $E + S \leftrightarrow C \rightarrow E + P$: A time-dependent solution has been derived for the case of a single enzyme in the system [85].
- Adding the reaction $E + P \rightarrow C$ to the previous system makes it reversible and a corresponding steady-state solution has been derived in [86].
- The system with added substrate input reaction $\varnothing \rightarrow S$ has been solved in steady state in [87].
- A similar system but with multiple substrate species competing for the enzyme has been solved in steady state in [88].

Most of these examples only consider steady-state solutions and all of them consider systems with few species. The lack of solutions for more complicated systems makes the development of approximation methods necessary.

4. Approximation methods

As discusses in the previous section, analytic solutions of the CME are known only for very restrictive classes of systems and few simple special cases. For most systems of practical interest, no analytic solutions are known to date. Stochastic simulation algorithms introduced in section 3.3 allow to simulate exact samples of the underlying stochastic process, but they quickly become computationally infeasible for larger systems. For these reasons significant effort has been spent in the literature on the development of approximation methods. We give here an introduction to a wide range of such methods. First, we give a detailed introduction to a few approximation methods that can be applied to (almost) arbitrary systems without any pre-knowledge of the system necessary. We derive the approximations, give examples and discuss their properties. Subsequently, we give an overview of other types of approximation methods developed in the literature.

The first method aims at approximating the solution of the CME, namely the *chemical Fokker–Planck equation* (CFPE) and the associated *chemical Langevin equation* (CLE), which we introduce in section 4.1. The CFPE/CLE define an approximating stochastic process. An alternative method for approximating the solution of the CME is given by the *system size expansion*, which constitutes a systematic expansion of the CME in the inverse volume size. The system size expansion includes the popular *linear noise approximation* (LNA) and also provide an efficient way of approximating the moments of a process. We discuss the system size expansion in section 4.2. Next, in section 4.3 we introduce a certain class of *moment closure approximations*, which approximate the moments of a process. In section 4.4 we show how such approximate moments can be used to construct distributions using the maximum entropy principle. Next, in section 4.5 we review software packages implementing the discussed approximation methods. Finally, in section 4.6 we give a brief overview of other approximation methods found in the literature. As we shall see, many of these methods use the CLE, the system size expansion or moment closure approximations as building blocks.

4.1. The chemical Langevin equation

The *chemical Langevin equation* (CLE) and the corresponding *chemical Fokker–Planck equation* (CFPE) constitute a popular diffusion approximation of the CME. Kramers and Moyal derived the CFPE by applying a Taylor expansion to the CME which upon truncation leads to a partial differential equation approximation of the CME [89, 90]: Suppose we let the variables in the CME in equation (15) become continuous and let $\mathbf{x} = (x_1, \ldots, x_N)$, where x_i is the continuous variable denoting the molecule number of species X_i. Performing a Taylor expansion to second order around $\mathbf{x}$ in the first term of the rhs of equation (15) gives (with $\mathbf{n}$ replaced by $\mathbf{x}$)

$$
\begin{aligned}
f_r(\mathbf{x} - \mathbf{S}_r)P(\mathbf{x} - \mathbf{S}_r, t) \approx & f_r(\mathbf{x})P(\mathbf{x}, t) - \sum_{i=1}^{N} S_{ir}\frac{\partial}{\partial x_i}[f_r(\mathbf{x})P(\mathbf{x}, t)] \\
& + \sum_{i,j=1}^{N} S_{ir}S_{jr}\frac{\partial}{\partial x_i}\frac{\partial}{\partial x_j}[f_r(\mathbf{x})P(\mathbf{x}, t)].
\end{aligned} \tag{35}
$$

Inserting this into the CME in equation (15), we see that the first term on the rhs of equation (35) cancels the last term on the rhs of equation (15) leading to the CFPE:

$$
\partial_t P(\mathbf{x}, t) = -\sum_{i=1}^{N} \frac{\partial}{\partial x_i}[A_i(\mathbf{x})P(\mathbf{x}, t)] + \frac{1}{2}\sum_{i,j=1}^{N}\frac{\partial}{\partial x_i}\frac{\partial}{\partial x_j}[B_{ij}(\mathbf{x})P(\mathbf{x}, t)], \tag{36}
$$

where the drift vector $\mathbf{A}$ and diffusion matrix $\mathbf{B}$ are respectively given by

$$
A_i(\mathbf{x}) = \sum_{r=1}^{R} S_{ir}f_r(\mathbf{x}), \tag{37}
$$

$$
B_{ij}(\mathbf{x}) = \sum_{r=1}^{R} S_{ir}S_{jr}f_r(\mathbf{x}). \tag{38}
$$

Note that the drift vector $\mathbf{A}(\mathbf{x})$ and diffusion matrix $\mathbf{B}(\mathbf{x})$ do not depend on time. Note also that whereas the state variables denote discrete molecule numbers in the CME, they denote continuous real numbers in the CFPE. The CFPE in equation (36) is equivalent to the CLE

$$
d\mathbf{x} = \mathbf{A}(\mathbf{x})dt + \mathbf{C}(\mathbf{x})d\mathbf{W}, \qquad \mathbf{C}(\mathbf{x})\mathbf{C}(\mathbf{x})^T = \mathbf{B}(\mathbf{x}), \tag{39}
$$

which is an *Ito stochastic differential equation* [15]. $\mathbf{W}$ in equation (39) is a multi-dimensional Wiener process. It can be shown [15] that the distribution of a process described by equation (39) agrees exactly with the solution of the corresponding Fokker–Planck equation in (36). One can thus interpret the CLE in equation (39) as a generator of realisations of the stochastic process described by the corresponding CFPE. In this sense, the CLE and CFPE are considered to be equivalent to each other. An alternative derivation, which leads directly to the CLE in equation (39), was given by Gillespie in [91].

Generally there exist different choices for $\mathbf{C}(\mathbf{x})$ in equation (39) corresponding to different factorisations of the matrix $\mathbf{B}(\mathbf{x})$; these lead to as many different representations of the CLE. One possibility is given by

$$dx_i = \sum_{r=1}^{R} S_{ir} f_r(\mathbf{x}) dt + \sum_{r=1}^{R} S_{ir} \sqrt{f_r(\mathbf{x})} dW_r, \quad i = 1, \ldots, N. \tag{40}$$

This representation of the CLE is the one most commonly used in the literature [91, 92].

Example. Let us come back to the gene expression system in figure 3 with reactions in equation (5) and consider the corresponding CFPE and CLE. Using the stoichiometric matrix in equation (10) and propensity vector in equation (16) we obtain for the drift vector and diffusion matrix defined in equation (37) and equation (38), respectively,

$$\mathbf{A}(\mathbf{x}) = \begin{pmatrix} -\dfrac{k_2}{\Omega} x_1 x_2 + k_3(1 - x_1) \\ k_1 x_1 - \dfrac{k_2}{\Omega} x_1 x_2 + k_3(1 - x_1) - k_4 x_2 \end{pmatrix}, \tag{41}$$

$$\mathbf{B}(\mathbf{x}) = \begin{pmatrix} \dfrac{k_2}{\Omega} x_1 x_2 + k_3(1 - x_1) & \dfrac{k_2}{\Omega} x_1 x_2 + k_3(1 - x_1) \\ \dfrac{k_2}{\Omega} x_1 x_2 + k_3(1 - x_1) & k_1 x_1 + \dfrac{k_2}{\Omega} x_1 x_2 + k_3(1 - x_1) + k_4 x_2 \end{pmatrix}, \tag{42}$$

where x_1 and x_2 are the (continuous) particle numbers of G^{on} and P, respectively. To obtain the CLE in equation (39), we have to compute $\mathbf{C}$, which is the square root of $\mathbf{B}$ and thus generally not uniquely defined. One possibility as given in equation (40) reads

$$C_{ij}(\mathbf{x}) = \begin{pmatrix} 0 & -\sqrt{\dfrac{k_2}{\Omega} x_1 x_2} & \sqrt{k_3(1 - x_1)} & 0 \\ \sqrt{k_1 x_1} & -\sqrt{\dfrac{k_2}{\Omega} x_1 x_2} & \sqrt{k_3(1 - x_1)} & -\sqrt{k_4 x_2} \end{pmatrix}. \tag{43}$$

This gives rise to the CLE

$$dx_1 = \left(-\frac{k_2}{\Omega} x_1 x_2 + k_3(1 - x_1)\right) dt - \sqrt{\frac{k_2}{\Omega} x_1 x_2}\, dW_2 + \sqrt{k_3(1 - x_1)}\, dW_3, \tag{44}$$

$$dx_2 = \left(k_1 x_1 - \frac{k_2}{\Omega} x_1 x_2 + k_3(1 - x_1) - k_4 x_2\right) dt$$
$$+ \sqrt{k_1 x_1}\, dW_1 - \sqrt{\frac{k_2}{\Omega} x_1 x_2}\, dW_2 + \sqrt{k_3(1 - x_1)}\, dW_3 - \sqrt{k_4 x_2}\, dW_4, \tag{45}$$

where the W_i are independent Wiener processes. Note that it does not make a difference if one changes the signs in front of the square roots in equation (43), or equivalently in front of the noise terms in equation (44), as long as one does so simultaneously for each occurrence of a specific term, i.e. changes the sign of whole columns in equation (43). To see that such changes are equivalent note that the diffusion matrix $\mathbf{B} = \mathbf{C}\mathbf{C}^T$ of the CFPE is invariant under such changes. This can also be seen directly from the CLE in equations (44) and (45) since the Wiener processes are symmetric.

4.1.1. Stochastic simulations. As for the CME, there are no analytic solutions known for the CLE for most systems. However, the computational cost of CLE simulations scales with the number of species, rather than with the rate of reaction events as CME simulations. This

means that CLE simulations are often more efficient than simulations of the CME, typically if the molecule counts in a system are not too small.

A popular method to simulate the CLE is the *Euler–Maruyama algorithm* which discretises time into intervals dt and simulates the process iteratively as [93]

$$x(t + \mathrm{d}t) = x(t) + A(x(t))\mathrm{d}t + \sqrt{B(x(t))\mathrm{d}t}\ \mathrm{d}w, \quad \mathrm{d}w \sim \mathcal{N}(0, 1), \tag{46}$$

where $\mathcal{N}(0, 1)$ is a normal distribution with mean 0 and variance 1. The smaller dt, the better the true process is approximated by equation (46) but also the slower the algorithm becomes. The right choice of dt is therefore a tradeoff between accuracy and efficiency. It is important to point out that more efficient simulation methods exist, see for example [93–96].

4.1.2. Properties and recent developments. The CLE is typically a good approximation whenever the molecule numbers of the system are not too small and is particularly useful if one is interested in time trajectories or distributions of a process. It can be shown that the differences between the CLE and the CME tend to zero in the limit of large molecule numbers [97]. In other words, the CLE becomes exact in the thermodynamic limit.

By multiplying the CFPE in equation (36) by $x_i \dots x_l$ and integrating over the whole state space, one obtains ordinary differential equation for the moment $\langle x_i \dots x_l \rangle$ of the process described by the CLE. Importantly, it turns out that the equations for moments of up to order two are *exactly the same as the corresponding equations derived from the CME.* These are given in equations (19) and (20). Note however that since they are generally coupled to higher order moments for which the evolution equations derived from the CLE and CME do not agree, the first two moments (and higher order moments) of the CLE do not generally agree with the ones of the CME. However, since the moment equations decouple for linear systems as shown in section 3.2.2, we obtain the important result that *the moments up to order two of the processes described by the CLE and CME agree exactly for linear reaction systems.*

Note that the CME has a natural boundary at zero molecule numbers, i.e. for a sensible initial condition with zero probability to have a negative number of molecules, this probability remains zero for all times. Until recently it has not been clear how this boundary condition behaves when approximating the discrete process underlying the CME by a continuous process using the CLE. As it turns out, this boundary issue leads to an ill-definedness of the CLE due to occurrences of square roots of negative expressions in finite time with finite probability [99, 100]. Since traditionally the domain of the CLE is (implicitly) assumed to be that of real numbers, the CLE is not well-defined in this case. More recently, it has been shown that this problem is independent of the chosen factorisation of the CFPE's noise matrix **B** (see equation (39)), which means that the CLE is not well-defined for real variables [87]. In this study it has been shown that the same can be expected for the majority of reaction systems.

Several modified versions of the CLE have been proposed that try to keep the state space real, for example [98, 99]. However, these are ad hoc modifications and have been found to introduce high inaccuracies for some non-linear reaction systems [87]. Importantly, they also have been found to violate the CLE's exactness for the moments up to order two for linear systems. Alternatively, the ill-defined problem can be solved by extending the state space of the CLE to complex variables. It has been proven that this leads to real-valued moments, real-valued autocorrelation functions and real-valued power spectra, and to restore the CLE's exactness for the moments up to order two for linear systems [87]. This complex CLE was found to be highly accurate for some non-linear systems in comparison to other modified versions. For one example the results from [87] are shown in figure 4. The complex-valued CLE has other drawbacks, however. For example, it does not directly give approximations of the process or of distributions. Rather, the results have to be projected to real space.
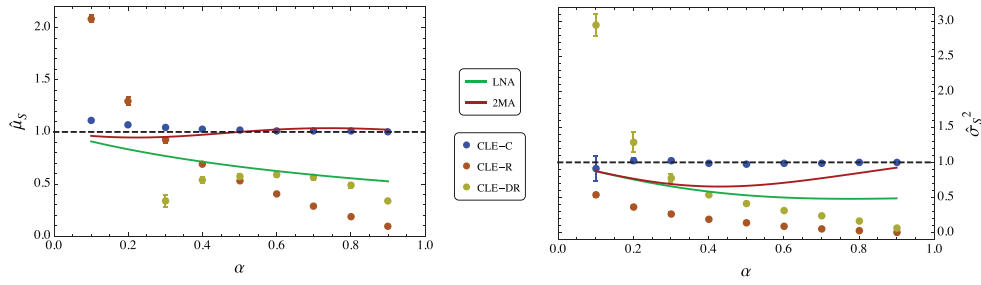
Figure 4. Comparison of different CLE implementations. The figures show the steady-state mean (left) and variance (right) of a substrate in a Michaelis–Menten reaction system as a function of the strength of the substrate-to-enzyme binding strength. The results are normalised by the exact CME result, which means that the horizontal dashed line corresponds to the exact value. CLE-C corresponds to the complex-valued CLE, CLE-R to a real-valued implementation with rejecting boundaries at zero molecules, and CLE-DR to the real-valued version proposed in [98]. LNA and 2MA correspond to the linear noise approximation and the second-order normal moment closure, which we will introduce in sections 4.2 and 4.3, respectively. We observe that both real-valued implementations give large deviations from the CME result, significantly more than the LNA and 2MA. The complex CLE on the other hand is significantly more accurate than all the other methods. Reprinted from [87], with the permission of AIP Publishing.

In recent years the CLE has been frequently used for approximating the dynamics of intermediate or high abundance species in so-called *hybrid methods*, see section 4.6.4. In this case, the probability of negative concentrations becomes very small. If the partitioning into low and high abundance species is done adaptively, the problem of square roots of negative expressions is avoided completely. In this case the complex-valued CLE automatically reduces to the real-valued CLE. The CLE is particularly useful in simulation based hybrid methods since it gives an approximation of the whole process (rather than for example only the process's moments).

Recently, a tensor-based method has been proposed in [101] for the direct numerical solution of the CFPE. This is a promising approach since it avoids computationally expensive ensemble averaging and has been used for sensitivity and bifurcation analysis [101]. Recall that the CLE becomes exact in the limit of large system sizes. As one may expect, the CLE therefore generally captures the multimodality of the CME's solution if the multimodality persists for large volumes, since this is the case if the deterministic rate equations are multi-stable. Surprisingly, however, it has recently been found that the CLE is also able to reproduce *noise-induced multimodality* of the CME's solution, i.e. multimodality that occurs only if the system volume is decreased below a certain critical value [102]. However, it was found that this is not the case for all reaction systems [103]. In [104], the CLE as a diffusion approximation was applied to Petri nets.

4.2. The system size expansion

Suppose we are not interested in approximating the whole process but only in its distribution or first few moments. Running stochastic simulations of the CLE seem like unnecessary computational effort in this case and the question arises if there exist more efficient approximation methods to achieve this. We next discuss the *system size expansion* which aims at

approximating the distribution or the first few moments of a process. The system size expansion is a perturbative expansion of the CME in the inverse system size originally developed by van Kampen [14, 105]. The idea is to separate concentrations into a deterministic part, given by the solution of the deterministic rate equations, and a part describing the fluctuations about the deterministic mean.

4.2.1. Derivation. The system size expansion splits the instantaneous particle numbers n_i of the CME into a deterministic part and a fluctuating part as

$$\frac{n_i}{\Omega} = \phi_i + \Omega^{-1/2}\,\epsilon_i, \quad i = 1, ..., N, \tag{47}$$

where Ω is the volume of the system, ϕ_i is the solution of the deterministic rate equations in (4) and we introduced the new variables ϵ_i representing fluctuations about the deterministic mean. Following [106], the system size expansion proceeds by transforming the master equation to these new variables ϵ_i. To this end, we rewrite the CME in equation (15) as

$$\partial_t P(\mathbf{n}, t) = \sum_{r=1}^{R}\left(\prod_{i=1}^{N} E_i^{-S_{ir}} - 1\right)f_r(\mathbf{n})P(\mathbf{n}, t), \tag{48}$$

where we introduced the step operators $E_i^{-S_{ir}}$, which are defined in terms of their action on a general function of the state space, $h(\mathbf{n}) = h(n_1, ..., n_N)$ as

$$E_i^{-S_{ir}}h(n_1, ..., n_i, ..., n_N) = h(n_1, ..., n_i - S_{ir}, ..., n_N). \tag{49}$$

The system size expansion now proceeds in three steps:

1. **Time derivative.** Due to the change of variables in equation (47) we need to transform the distribution $P(\mathbf{n}, t)$ into a distribution $\Pi(\epsilon, t)$ of the new variables. Note that the time derivate in the CME in equation (48) is taken at constant $\mathbf{n}$, which implies $\partial \epsilon_i/\partial_t = -\Omega^{1/2}\,\partial \phi_i/\partial_t$. This in turn leads to

$$\partial_t P(\mathbf{n}, t) = \frac{\partial \Pi(\epsilon, t)}{\partial t} - \Omega^{1/2}\sum_{i=1}^{N}\frac{\partial \phi_i}{\partial t}\frac{\partial \Pi(\epsilon, t)}{\partial \epsilon_i}. \tag{50}$$

2. **Step operator.** Using equation (47) in the definition of the step operator in equation (49), we obtain

$$\prod_{i=1}^{N} E_i^{-S_{ir}}h(\epsilon_1, , ..., \epsilon_N) = h(\epsilon_1 - \Omega^{-1/2}\,S_{1r}, ..., \epsilon_N - \Omega^{-1/2}\,S_{Nr}). \tag{51}$$

Expanding the rhs in powers of $\Omega^{-1/2}$ we can expand the term of the CME involving the step operators as

$$\prod_{i=1}^{N} E_i^{-S_{ir}} - 1 = -\Omega^{-1/2}\sum_{i=1}^{N} S_{ir}\frac{\partial}{\partial \epsilon_i} + \frac{\Omega^{-1}}{2}\sum_{i,j=1}^{N} S_{ir}S_{jr}\frac{\partial^2}{\partial \epsilon_i \partial \epsilon_j}$$
$$-\frac{\Omega^{-3/2}}{6}\sum_{i,j,l=1}^{N} S_{ir}S_{jr}S_{lr}\frac{\partial^3}{\partial \epsilon_i \partial \epsilon_j \partial \epsilon_l} + O(\Omega^{-2}). \tag{52}$$

3. **Propensity functions.** We assume that the propensity functions in the CME can be expanded as

$$f_r(\mathbf{n}) = \Omega \sum_{i=0}^{\infty} \Omega^{-i} f_r^{(i)}(\frac{\mathbf{n}}{\Omega}). \tag{53}$$

If we assume mass-action kinetics for which the propensity functions take the form in equation (12), the propensity functions are polynomials and it is easy to see that such an expansion always exists. The same is also true for many propensity functions that are not of mass-action kinetics type, such as Michaelis–Menten and Hill-type propensity functions. Using equation (47) and expanding the rhs of equation (53) in $\Omega^{-1/2}$ we obtain

$$f_r(\mathbf{n}) = g_r(\phi) + \Omega^{-1/2} \sum_{i=1}^{N} \epsilon_i \frac{\partial g_r(\phi)}{\partial \phi_i}$$
$$+ \Omega^{-1} \left(\frac{1}{2} \sum_{i,j=1}^{N} \epsilon_i \epsilon_j \frac{\partial^2 g_r(\phi)}{\partial \phi_i \partial \phi_j} - f_r^{(1)} \right) + O(\Omega^{-3/2}), \tag{54}$$

where we have identified $f_r^{(0)}(\phi) = g_r(\phi)$ and $g_r(\phi)$ is the macroscopic rate function of the rth reaction introduced in equation (2) in context of the deterministic rate equations.

Applying equations (50), (52) and (54) to equation (48) we obtain

$$\partial_t \Pi(\epsilon, t) = [\mathcal{L}^{(0)} + \Omega^{-1/2} \mathcal{L}^{(1)} + \Omega^{-1} \mathcal{L}^{(2)} + O(\Omega^{-3/2})] \Pi(\epsilon, t), \tag{55}$$

where we defined

$$\mathcal{L}^{(0)} = -\partial_i J^{(0)j}_{i} \epsilon_j + \frac{1}{2} D^{(0)}_{ij} \partial_i \partial_j, \tag{56}$$

$$\mathcal{L}^{(1)} + \Omega^{-1/2} \left(\frac{1}{2} J^{(0)q}_{i,j} \partial_i \partial_j \epsilon_q - \frac{1}{2} J^{(0)q,r}_{i} \partial_i \epsilon_q \epsilon_r - \frac{1}{6} D^{(0)}_{i,j,k} \partial_i \partial_j \partial_k - D^{(1)}_i \partial_i \right) \Big] \Pi, \tag{57}$$

$$\mathcal{L}^{(2)} + \Omega^{-1} \Bigg(-\frac{1}{6} J^{(0)q}_{i,j,k} \partial_i \partial_j \partial_k \epsilon_q + \frac{1}{4} J^{(0)q,r}_{i,j} \partial_i \partial_j \epsilon_q \epsilon_r - \frac{1}{6} J^{(0)q,r,s}_{i} \partial_i \epsilon_q \epsilon_r \epsilon_s$$
$$- J^{(1)q}_{i} \partial_i \epsilon_q + \frac{1}{2} D^{(1)}_{i,j} \partial_i \partial_j + \frac{1}{24} D^{(0)}_{i,j,k,l} \partial_i \partial_j \partial_k \partial_l \Bigg). \tag{58}$$

Here we use the shorthand $\partial_i = \partial/\partial\epsilon_i$ and assume summations over doubly occurring indices for notational simplicity. We note that in equation (55) also two terms of order $\Omega^{1/2}$ occur. However, these terms cancel each other because they just correspond to the deterministic rate equations. In equations (56)–(58) we defined

$$D^{(n)}_{i,\ldots k} = \sum_{r=1}^{R} S_{ir} \ldots S_{kr} f_r^{(n)}(\phi), \tag{59}$$

$$J^{(n)l,\ldots,m}_{i,\ldots,k} = \partial_{\phi_l} \ldots \partial_{\phi_m} D^{(n)}_{i,\ldots k}. \tag{60}$$

Note that both the $\mathbf{J}^{(i)}$ and $\mathbf{D}^{(i)}$ matrices depend on the solution ϕ of the rate equations in (4) and are thus generally time dependent. Comparing with equation (4) we find that $\mathbf{J}^{(0)}$ in equation (60) is just the Jacobian of the rate equations.

4.2.2. The linear noise approximation. The popular *linear noise approximation* (LNA) [14, 105] is obtained if we truncate the expansion in equation (55) to zeroth order, leading to

$$\partial_t \Pi(\boldsymbol{\epsilon}, t) = \left[-\sum_{i=1}^{N} \frac{\partial}{\partial \epsilon_i} \sum_{j=1}^{N} J^{(0)j}_{\ i} \epsilon_j + \frac{1}{2} \sum_{i,j=1}^{N} D^{(0)}_{ij} \frac{\partial}{\partial \epsilon_i} \frac{\partial}{\partial \epsilon_j} \right] \Pi(\boldsymbol{\epsilon}, t) + O(\Omega^{-1/2}). \tag{61}$$

Equation (61) is a Fokker–Planck equation with drift and diffusion linear and constant in $\boldsymbol{\epsilon}$, respectively, and hence has a multivariate normal solution under appropriate initial conditions. By multiplying equation (61) with ϵ_i and $\epsilon_i \epsilon_j$ and integrating over all $\boldsymbol{\epsilon}$, one obtains ordinary differential equations for the first and second-order moments, $\langle \epsilon_i \rangle$ and $\langle \epsilon_i \epsilon_j \rangle$, respectively. By doing so one finds that if the mean is initially zero, $\langle \epsilon_i \rangle|_{t=0} = 0$, it remains zero for all times. This is normally the case since otherwise the initial condition of the rate equations would not agree with the initial mean value of the stochastic system. If we additionally assume deterministic initial conditions, we also have $\langle \epsilon_i \epsilon_j \rangle|_{t=0} = 0$. The solution of equation (61) is thus a multivariate normal distribution with zero mean. Since $\boldsymbol{n}$ and $\boldsymbol{\epsilon}$ are related by the linear transformation given in equation (47), the distribution of $\boldsymbol{n}$ is also given by a multivariate normal distribution. The mean of the latter satisfies the rate equations in equation (4) and the covariance $\boldsymbol{\Sigma}$ defined as $\Sigma_{ij} = \langle n_i n_j \rangle - \langle n_i \rangle \langle n_j \rangle$ fulfils

$$\partial_t \boldsymbol{\Sigma} = \boldsymbol{J}^{(0)} \boldsymbol{\Sigma} + \boldsymbol{\Sigma} \boldsymbol{J}^{(0)T} + \Omega^{-1} \boldsymbol{D}^{(0)}. \tag{62}$$

The LNA describes the lowest order fluctuations of the system size expansion about the deterministic mean and is valid in the limit of large volumes, i.e. the thermodynamic limit.

An alternative derivation of the LNA is given in [15], where it is derived from the chemical Langevin equation given in equation (39).

Example. Let us consider the rate equations and LNA for the gene system in figure 3. The corresponding stoichiometric matrix, the macroscopic rate vector and the rate equations were already derived in section 3.1 and are given in equations (10) and (11), respectively. The LNA defined in equation (61) requires matrices $\mathbf{J}^{(0)}$ and $\mathbf{D}^{(0)}$ defined in equations (59) and (60), respectively, for which we obtain

$$\mathbf{D}^{(0)} = \begin{pmatrix} k_2\phi_1\phi_2 + k_3(\frac{1}{\Omega} - \phi_1) & k_2\phi_1\phi_2 + k_3(\frac{1}{\Omega} - \phi_1) \\ k_2\phi_1\phi_2 + k_3(\frac{1}{\Omega} - \phi_1) & k_1\phi_1 + k_2\phi_1\phi_2 + k_3(\frac{1}{\Omega} - \phi_1) + k_4\phi_2 \end{pmatrix}, \tag{63}$$

$$\mathbf{J}^{(0)} = \begin{pmatrix} -k_2\phi_2 - k_3 & -k_2\phi_1 \\ k_1 - k_2\phi_1 - k_3 & -k_2\phi_1 - k_4 \end{pmatrix}. \tag{64}$$

Note that $\mathbf{D}^{(0)}$ and $\mathbf{J}^{(0)}$ are functions of the time-dependent solutions ϕ_1 and ϕ_2 of the rate equations in (11). The solution of the LNA is a normal distribution in $\mathbf{n} = (n_1, n_2)$. Its mean is obtained by (numerically) solving the rate equations in (11), and the covariance by subsequently solving equation (62) using equations (63) and (64).

4.2.3. Higher order corrections. Suppose we want to include higher orders in $\Omega^{-1/2}$ in equation (55) beyond the LNA. In contrast to the LNA, now the PDE in equation (55) can generally not be solved analytically. However, we can derive ordinary differential equations for the

moments of the system accurate to the corresponding order in $\Omega^{-1/2}$, as follows. Assuming an expansion of the distribution $\Pi(\epsilon, t)$ as

$$\Pi(\epsilon, t) = \sum_{i=0}^{\infty} \Omega^{-i/2} \Pi^{(i)}(\epsilon, t), \tag{65}$$

one obtains an expansion of the moments

$$\langle \epsilon_j \dots \epsilon_k \rangle = \sum_{i=0}^{\infty} [\epsilon_j \dots \epsilon_k]^{(i)} \Omega^{-i/2}, \qquad [\epsilon_j \dots \epsilon_k]^{(i)} = \int d\epsilon \; \epsilon_j \dots \epsilon_k \Pi^{(i)}(\epsilon, t). \tag{66}$$

Multiplying equation (55) with $\epsilon_j \dots \epsilon_k$, using equations (65) and (66) and integrating over all ϵ, we obtain

$$\partial_t [\epsilon_i]_0 = J^{(0)q}_{i} [\epsilon_q]_0, \tag{67}$$

$$\partial_t [\epsilon_i]_1 = J^{(0)q}_{i} [\epsilon_q]_1 + D^{(1)}_i + \frac{1}{2} J^{(0)q,r}_{i} [\epsilon_q \epsilon_r]_0, \tag{68}$$

$$\partial_t [\epsilon_i]_2 = J^{(0)q}_{i} [\epsilon_q]_2 + \frac{1}{2} J^{(0)q,r}_{i} [\epsilon_q \epsilon_r]_1 + \frac{1}{6} J^{(0)q,r,s}_{i} [\epsilon_q \epsilon_r \epsilon_s]_0 + J^{(1)q}_{i} [\epsilon_q]_0, \tag{69}$$

$$\partial_t [\epsilon_i \epsilon_j]_0 = J^{(0)q}_{i} [\epsilon_j \epsilon_q]_0 + J^{(0)q}_{j} [\epsilon_i \epsilon_q]_0 + D^{(0)}_{ij}, \tag{70}$$

$$\begin{aligned}
\partial_t [\epsilon_i \epsilon_j]_1 = {} & J^{(0)q}_{i} [\epsilon_j \epsilon_q]_1 + J^{(0)q}_{j} [\epsilon_i \epsilon_q]_1 + D^{(1)}_i [\epsilon_j]_0 + D^{(1)}_j [\epsilon_i]_0 \\
& + \frac{1}{2} J^{(0)q,r}_{i} [\epsilon_j \epsilon_q \epsilon_r]_0 + \frac{1}{2} J^{(0)q,r}_{j} [\epsilon_i \epsilon_q \epsilon_r]_0 + J^{(0)q}_{i,j} [\epsilon_q]_0,
\end{aligned} \tag{71}$$

$$\begin{aligned}
\partial_t [\epsilon_i \epsilon_j]_2 = {} & J^{(0)q}_{i} [\epsilon_j \epsilon_q]_2 + J^{(0)q}_{j} [\epsilon_i \epsilon_q]_2 + D^{(1)}_i [\epsilon_j]_1 + D^{(1)}_j [\epsilon_i]_1 \\
& + \frac{1}{2} J^{(0)q,r}_{i} [\epsilon_j \epsilon_q \epsilon_r]_1 + \frac{1}{2} J^{(0)q,r}_{j} [\epsilon_i \epsilon_q \epsilon_r]_1 + J^{(0)q}_{i,j} [\epsilon_q]_1 \\
& + D^{(1)}_{ij} + \frac{1}{6} J^{(0)q,r,s}_{i} [\epsilon_j \epsilon_q \epsilon_r \epsilon_s]_0 + \frac{1}{6} J^{(0)q,r,s}_{j} [\epsilon_i \epsilon_q \epsilon_r \epsilon_s]_0 \\
& + \frac{1}{2} J^{(0)q,r}_{i,j} [\epsilon_q \epsilon_r]_0 + J^{(1)q}_{i} [\epsilon_j \epsilon_q]_0 + J^{(1)q}_{j} [\epsilon_i \epsilon_q]_0,
\end{aligned} \tag{72}$$

$$\partial_t [\epsilon_i \epsilon_j \epsilon_k]_0 = D_{ij} [\epsilon_k]_0 + (i \leftrightarrow j \leftrightarrow k) + J^{(0)q}_{i} [\epsilon_j \epsilon_k \epsilon_q]_0 + (i \leftrightarrow j \leftrightarrow k), \tag{73}$$

$$\begin{aligned}
\partial_t [\epsilon_i \epsilon_j \epsilon_k]_1 = {} & D_{ij} [\epsilon_k]_1 + (i \leftrightarrow j \leftrightarrow k) + J^{(0)q}_{i} [\epsilon_j \epsilon_k \epsilon_q]_1 + (i \leftrightarrow j \leftrightarrow k) \\
& + D^{(1)}_i [\epsilon_j \epsilon_k]_0 + (i \leftrightarrow j \leftrightarrow k) + D_{ijk} + \frac{1}{2} J^{(0)q,r}_{i} [\epsilon_j \epsilon_k \epsilon_q \epsilon_r]_0 \\
& + J^{(0)q}_{i,j} [\epsilon_k \epsilon_q]_0 + (i \leftrightarrow j \leftrightarrow k),
\end{aligned} \tag{74}$$

$$\partial_t [\epsilon_i \epsilon_j \epsilon_k \epsilon_l]_0 = D_{ij} [\epsilon_k \epsilon_l]_0 + (i \leftrightarrow j \leftrightarrow k \leftrightarrow l) + J^{(0)q}_{i} [\epsilon_j \epsilon_k \epsilon_l \epsilon_q]_0 + (i \leftrightarrow j \leftrightarrow k \leftrightarrow l). \tag{75}$$

Here, $(i \leftrightarrow j \leftrightarrow k)$ denotes a sum over the previous term over all cyclic permutations of the indices i,j and k, and similarly for four indices. For example, we have $D_{ij}[\epsilon_k]_1 + (i \leftrightarrow j \leftrightarrow k) = D_{ij}[\epsilon_k]_1 + D_{ki}[\epsilon_j]_1 + D_{jk}[\epsilon_i]_1$. We find that for deterministic initial conditions $[\epsilon_i]_0 = [\epsilon_i]_2 = [\epsilon_i\epsilon_j]_1 = [\epsilon_i\epsilon_j\epsilon_k]_0 = 0$ holds for all times. The moments in molecule numbers n_i can now be obtained from the moments of the ϵ_i variables by using the ansatz in equation (47). For the mean and covariance this leads to

$$\left\langle \frac{n_i}{\Omega} \right\rangle = \phi_i + \Omega^{-1}[\epsilon_i]_1 + O(\Omega^{-2}), \tag{76}$$

$$\left\langle \frac{n_i}{\Omega} \frac{n_j}{\Omega} \right\rangle - \left\langle \frac{n_i}{\Omega} \right\rangle \left\langle \frac{n_j}{\Omega} \right\rangle = \Omega^{-1}[\epsilon_i\epsilon_j]_0 + \Omega^{-2}([\epsilon_i\epsilon_j]_2 - [\epsilon_i]_1[\epsilon_j]_1) + O(\Omega^{-3}). \tag{77}$$

To order Ω^0, equation (76) corresponds to the macroscopic rate equations, while equation (77) to order Ω^{-1} corresponds to the LNA estimate for the covariance. Including terms of order Ω^{-1} in equation (76) gives the leading order corrections to the mean given by the rate equations. To this order, the system size expansion equations have been called 'effective mesoscopic rate equations' in the literature [107]. The next leading order corrections to the covariance, which have been called 'inverse omega square' in the literature [106], are obtained by keeping terms of order Ω^{-2} in equation (77).

4.2.4. Properties and recent developments. As long as one is only interested in the LNA or higher order corrections to the moments, rather than the distributions of higher order truncations, the system size expansion amounts to the solution of finite sets of ordinary differential equations. It is thus generally significantly more efficient than stochastic simulations of the CME or of the CLE.

Since the system size expansion is an expansion around the deterministic mean, it cannot be used for deterministically multistable systems, i.e. systems whose rate equations have two or more positive stable steady states, unless one is only interested in the short-time behaviour of a process (for example for the purpose of inference from time-series data, see section 6).

In many cases, the lowest order of the system size expansion, i.e. the LNA, already gives remarkably accurate results and has been used succesfully in various applications [109–113]. However, for some systems it gives rise to significant deviations from the exact result [114]. Higher order approximations beyond the LNA have frequently been found to be highly accurate for such systems. Examples include oscillations in networks of coupled autocatalytic reactions [115, 116], predator-prey systems [117] and circadian oscillators [118].

We note that for truncations of higher orders than zero, the resulting PDEs involve higher order derivatives than two and hence have no probabilistic interpretation due to non positive-definite solutions [119]. Moreover, the solutions to the PDEs are generally not longer known in this case. However, for systems with only one species a general solution to all orders has recently been derived [108]. A similar approach has been developed in [120] for discrete-time models in neuroscience. Note that the variables described by the system size expansion are typically assumed to be continuous. In [108] it has been found that if the support is assumed to be continuous, the first few leading orders of the system size expansion lead to oscillations in the tail of the Poisson distribution for a linear birth-death process. In contrast, if one assumes a discrete support, the distribution is captured accurately. In the same work, a modified system
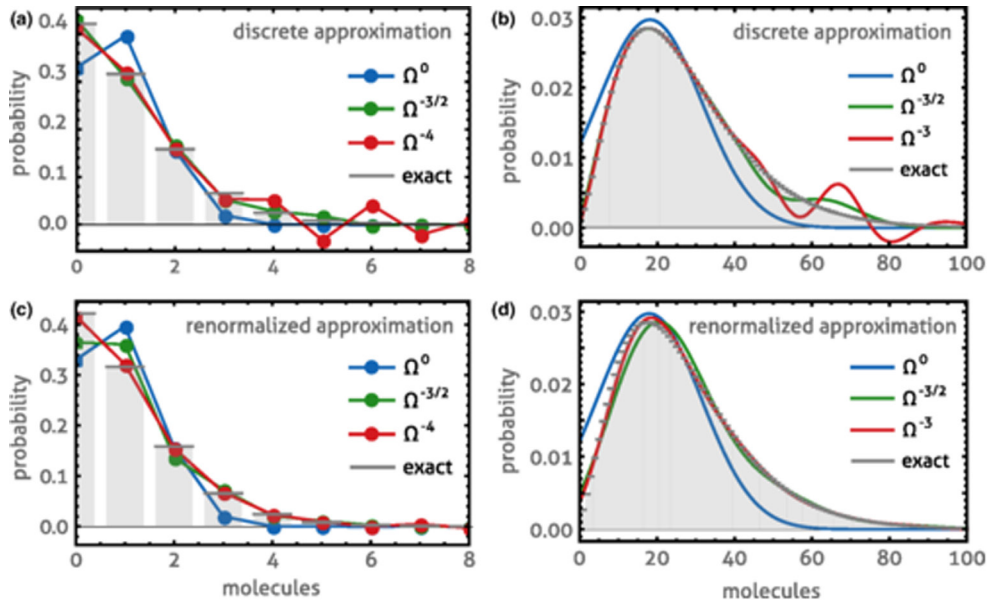
Figure 5. Distribution of non-linear birth-death process. The figures show the steady-state distribution for a non-linear birth-death process for two parameter sets, one corresponding to a small mean number of molecules (left panel) and one corresponding to a large mean number of molecules (right panel). Results are shown for different orders of the system size expansion with discrete support (top panel) as well as the renormalised approximation (see main text for a definition). The studied system corresponds to a birth-death process where the propensity of the death process is of Michaelis–Menten type. Such a system is obtained when reducing a Michaelis–Menten system under time-scale separation. This will be demonstrated in section 4.6.3. Reprinted figure with permission from [108]. Copyright 2015 by the American Physical Society.

size expansion has been employed. Instead of expanding the variables of the CME around the deterministic mean value given by the rate equations, one expands the variables around the mean value given by the system size expansion itself to the considered order (which is not known *a priori*). The authors call this method 'renormalised approximation'. This approach has been found to give an improved approximation of the distribution for a non-linear process as shown in figure 5.

4.3. Moment closure approximations

Another popular class of methods that approximate the first few moments of a process are so-called *moment closure approximations*. We present a popular class of such methods here that are particularly easy to derive and implement.

4.3.1. General formulation. In section 3.2.2 we showed that the CME gives rise to ordinary differential equations for the moments of a process. For linear reaction systems, these equations can be solved numerically up to a certain desired order. For non-linear systems, however, we found that the moment equations of a certain order couple to higher order equations, leading to an infinite hierarchy of equations which can hence not be solved directly. Moment

closure approximations truncate this infinite set of equations at a certain order M in some approximate way.

One popular class of moment closure approximations close the moment equations by expressing all moments above a certain order M as functions of lower-order moments. One way to achieve this is for example by assuming the distribution of the system to have a particular functional form, for example a normal distribution. This decouples the equations of the moments up to order M from higher-order moments, and hence leads to a finite set of coupled ordinary differential equations which can then be solved (numerically). We refer to such a moment closure as 'Mth order moment closure'.

Let

$$y_{i_1,\ldots,i_k} = \langle n_{i_1} \ldots n_{i_k} \rangle, \tag{78}$$

$$z_{i_1,\ldots,i_k} = \begin{cases} \langle (n_{i_1} - y_{i_1}) \ldots (n_{i_k} - y_{i_k}) \rangle & \text{if} \quad k \geqslant 2, \\ y_{i_1} & \text{if} \quad k = 1, \end{cases} \tag{79}$$

$$c_{i_1,\ldots,i_k} = \partial_{s_{i_1}} \ldots \partial_{s_{i_k}} g(s_1, \ldots, s_N)\big|_{s_1,\ldots,s_N=0}, \tag{80}$$

denote raw moments, central moments and cumulants of order k, respectively. We call $y_{i_1,\ldots,i_k}$ a 'diagonal moment' if $i_l = i_m$ for all $l, m \in \{1, \ldots, k\}$, and a 'mixed moment' otherwise, and similarly for central moments and cumulants. The function $g(s)$ in equation (80) is the *cumulant generating function* and is defined as

$$g(s_1, \ldots, s_N) = \log\langle \exp(s_1 n_1 + \ldots + s_N n_N) \rangle. \tag{81}$$

We note that all three types of moments are respectively invariant under permutations of their indices. Therefore, only one representative combination of each permutation class has to be considered. Taking this symmetry into account significantly reduces the number of variables and moment equations.

Some popular moment closure methods can be defined as

- 'Normal moment closure' [57, 121–123] (also called 'cumulant neglect moment closure' in the literature): all cumulants above order M are set to zero, i.e.

$$c_{i_1,\ldots,i_k} = 0, \quad \text{for} \quad k > M. \tag{82}$$

  Note that the cumulants of order higher than two are zero for a normal distribution, hence the name 'normal moment closure'. We refer to the normal moment closure approximations to second and third order as '2MA' and '3MA', respectively.

- 'Poisson moment closure' [124]: the cumulants of a one-dimensional Poisson distribution are all equal to the mean value. We assume here a multivariate distribution to be a product of uni-variate Poisson distributions. Accordingly, for the Poisson moment closure approximation of order M we set all diagonal cumulants to the corresponding mean and all mixed cumulants to zero, i.e.

$$c_{i_1,\ldots,i_k} = y_i, \quad \text{for} \quad k > M \quad \text{and} \quad i_1, \ldots, i_k = i, \quad \text{for some} \quad i \in \{1, \ldots, N\}, \tag{83}$$

$$c_{i_1,\ldots,i_k} = 0, \quad \text{for} \quad k > M \quad \text{and} \quad i_m \neq i_n \quad \text{for some} \quad m, n \in \{1, \ldots, K\}. \tag{84}$$

- 'Log-normal moment closure' [125]: let $\mathbf{m}$ and $\mathbf{S}$ be the mean vector and covariance matrix of a multi-dimensional normal random variable. Then the logarithm of this random variable has a multivariate log-normal distribution whose moments can be expressed in terms of $\mathbf{m}$ and $\mathbf{S}$ as [126]

$$y_{i_1,\ldots,i_k} = \exp\left(\mathbf{v}^T\mathbf{m} + \frac{1}{2}\mathbf{v}^T\mathbf{S}\mathbf{v}\right), \quad \text{for} \quad k > M, \tag{85}$$

where $\mathbf{v} = (g_1, \ldots, g_N)$, and g_m is the number of i_j's having the value m. This allows one to express $\mathbf{m}$ and $\mathbf{S}$ in terms of the first two moments y_i and $y_{i,j}$ which then in turn allows one to express higher-order moments in terms of y_i and $y_{i,j}$, too.

- 'Central-moment-neglect moment closure' (also called 'low dispersion moment closure' in the literature) [127]: all central moments above order M are set to zero:

$$z_{i_1,\ldots,i_k} = 0, \quad \text{for} \quad k > M. \tag{86}$$

- 'Derivative matching': the idea of this method is to express moments above order M by lower order moments in such a way that the time derivatives of the moments of the closed system approximate the time derivatives of the moments up to order M of the exact system at some initial time point. In [128] a method is derived that allows one to produce the corresponding expressions.

Each of the closure methods allows one to express all raw moments above a specified order M in terms of lower order moments and hence to close the moment equations. Note that third order central moments and third order cumulants are identical. Therefore, whenever the moment equation of up to order two do not depend on moments of order higher than three, the normal and central-moment-neglect moment closure are identical. Similarly, the normal and Poisson moment closure approximations can be equivalent for certain systems.

Example. As an example, consider again the gene system in figure 3. The moment equations up to order two for this system are given in equations (21)–(25). To close these we need to express the third order moments $y_{1,1,2}$ and $y_{1,2,2}$ in terms of lower order moments. We do so by means of the normal moment closure defined in equation (82) to second order, i.e. we set the third order cumulants $c_{1,1,2}$ and $c_{1,2,2}$ to zero:

$$c_{1,1,2} = y_{1,1,2} - 2y_1 y_{1,2} - y_2 y_{1,1} + 2y_2 y_1^2, \tag{87}$$

$$c_{1,2,2} = y_{1,2,2} - 2y_2 y_{1,2} - y_1 y_{2,2} + 2y_1 y_2^2. \tag{88}$$

Setting these to zero and rearranging gives

$$y_{1,1,2} = 2y_1 y_{1,2} + y_2 y_{1,1} - 2y_2 y_1^2, \tag{89}$$

$$y_{1,2,2} = 2y_2 y_{1,2} + y_1 y_{2,2} - 2y_1 y_2^2. \tag{90}$$

Using these expressions in equations (21)–(25) the equations decouple from higher order moments. We give here the resulting equations in terms of central rather than raw moments (see equation (79))

$$\partial_t z_1 = -\frac{k_2}{\Omega}(z_{1,2} + z_1 z_2) + k_3(1 - z_1), \tag{91}$$

$$\partial_t z_2 = k_1 z_1 - \frac{k_2}{\Omega}(z_{1,2} - z_1 z_2) + k_3(1 - z_1) - k_4 z_2, \tag{92}$$

$$\partial_t z_{1,1} = \frac{k_2}{\Omega}(-2z_2 z_{1,1} - 2z_1 z_{1,2} + z_{1,2} + z_1 z_2) + y_{1,2}) + k_3(-2z_{1,1} - z_1 + 1), \tag{93}$$

$$\partial_t z_{1,2} = k_1 z_{1,1} + \frac{k_2}{\Omega}(-z_2 z_{1,1} - z_1 z_{1,2} - z_2 z_{1,2} - z_1 z_{2,2} + + z_{1,2} + z_1 z_2)$$
$$+ k_3(z_{1,2} - z_{1,1} - z_1 + 1) - k_4 z_{1,2}, \tag{94}$$

$$\partial_t z_{2,2} = k_1(2z_{1,2} + z_1) + \frac{k_2}{\Omega}(-2z_1 z_{2,2} - 2z_2 z_{1,2} + z_{1,2} + z_1 z_2)$$
$$+ k_3(-2z_{1,2} - z_1 + 1) + k_4(-2z_{2,2} + z_2). \tag{95}$$

Note that the equations do not depend on third or higher order moments and hence are closed and can be integrated.

4.3.2. Properties and recent developments. Moment closures are a popular class of approximations of the moments of the CME with many useful properties. First of all, they are easy to derive and implement. Moreover, since they amount to solving a finite set of ordinary differential equations no ensemble averaging is needed, which means they are computationally significantly more efficient than stochastic simulations of the CME or the CLE, and comparable to the system size expansion. However, this computational efficiency comes at a cost: moment closure methods only give approximate moments and not an approximation of the process or distributions. One advantage of moment closures over the system size expansion is that they can be applied to deterministically multistable systems. However, care must be taken for such system since moment closures can lead to unphysical results, see below.

In contrast to the system size expansion, moment closures are not an expansion in any small parameter, and one can therefore generally not expect that increasing the closure order leads to a higher accuracy. Numerical case studies suggest that this is often indeed the case, however [129–131]. One explanation for this was given in [129] where it has been shown that increasing orders of the normal moment closure agree with increasing orders of the system size expansion in the limit of large volumes. These results can to some extent be generalised to other moment closure methods. For monostable systems, one may therefore expect higher order closures to become more accurate for large enough system sizes. For small system sizes on the other hand, the closures can generally not be expected to converge.

Moreover, since moment closure approximations are an ad hoc approximation, it is not even clear if they always give physically meaningful results. As it turns out, this is not always the case, but they sometimes give rise to unphysical behaviour, such as negative mean values, negative variances or negative higher central moments, as well as diverging trajectories [124, 131]. In numerical case studies, it has been found for several non-linear reaction systems that the normal, log-normal, Poisson and central-moment neglect moment closure show such unphysical behaviour for system volumes *below a certain critical volume* [131, 132]. An example of such unphysical behaviour is shown in figure 6. This may be expected, since smaller volumes correspond to stronger non-linearity and hence stronger fluctuations. However, it has been found for deterministically oscillatory and deterministically multistable systems, that the moment closure methods give rise to non-physical oscillations and non-physical
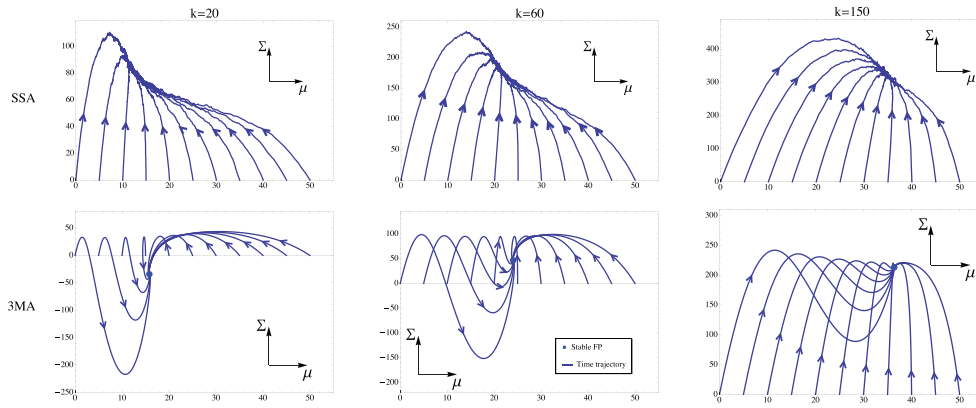
Figure 6. Time trajectories of moment closure with negative variance. The figures show time trajectories in the mean-variance ($\mu - \Sigma$) plane of a protein P in a system with bursty protein production. The reactions are $\varnothing \longrightarrow mP$, $\quad m \in \mathbb{N} \setminus \{0\}$, $\quad P + P \longrightarrow \varnothing$, where the burst size m is a geometrically distributed random number. The top panel shows results obtained from exact stochastic simulations (SSA) and the lower panel results obtained using the normal moment closure to third order (3MA). The parameter k corresponds to the system size. We find that the results from stochastic simulations converge to a positive fixed point for all system sizes and all initial conditions, and that the mean and variance remain non-negative for all times. The 3MA, in contrast, converges to a fixed point with negative variance for the smallest system size $k = 20$. For $k = 60$, the fixed point is now positive, but some initial conditions still lead to trajectories that have a negative variance for some time and are hence physically not meaningful. Only for the largest system size $k = 150$ do all trajectories lead to non-negative trajectories. Reprinted from [131], with the permission of AIP Publishing.

multistability, respectively, for system volumes *above a certain critical volume* [131, 132]. In [132], the normal moment closure was found to give physically meaningful results for larger ranges of system volumes than the other three methods. In terms of accuracy, the four methods performed similar to each other. We emphasise that these were numerical case studies, and it remains open to what extent these results can be generalised to arbitrary reaction systems. In [133] and [134], for example, the log-normal moment closure was found to be significantly more accurate than the normal and central-moment-neglect moment closure for several reaction systems.

4.4. Construction of distributions from moments

The system size expansion and moment closure approximations generally only provide approximations of the moments of a process. Suppose we are however interested in approximating distributions of a process. This can be achieved by running Monte Carlo simulations using the SSA or simulations of the chemical Langevin equation. These methods are computationally quite expensive, however. An alternative approach is to first compute approximations of the moments up to a certain order by means of the system size expansion or moment closure methods, and subsequently constructing a distribution from these moments. If the second step can be done efficiently, this can provide a significantly more efficient method than stochastic simulations since one avoids ensemble averaging. One popular method to construct distributions from moments relies on the *maximum entropy principle*.
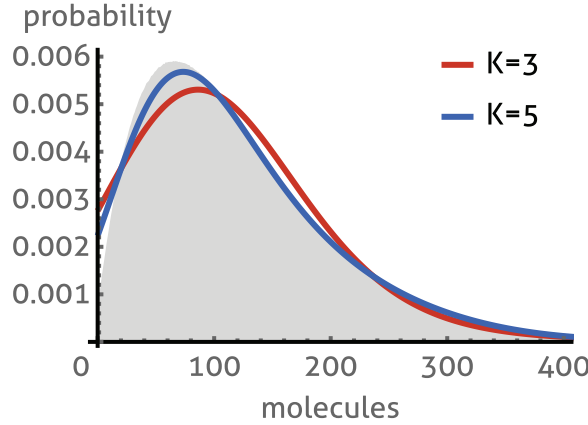
Figure 7. Construction of distribution via maximum entropy. The figure shows the marginal steady-state distribution of a protein in a chemical reaction system with bursty protein production. The grey area shows the exact result computed using the SSA, while the coloured lines show the results obtained using the central-moment-neglect moment closure to order $K = 3$ (red) and $K = 5$ (blue) in combination with maximum entropy. We observe that the approximations accurately capture the skewed distribution, with increasing accuracy at higher order. The figure is taken from [137].

For simplicity, assume we have a one-dimensional problem with discrete-valued variable n and that we have approximate values for the first K moments $\mu_1, \ldots, \mu_K$, for example obtained by means of the system size expansion. The goal is to construct a distribution $P(n)$ matching these moments. The entropy H of a distribution $P(n)$ is defined as [135]

$$H(P) = -\sum_n P(n) \log(P(n)). \tag{96}$$

The maximum entropy method aims at finding a distribution $P(n)$ in a certain family of distributions that maximises the entropy. In our case, the family of distributions is given by the constraint that the first K moments of the distribution should match $\mu_1, \ldots, \mu_K$. This is a non-linear constrained optimisation problem, which can be solved by means of Lagrange multipliers as follows. We define the Lagrangian

$$\mathcal{L} = H(P) - \sum_{i=0}^{K} \lambda_i \left(\sum_n n^i P(n) - \mu_i \right). \tag{97}$$

Optimising $\mathcal{L}$ with respect to P and the λ_i corresponds to solving the constrained optimisation problem. The variation of $\mathcal{L}$ with respect to $P(n)$ reads

$$\frac{\delta \mathcal{L}}{\delta P(n)} = -\log(P(n)) - 1 - \sum_{i=0}^{K} \lambda_i n^i. \tag{98}$$

Setting this to zero we obtain

$$P(n) = \frac{1}{Z} \exp(-\sum_{i=1}^{K} \lambda_i n^i),$$

$$Z = \exp(1 + \lambda_0) = \sum_n \exp(-\sum_{i=1}^{K} \lambda_i n^i), \tag{99}$$

with normalisation constant Z. Inserting equation (99) into equation (97) one obtains an unconstrained optimisation problem which can be efficiently solved using standard numerical methods. The maximum entropy method for the construction of the marginal distributions of single species has recently been used in combination with moment closure methods and the system size expansion in [136, 137]. Figure 7 shows the result for one example system.

A related approach has been developed in [138], where the maximum entropy principle has been used directly to close the moment equations. For time-dependent approximations this method is computationally expensive because the moment equations have to be solved iteratively in small time steps, and in each time step a multi-variate optimisation problem has to be solved. For steady-state approximations, however, the method was found to be efficient and accurate.

4.5. Software

Several software packages for using the discussed approximation methods, as well as for deterministic rate equations and stochastic simulation algorithms, are freely available. For exact stochastic simulations, for example, available packages include the Java-based program Dizzy [139], stand-alone COPASI [140], stand-alone StochKit [141] and the python package StochPy [142]. The system size expansion for orders beyond the LNA is implemented in the stand-alone package iNA [143], which also allows one to perform exact stochastic simulations. Various moment closure approximations are available in the Matlab toolbox StochDynTools [144], and the normal moment closure in Python package MomentClosure [145]. These packages require programming knowledge and are only applicable to mass-action propensity functions. The mathematica package MOCA [132] extends the applicability to non-polynomial and time-dependent propensity functions and does not require any programming skills. Similarly, the python package MEANS [146] extends moment closure methods to non-mass-action propensity functions. The Matlab package CERENA [147] implements several of the mentioned methods, including exact stochastic simulations, the system size expansion and moment closure methods. For a detailed overview of available software packages and their capabilities see figure 1 in [147].

4.6. Other approximations

One reason why the CLE, the system size expansion and moment closure approximations discussed above are so popular is that they are easy to implement, do not require any pre-knowledge of the system (except monostability for the system size expansion), are generally efficient computationally, and often give accurate approximations. The three methods have been frequently applied successfully in the literature. However, there are many scenarios where the three methods give quite inaccurate results. In particular if one or several of the species occur in very low copy numbers, the three methods often perform poorly. A large number of other approximation methods have been developed in the literature, and we give an overview of these methods here. Many of these methods are more sophisticated but only apply to certain classes of systems and require pre-knowledge and/or fine-tuning. As we shall see, the CLE, the system size expansion or moment closure approximations form building blocks of many of these methods.

4.6.1. State space truncation. In section 3.2 we saw that the CME can be solved exactly by matrix exponentiation whenever the state space is finite (see equation (32)). However, for many chemical systems of interest the state space is infinite and the solution in equation (32)

can not be computed. The idea of the *finite state projection algorithm* is to truncate the state space to a finite subspace and use matrix exponentiation to obtain an approximation of the distribution on this subspace [148]. It also provides an estimate of the error, which can be systematically reduced by increasing the truncated space.

While this is an efficient method in some cases, often a large truncated space has to be chosen to achieve a reasonable accuracy, making matrix exponentiation intractable, despite recent progress on numerical algorithms, see [16, 66] for overviews. Similarly, sometimes the state space of a system is finite but too large to compute the matrix exponential. Accordingly, several modified versions of the finite state projection algorithm have been developed, see [149] for a review.

4.6.2. Tau-leaping. *Tau-leaping methods* are approximate ways of simulating a chemical network with the goal of being more efficient than exact stochastic simulations [150]. The basic idea is to 'leap' along time in certain steps τ during which several reaction events occur, thereby avoiding the simulation of each single reaction event as exact stochastic simulations do. The time step has to be chosen small enough such that the propensity functions do not change significantly during one time step. In that case different reactions become independent of each other, and the number of reaction events $N_r(\tau; \mathbf{n}, t)$ of the rth reaction during the time interval τ given that the state at time t is $\mathbf{n}$ becomes a Poisson random variable $\mathcal{P}(f_r(\mathbf{n})\tau)$ with mean $f_r(\mathbf{n})\tau$. Accordingly, the tau-leaping algorithm updates the state $\mathbf{n}(t)$ of the system iteratively as

$$\mathbf{n}(t + \tau) = \mathbf{n}(t) + \sum_{r=1}^{R} \mathbf{S}_r \mathcal{P}(f_r(\mathbf{n})\tau), \qquad (100)$$

where $\mathbf{S}_r$ is the rth row of the stiochiometric matrix. If many reactions are happening per time step the algorithm is more efficient than exact simulations. Increasing τ increases the algorithm's efficiency but obviously decreases its accuracy. If a system contains species with very low particle numbers τ may have to be chosen smaller than the average inter-reaction time to ensure approximately constant propensity functions, which would lead to the algorithm becoming less efficient than exact simulations. Roughly speaking, tau-leaping therefore works best for systems with not too small average molecule numbers.

Despite its simplicity, implementing the tau-leaping method bears several difficulties. Most importantly, the trade-off between accuracy and efficiency makes the choice of the step size τ non-trivial. A bad choice can lead to highly inefficient or inaccurate results and also to other problems, such as negative molecule numbers. Recent years have therefore seen a wide range of studies developing modified versions of the tau-leaping method that aim at solving these problems, including implicit methods [151, 152], binomial methods [153–156], multinomial methods [157], K/R-leaping [156, 158] and a post-hoc correcting method [159]. See [8, 9, 160] for overviews of these methods.

4.6.3. Time-scale separation. Many biochemical systems involve processes with highly varying time scales. Given a certain separation of time scales in a system, it is often possible to derive a reduced model of the system which either allows more efficient simulations or more accurate approximations. Such methods have been well-known for deterministic ordinary differential equation models of biochemical kinetics for several decades [161] but have only more recently been developed for stochastic methods [162]. We start here by describing the deterministic setting.

The deterministic case. In the following we describe two popular methods for reducing deterministic rate equations under time-scale separation. The first method assumes that the chemical reactions in a given system can be divided into 'slow' and 'fast' *reactions*, i.e. reactions

that happen very infrequently or very frequently on a certain time scale of interest. In this case it is sometimes possible to derive a reduced model by eliminating the fast reactions, typically by assuming that certain reactions balance each other. Such methods are often called *quasi-equilibrium approximations* (QEA) [163].

The second method separates a system into 'slow' and 'fast' *species*, rather than reactions. The fast species are assumed to be asymptotically in steady state on the time scale of the slow species. This is known as the *quasi-steady-state approximation* (QSSA). The idea is to eliminate the fast species from the system and to include their steady-state effect on the slow species.

Example. Let us illustrate the QEA and QSSA in the deterministic setting by means of a simple example, the well-studied Michaelis–Menten system:

$$S + E \; \underset{k_2}{\overset{k_1}{\rightleftharpoons}} \; C \overset{k_3}{\longrightarrow} E + P. \tag{101}$$

A substrate S reversibly binds to an enzyme E to form the substrate-enzyme complex C, from which a product molecule P becomes catalysed. The corresponding rate equations are

$$
\begin{aligned}
\partial_t \phi_S &= -k_1 \phi_S \phi_E + k_2 \phi_C, \\
\partial_t \phi_E &= -k_1 \phi_S \phi_E + (k_2 + k_3) \phi_C, \\
\partial_t \phi_C &= k_1 \phi_S \phi_E - (k_2 + k_3) \phi_C, \\
\partial_t \phi_P &= k_3 \phi_C,
\end{aligned}
\tag{102}
$$

where ϕ_S, ϕ_E, ϕ_C and ϕ_P denote the concentrations of S, E, C and P, respectively. Let us first make the assumption that the binding and unbinding of the substrate and the enzyme in equation (101) are fast reactions and that the catalysis reaction is slow. This is the case whenever $k_3 \ll k_2$. On the time scale of the slow reaction, this means that the two fast reactions balance each other, i.e.

$$k_1 \phi_S \phi_E \approx k_2 \phi_C. \tag{103}$$

This corresponds to the *quasi-equilibrium approximations* (QEA). Note that the total concentration E^0 of enzyme molecules is conserved, i.e. $\phi_E + \phi_C = E^0$. Using this to eliminate ϕ_E in equation (103) one obtains $\phi_C = (\phi_S E^0)/(\phi_S + K_M)$, where we defined the *Michaelis–Menten constant* $K_M = k_2/k_1$. Let $v = \partial_t \phi_P = k_3 \phi_C$ be the production rate of the product P. According to equation (102) this becomes

$$v = \frac{k_3 \phi_S E^0}{\phi_S + K_M}, \tag{104}$$

which is the well-known Michaelis–Menten equation [164].

Next, we illustrate the *quasi-steady-state approximation* (QSSA). To this end, we assume that the complex is approximately in steady state:

$$\partial_t \phi_C \approx 0. \tag{105}$$

It can be shown that this is a reasonable approximation whenever $\phi_S + K'_M \gg \phi_E$ with $K'_M = (k_2 + k_3)/k_1$ [165]. This corresponds to the complex C and free enzyme E being fast species, and the substrate S being a slow species.

Using equation (105) in equation (102) and the conservation law $\phi_E + \phi_C = E^0$ we obtain
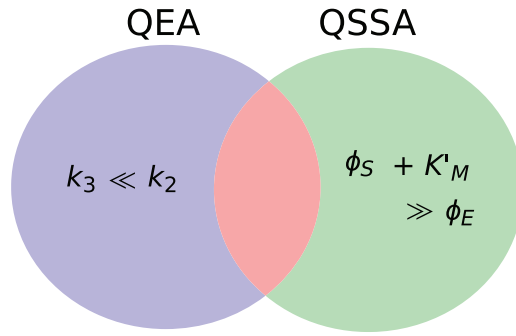
Figure 8. Visualisation of different time-scale separation regimes. The figure visualises the two different time-scale separation regimes for the Michaelis–Menten reaction system with reactions in equation (101). The blue area represents the regime of parameter space where only the quasi-equilibrium assumption holds, the green area where only the quasi-steady-state assumption holds, and the red area where both assumptions hold.

$$v = \frac{k_3 \phi_S E^0}{\phi_S + K'_M}. \tag{106}$$

This is also called *Briggs–Haldane kinetics* [166]. Note that equation (106) has the same form as equation (104) but with a different constant K'_M, showing that the QEA and QSSA are generally not equivalent. Depending on the parameters, only one of the two conditions (or none) may be satisfied. This is illustrated in figure 8.

The stochastic case. For stochastic systems described by the CME, reductions based on time-scale separations are not as straightforward as in the deterministic case. Ideally, one would like to derive a reduced CME and/or a stochastic simulation algorithm for a reduced system. As it turns out, under the quasi-equilibrium assumption of fast and slow reactions it is indeed possible to derive a reduced CME [167]. In contrast, the quasi-steady-state assumption of fast and slow species does not necessarily allow to derive a reduced master equation since it leads to a non-Markovian process for the fast species [168]. One can use a reduced master equation with non mass-action propensity functions that corresponds to the reduced deterministic system under the quasi-steady-state assumption. This heuristic master equation has been found to be accurate for some examples [169–171]. However, even if the quasi-steady-state assumption is valid for the deterministic system, the reduced master equation can be highly inaccurate [172, 173]. The validity of a reduction in the deterministic case does not generally imply a valid reduction in the stochastic case. The relationship between the two descriptions with respect to time-scale separation approximations has for example been studied in [173]. In [174] a reduced description of stochastic models under the quasi-steady-state assumption was derived by means of the linear noise approximation.

It is however possible to derive reduced descriptions based on quasi-steady-state assumptions for a stochastic system if one requires stronger conditions on a system than the conditions needed to reduce the deterministic system. In [175], for example, reactions are split into fast and slow reactions, and slow species are defined as those involved in slow reactions only and fast species as those participating in at least one fast reaction and any number of slow reactions. As mentioned before, the fast variables conditioned on the slow ones are not

Markovian. The authors in [175] solve this problem by approximating the fast process by a virtual fast process that is affected by fast reactions only and therefore Markovian.

A large number of other approximation methods for CME type systems based on time-scale separations have been developed in recent years, see for example [167, 176–185]. Most of these methods split up the full CME into a CME for the fast variables conditioned on the slow reactions/species, and a CME for the slow reactions/variables with marginalised fast variables. On the time scale of the slow dynamics, the conditional CME of the fast dynamics is assumed to quickly reach steady state. Therefore, the CME of the slow dynamics only depends on the steady-state distribution of the fast dynamics. Most methods rely on SSA simulations of the slow dynamics and assume that the fast dynamics is in steady state [175–179, 183].

For some systems, the reduced CMEs under time-scale separation have been solved exactly, for example for gene expression [186–188]. In [189] a perturbative expansion based on time-scale separation for a two-stage gene expression model is derived that allows one to systematically include higher order corrections to the solution.

The studies mentioned so far mainly rely on stochastic simulations of the reduced equations or on analytic solutions for special cases. An alternative approach is to combine time-scale separation reductions with other approximations, such as approximating the fast variables by the chemical Langevin equation [183, 190, 191], deterministic rate equations [190], or moment closure methods [180]. Other methods combine time-scale separation with the finite state projection algorithm [192] (see section 4.6.1), tau-leaping [193], or the linear noise approximation [174, 194, 195]. An adiabatic approximation derived form a stochastic path integral description has been developed in [196]. Other methods derive reduced models for systems that allow a partition in more than two typical time scales, see for example [179, 197]. In [198] a conditional linear noise approximation for gene expression systems with finite, slow promoter states has been derived.

4.6.4. Hybrid methods. In many biochemical reaction networks of practical interest no time-scale separation assumptions apply. However, often some species occur in low and others in high copy numbers, which motivates the combination of different simulation and/or approximation methods for these two groups of species, similar to the time-scale separation case. One possibility would be to partition species into a discretely modelled group and a continuously modelled group, and accordingly reactions into discrete reactions involving discrete species and continuous reactions that do not involve discrete species. In this case species that are involved in both continuous and discrete reactions describe diffusion-jump processes. Such methods are broadly referred to as *hybrid methods*. Note that many of the methods discussed in section 4.6.3 fall under this definition. We give here an overview of hybrid methods that do not rely on separation of time scales, although a clear distinction between the two cases is often not possible.

Consider the case where we split the species of a system into a low abundance group which we model discretely using the SSA and a high abundance group which we model continuously using deterministic rate equations or chemical Langevin equations. Accordingly, we group reactions into discrete reactions involving discrete species and continuous reactions that do not involve discrete species. Between two discrete reaction events, the continuous variables follow rate equations or chemical Langevin equations which can be solved numerically or simulated in a standard way. The propensity functions of the discrete reactions, however, may depend on the continuous variables and hence depend on time in such a hybrid approach. This is akin to the case of extrinsic noise discussed in section 3.3.1, where a discrete system's propensities where assumed to depend on some external stochastic process. Correspondingly, the discrete system of our hybrid system can not be simulated by a standard SSA. As in the extrinsic noise case, one possibility is to numerically integrate the time-dependent propensity functions over time until a target value is

reached and the corresponding reaction occurs. The discrete system is then updated accordingly which may also change the propensity functions of the continuous reactions.

A key challenge of such an approach is to decide which species should be modelled discretely and which continuously, in particular since some species may fluctuate strongly during a simulation. More sophisticated algorithms therefore partition species and reactions adaptively during simulation. Many different methods addressing these and other issues have been developed in the literature, see for example [199–211]. They differ mainly in how the partitioning is conducted and in the simulation methods for the resulting reduced system. Simulation-based approaches include combination of the SSA for the low abundance species with tau-leaping [204], chemical Langevin equations [202, 206, 208, 209, 212], the chemical Fokker–Planck equation [210] and deterministic rate equation approximations [199–201, 205, 211] for the abundance species. Other approaches split species into more than two sets [203].

Since many of these methods are heuristic, it is not straightforward to assess their performance or to prove their convergence to the exact system in some limit. In some cases, this is possible, however. For example, error bounds for some hybrid methods have been derived [208, 213, 214]. In [215] the convergence of different types of hybrid methods to the exact system have been studied, and in [184] criteria have been developed for the convergence of a CME to discrete-deterministic hybrid approximations. An error analysis of various methods has recently been conducted in [216].

While simulation-based hybrid methods are often orders of magnitude more efficient than the standard SSA, they all still rely on stochastic simulations for the low abundance species and some of them also for the high abundance species. They therefore can still become computationally expensive depending on the studied system. Some methods aim at circumventing expensive simulations by applying additional approximations to the reduced systems. In [217], for example, the dynamics of the large abundance species is formulated in terms of moment equations conditional on the low abundance (discrete) species. For non-linear systems these can be closed by means of moment closure approximations (see section 4.3). This method amounts to the solution of a differential algebraic equation which is difficult to solve. The authors in [217] propose several simulation based algorithms. This method has shown to give accurate approximations to the distributions and moments of some gene systems, but its implementation is non-trivial and requires additional approximations and/or simulations. In [218] a simpler conditional moment closure has been proposed for two-state gene systems. Here, the conditional moment equations are closed by means of the second-order normal closure or derivative matching (see section 4.3). This method amounts to solving a coupled set of ordinary differential equations. In [219] a hybrid method is developed where the abundance species is described by the (non-conditional) rate equations. This is a simplification of the previously mentioned methods that model the abundance species by rate equations conditional on the low abundance species. However, this simplification allows one to derive analytic solutions of the CME for the low abundance species for certain systems [219].

5. Comparison of approximation methods

In the previous section we gave a detailed introduction to three popular approximation methods of the CME: the chemical Langevin equation (CLE, section 4.1), the system size expansion (section 4.2) and moment closure approximations (section 4.3). These methods have been successfully used in many applications in the literature, but there exist only very few studies comparing their accuracy. It thus remains unclear how the different methods compare to each other.

In sections 4.1–4.3 we discussed advantages and disadvantages of the different methods. Here, we perform a numerical case study to enable the reader to understand the differences between the methods. First, we study an enzyme reaction system of the Michaelis–Menten type in section 5.1 which can be viewed as a catalysed degradation of a spontaneously produced protein. We then extend the model in section 5.2 by including transcription of mRNA and translation of the protein from mRNA, which allows for bursts in protein production. Finally in section 5.3 we summarise the results and give an overview of advantages and disadvantages of the different methods.

Implementation. Implementation details for the different methods are:

- **CLE**: we simulate the CLE using the Euler–Maruyama algorithm introduced in section 4.1.1. As pointed out in section 4.1, the CLE is traditionally defined for real variables, but suffers from the occurrences of square roots of negative expressions for which it is not well-defined. We therefore implement two versions of the CLE here: (i) a real-valued implementation which keeps the variables positive by rejecting Euler–Maruyama steps that lead to negative variables. We term this version 'CLE-R'; (ii) the recently proposed complex CLE which is defined for complex variables and has been shown to give real-valued moments in the ensemble average [87]. We term this version 'CLE-C'.
  The Euler–Maruyama algorithm requires a step size dt for time discretisation. For the simulation of steady-state moments or distributions, we take a total number of M samples from a single long trajectory at time steps separated by Δt. The chosen values for these parameters will be given for each of the examples.
- **System size expansion and stochastic simulation algorithm**: we use the stand-alone software package iNA [143] for both the system size expansion and the stochastic simulation algorithm (section 3.3). We study the zeroth order system size expansion, i.e. the linear noise approximation (termed 'LNA', section 4.2.2), as well as the first order corrections to the mean and variance (both termed 'SSE-1') given in equations (76) and (77), respectively.
- **Moment closure approximations**: we study here the second-order normal moment closure (termed '2MA') introduced in section 4.3, which is probably the most commonly used one in the literature. For its implementation we use the Mathematica software package MOCA [132].

5.1. Enzymatic protein degradation

Let us consider the well-known Michaelis–Menten system with reactions

$$\varnothing \xrightarrow{k_0} S, \quad S + E \underset{k_2}{\overset{k_1}{\rightleftharpoons}} C \xrightarrow{k_3} E + P. \tag{107}$$
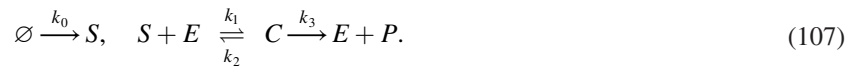
We studied this system without the first reaction in section 4.6.3 in the context of time-scale separations in a deterministic setting. A substrate molecule S is created spontaneously and binds reversibly to a free enzyme molecule E to form a complex C, which catalyses the substrate into a product molecule P. We consider S to be a protein in the following.

Here, we are interested in the accuracy of the different approximation methods as compared to exact stochastic simulations of the corresponding stochastic system. Note that the total number of enzymes E^0 is conserved. The system has a steady state in the protein numbers if and only if $\alpha \equiv k_0 \Omega / (E^0 k_3) < 1$, which means that the input rate must be smaller than the maximum turnover rate. The parameter α can hence be viewed as a saturation factor.
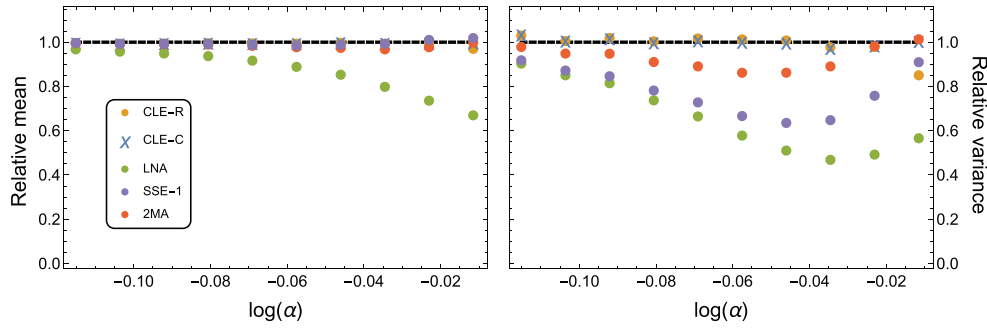
Figure 9. Steady-state mean and variance of protein for enzyme system in equation (107). The figures show the steady-state mean (left) and variance (right) of the protein S of the enzyme system in equation (107), as a function of the saturation parameter $\alpha = k_0\,\Omega/(E^0 k_3)$ on logarithmic scale. The values obtained by the approximation methods are normalised by the result obtained from stochastic simulations using the SSA, which means that the dashed black line corresponds to the exact result (up to sampling error). The parameters used are $E^0 = 2500, k_1 = 40, k_2 = 8, k_3 = 60, \Omega = 1$ and we vary k_0 according to the shown values of α. The simulation parameters for the two CLE implementations are $dt = 10^{-5}, \Delta t = 1$ and $M = 10^5$.

Figure 9 shows the relative mean and variance of the protein S as a function of α for a system with a large number of total enzymes, $E^0 = 2500$. We find that both the real and complex CLE implementations (CLE-R and CLE-C), the first order system size expansion (SSE-1) and the second-order normal moment closure (2MA) give good approximations for the mean value. Only the LNA, which corresponds to the deterministic rate equations for mean values, shows significant deviations from the exact result. For the variance we observe larger deviations for all methods, with the LNA again being the least accurate. The two CLE implementations show the best performance, being more accurate than the 2MA, which in turn is more accurate than the SSE-1. Note that the CLE-R and the CLE-C give very similar results, meaning that the inclusion of a rejecting boundary has a negligible effect in this case.

Overall, we find that the approximation methods give rise to larger errors for larger values of α, which can be explained as follows. As mentioned before, the system only possesses a steady state for $\alpha < 1$ and becomes unstable for $\alpha > 1$. We therefore expect larger fluctuations for values of α close to unity. Moreover, since most of the enzymes are in the complex state C in this limit, we expect a skewed distribution for the substrate. In the other limit, $\alpha \to 0$, most enzymes are in the free state E, which reduces the non-linear effect of the bimolecular reaction $S + E \to C$. We therefore expect the approximation methods to perform well in this limit, and to lead to larger errors for $\alpha \to 1$. This is exactly what we observe in figure 9.

Next, we study the steady-state probability distribution of the protein as predicted by the SSA, CLE-R, CLE-C and the LNA. Note that the CLE-C gives complex-valued samples. To obtain a distribution in real variables, we take the real parts of these samples. We reduce the total number of enzymes to $E^0 = 60$ to study small molecule number effects. The results are shown in figure 10. We find that the LNA strongly underestimates the true mean value (obtained using the SSA) and does not reproduce the distribution very accurately. However, surprisingly, the real-valued CLE (CLE-R), does even worse than the LNA, both in terms of mean value and distribution. The complex valued CLE (CLE-C) on the other hand, predicts the true mean value with a negligible error, and even captures the highly skewed distribution very well (right panel of figure 10). This demonstrates that a naive fixing of the boundary problem of the CLE (CLE-R) can lead to highly inaccurate results.
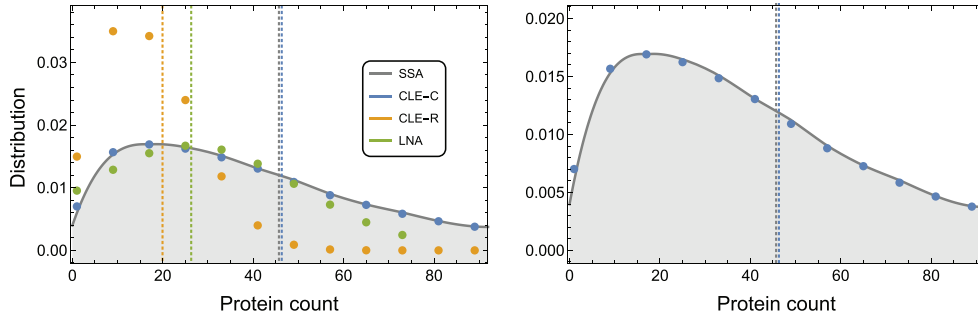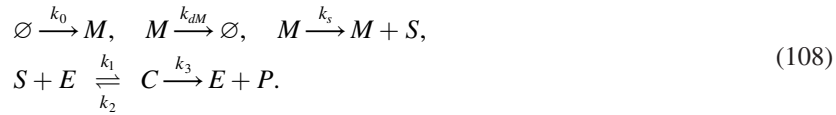
Figure 10. Steady-state distribution of protein for enzyme system in equation (107). Left: distribution obtained using the CLE-R, CLE-C, LNA and SSA. Right: only CLE-C and SSA. The vertical lines indicate the corresponding steady-state mean values. The mean values of the 2MA and SSE-1 are very close to the SSA result and are not shown here. The used parameters are $E^0 = 60, k_0 = 115, k_1 = 3.5, k_2 = 2, k_3 = 2, \Omega = 1$. Simulation parameters for the CLE-R and CLE-C are $dt = 10^{-5}, \Delta t = 1$ and $M = 10^5$.

5.2. Bursty protein production

We now extend the system in equation (107) by considering the protein S to be produced in a gene expression motif, i.e. via transcription of mRNA M followed by translation to protein:

$$\varnothing \xrightarrow{k_0} M, \quad M \xrightarrow{k_{dM}} \varnothing, \quad M \xrightarrow{k_s} M + S,$$
$$S + E \underset{k_2}{\overset{k_1}{\rightleftharpoons}} C \xrightarrow{k_3} E + P. \tag{108}$$

Depending on the parameters k_{dM} and k_s this system can give rise to bursts in the production of protein S, namely whenever each produced mRNA molecule produces on average several proteins during its lifetime. The average number of proteins per mRNA molecule is given by $b = k_s/k_{dM}$ and also called 'burst size' [220]. Large b correspond to large bursts which in turn lead to large fluctuations.

Here, we are interested in the accuracy of the different approximation methods as a function of the burst size b, since we expect larger errors for larger fluctuations. To isolate the effect of the burst size b, we vary k_0 and k_{dM} such that the average production rate of protein $k_0 k_s/k_{dM} = 115$ in steady-state conditions is held constant and equal to the production rate used in the previous section in figure 10.

Figure 11 shows the mean and variance of the protein S as a function of b as predicted by the different approximation methods. Similar to the previous section in figure 9 we find that the LNA performs worse than the other methods for the mean and the variance, with the exception of the real-valued CLE (CLE-R) which performs even worse. Similarly to figure 10, we find here that the complex-valued CLE (CLE-C) performs surprisingly well, demonstrating again the significance of the boundary problem of the CLE. In fact, the CLE-C performs significantly better than all the other methods, including the 2MA and SSE-1. The latter two perform similarly and both significantly better than the LNA. One should keep in mind, however, that the 2MA, SSE-1 and LNA are computationally significantly more efficient than the CLE-C since they do not rely on sampling.

Next, we consider the steady-state distribution of the protein for a large burst size of $b = 7.5$. Figure 12 shows the results predicted by the stochastic simulation algorithm, the two CLE implementations, and the LNA. We find that the CLE-R and LNA give highly
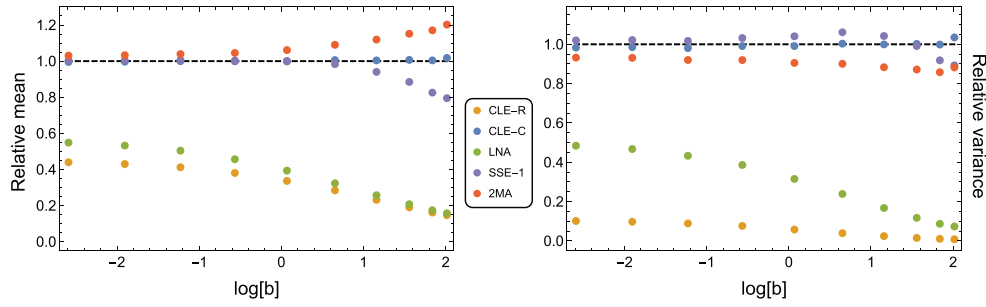
Figure 11. Steady-state mean and variance of protein for extended system in equation (108). The figures show the steady-state mean (left) and variance (right) of the protein S of the enzyme system in (108), as a function of the burst size $b = k_s/k_{dM}$ on logarithmic scale. The values obtained by the approximation methods are normalised by the result obtained from stochastic simulations (SSA), which means that the dashed black line corresponds to the exact result (up to simulation error). The used parameters are $E^0 = 60, k_s = 1.5, k_1 = 3.5, k_2 = 2, k_3 = 2, \Omega = 1$. k_0 and k_{dM} are varied to vary the burst size $b = k_s/k_{dM}$ such that the average protein production rate $k_0 k_s/k_{dM} = 115$ is held constant. Simulation parameters for the CLE-R and CLE-C are $dt = 10^{-4}, \Delta t = 10$ and $M = 10^4$.

inaccurate results, both for the mean value and the distribution, with even larger deviations than for the simpler system in the previous section in figure 10. This is to be expected, since the system now leads to large bursts in protein production, while the average production rate of protein and the rate constants of the enzymatic reactions are the same as in figure 10. Similar to the previous section, we find here that the complex-valued CLE gives a highly accurate approximation for the distribution, despite the large skewness of the later. Once again we find that the boundary problem of the CLE can lead to highly inaccurate results if not treated carefully.

5.3. Discussion

5.3.1. Numerical results. In this section we gave a numerical comparison of the CLE, the LNA, the normal moment closure to second order (2MA) as well as the next leading order corrections to mean and variance of the system size expansion beyond the LNA (SSE-1). We implemented two versions of the CLE: a real-valued version with rejecting boundary and a complex-valued version. We considered an enzymatic protein degradation system (section 5.1) and an extension thereof including bursty protein production (section 5.2).

One important observation we made is the large discrepancy between the two CLE implementations: while we found a negligible difference in the case of large total enzyme numbers (figure 9), the complex version was significantly more accurate in the case of smaller total enzyme numbers (figures 10–12). Crucially, in the latter cases, the real-valued CLE performed worse than all the other methods, while the complex-valued CLE performed better than all the other methods. This illustrates the boundary problem of the CLE, and how significant inaccuracies can arise when fixing it in a naive way.

Another important observation is that the LNA was throughout found to be less accurate than all the other methods (except the CLE-R). This result is not very surprising, since the LNA corresponds to the deterministic rate equations on the mean level, and gives the zeroth order fluctuations (in terms of the system size expansion) about the deterministic mean for the variance. The SSE-1 includes the leading order corrections to both the mean and variance.
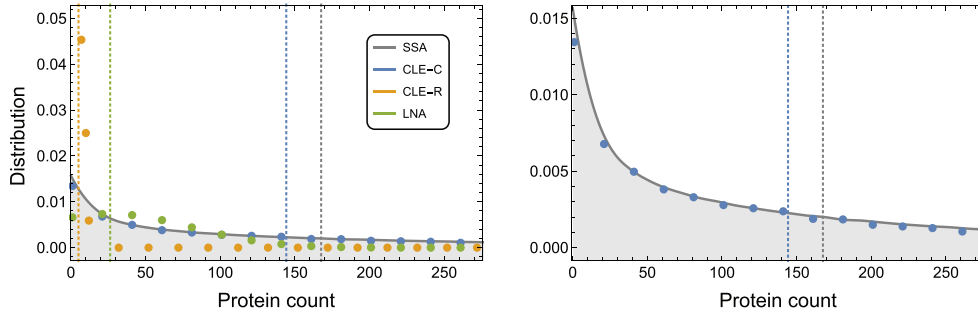
Figure 12. Steady-state distribution of protein for extended system in equation (108). Left: distribution obtained using the CLE-R, CLE-C, LNA and SSA. Right: only CLE-C and SSA. The vertical lines indicate the corresponding steady-state mean values. The used parameters are $E^0 = 60, k_s = 1.5, k_0 = 15.33, k_{dM} = 0.2,$ $k_1 = 3.5, k_2 = 2, k_3 = 2, \Omega = 1,$ corresponding to a burst size of $b = k_s/k_{dM} = 7.5.$ Simulation parameters for the CLE-R and CLE-C are $dt = 10^{-4}, \Delta t = 10$ and $M = 10^4.$

Similarly, the 2MA and CLE capture effects of fluctuations on the mean. It is hence not surprising that these methods perform better than the LNA.

In terms of the mean and variance (figures 9 and 11), we found that the SSE-1 and 2MA performed similar to each other, except for the variance in figure 9 where 2MA was more accurate. Surprisingly however, the CLE-C was found to be significantly more accurate than the 2MA and SSE-1 in most cases.

In terms of distributions (figures 10 and 12) we found the CLE-C to be highly accurate, even though the distributions were highly skewed. The LNA, which predicts a Gaussian distribution, was obviously not able to capture these skewed distributions.

5.3.2. Advantages and disadvantages. The advantage of the CLE over the other methods is that it gives approximations of the process and distributions in contrast to moment closure methods and the system size expansion. Moment closure approximations give only approximations to the first few moments of a process. The system size expansion in principle predicts distributions. However, closed-form solutions for higher orders beyond the LNA have so far only been derived for one-dimensional systems [108]. It is not clear if the same is possible for multi-species systems. The higher orders of the system size expansion have therefore mainly been used to approximate the moments of a process. If one is interested in approximating the whole process or its distributions, the CLE is therefore a useful method.

Suppose now that we are only interested in the moments of a process. In the numerical case study performed before, we found the CLE to be more accurate than the LNA, the SSE-1 and the 2MA. However, the CLE is computationally significantly more expensive than the other methods. While the CLE requires a large number of stochastic simulations to obtain the moments of a process, the other methods only require the numerical solution of a finite set of ODEs and are hence typically orders of magnitude faster. Moreover, if defined for real-valued variables, the CLE suffers from a boundary problem at zero molecule numbers, and real-valued modifications lead to inaccurate results. The boundary problem is solved by extending the CLE to complex-valued variables, which is however less efficient to simulate [87]. Due to these reasons, it seems preferable to use the system size expansion or moment closure approximations if one is only interested in the moments of a process.

Next, the question arises if the system size expansion or moment closure approximations are preferable. While the system size expansion is a systematic expansion in a small parameter,

moment closure approximations are an ad hoc approximation. This fact makes the system size expansion more appealing, since it is guaranteed to be accurate for large system volumes. The same cannot generally be expected to be true for moment closure approximations. On the other hand, the system size expansion has the disadvantage that it is not applicable to systems that are deterministically multi-stable, a limitation not shared by moment closure methods. Moreover, higher order corrections of the system size expansion are significantly harder to derive and implement than higher order moment closure methods. Current software packages implementing the system size expansion only implement two orders beyond the LNA for the mean, and one order beyond the LNA for the covariance [143, 147]. Moment closure approximations, on the other hand, are implemented to various orders [132, 146, 147]. Due to these reasons, it depends on the problem at hand as to decide which method is preferable.

6. Inference

So far, we have focussed on the *forward problem* of approximating marginal distributions of a fully specified process. Such distributions depend naturally on the parametrisation of the process: it is not uncommon for e.g. steady-state distributions to exhibit qualitatively different behaviours depending on the specific value of reaction propensities. In many concrete applications, the model parameters may only be known approximately: direct measurements of kinetic reaction parameters are difficult to obtain, and, even in cases when good estimates are available, *in vivo* parameters of a concrete system embedded in a cell may be influenced by a plethora of additional factors, leading to significant uncertainty. It is therefore of considerable interest to also address the *inverse problem*: using (noisy) observations of a system to constrain the uncertainty over model parameters and/ or predictions. This is a well-studied problem in statistics and machine learning: we give here a brief review of recent developments within the *Bayesian* approach to solving this inverse problem, with particular attention to methodologies which have employed the approximation methods described earlier.

6.1. General problem formulation

The setup we will consider is the following: we consider a stochastic process $p(\mathbf{x}_{0:T}|\theta)$ as a measure over the space of trajectories $\mathbf{x}_{0:T}$ of the system in the time interval $[0, T]$, with $\mathbf{x}_t$ representing the value of the state variable at time t. In the case of a CME system, such trajectories will be piecewise constant functions from $[0,T]$ onto a discrete space, while for a continuous approximation (e.g. CLE or LNA) $\mathbf{x}_t$ will be real-valued. The stochastic process is assumed to depend on a set of kinetic parameters θ, whose *a priori* uncertainty is captured by a prior distribution $p(\theta)$. Additionally, we assume the existence of a measurement process which associates each trajectory $\mathbf{x}_{0:T}$ with an observed random variable $\mathbf{y}$; in the simplest case, the observed variable $\mathbf{y}$ may just be thought of as the state of the system at a particular set of time points, corrupted by random observation noise. We account for such experimental errors through an *observation model* encoded in a probability distribution $p(\mathbf{y}|\mathbf{x}_{0:T})$, i.e. the likelihood of a measurement given the true state of the system, which may depend on additional parameters (here omitted for notational conciseness). We restrict our interest here to the case where the observations are in the form of a time series of state variable observations with independent and identically distributed noise, i.e. measurements of all or of a subgroup of the species in the system. Hence, the observation vector will take the form $\mathbf{y} = (y_0, \ldots, y_T)$ for some discrete time points 0,...,T. More general cases where the observations take other forms, such as continuous-time constraints or penalties over particular areas of the state space, are treated for example in [221, 222].

Also, we consider all observations to come from a single trajectory, or from replicate trajectories with the same (unknown) kinetic parameter values; in other words, we exclude the case where parameters can also be variable, e.g. to accommodate extrinsic noise between cells [223].

The general inference problem is then the problem of computing the joint posterior measure $p(\mathbf{x}_{0:T}, \theta | \mathbf{y})$ over the set of trajectories and parameters of the system. This is formally obtained by applying Bayes' rule

$$p(\mathbf{x}_{0:T}, \theta | \mathbf{y}) \propto p(\mathbf{y} | \mathbf{x}_{0:T}) p(\mathbf{x}_{0:T} | \theta) p(\theta). \tag{109}$$

This joint posterior provides information both about the parameters and about the state of the system at all time points during the specific trajectory for which data was collected. In most systems of interest, computation of the normalisation constant of the posterior distribution is analytically intractable, due to the requirement of performing very large sums/ high dimensional integrals. Much of the research in Bayesian statistics and machine learning is therefore focussed on computing efficient approximations to posterior distributions: such approximations can be analytic, usually obtained by variational minimisation of a divergence functional, or sampling based [224]. This latter class has received particular attention in recent years, and is predicated on constructing a Markov chain which has the required posterior as an invariant distribution. This implies that, asymptotically, the Markov chain will sample from the correct posterior distribution, enabling Monte Carlo computation of any desired statistic (*Markov chain Monte Carlo, MCMC*).

6.2. The forward–backward algorithm

Inference in dynamical systems is based on the fundamental factorisation of the single time posterior marginal

$$p(\mathbf{x}_t | \mathbf{y}_0, \ldots \mathbf{y}_T) \propto p(\mathbf{x}_t | \mathbf{y}_{i \leqslant t}) p(\mathbf{y}_{i > t} | \mathbf{x}_t). \tag{110}$$

This factorisation, which is a simple consequence of the Markovian assumption and the product rule of probability [224], states that the posterior probability at time t is a product of the posterior $p(\mathbf{x}_t | \mathbf{y}_{i \leqslant t})$ based on the data seen so far up to time t (the so called *filtering distribution*) and the likelihood $p(\mathbf{y}_{i > t} | \mathbf{x}_t)$ of future data conditioned on the current state. The factors in equation (110) can be computed iteratively using the celebrated *forward–backward algorithm* [225].

The forward part, which is also referred to as *filtering* [226], works as follows. Assume that we know the *posterior* $p(x_{i-1} | y_{i-1}, \ldots y_0)$ at time step $i - 1$, and that we can solve the system forward in time (i.e. solve the CME) to obtain the transition probability $p(x_i | x_{i-1})$ and hence the *predictive distribution* $p(\mathbf{x}_i | \mathbf{y}_{i-1}, \ldots \mathbf{y}_0) = \int \mathrm{d}\mathbf{x}_{i-1} p(\mathbf{x}_i | \mathbf{x}_{i-1}) p(\mathbf{x}_{i-1} | \mathbf{y}_{i-1}, \ldots \mathbf{y}_0)$. The posterior of time step i is then obtained by taking the measurement $\mathbf{y}_i$ at time point i into account by means of the the *Bayesian measurement update*

$$
\begin{aligned}
p(\mathbf{x}_i | \mathbf{y}_i, \ldots \mathbf{y}_0) &= \frac{p(\mathbf{y}_i | \mathbf{x}_i, \mathbf{y}_{i-1}, \ldots \mathbf{y}_0) p(\mathbf{x}_i | \mathbf{y}_{i-1}, \ldots \mathbf{y}_0)}{p(\mathbf{y}_i | \mathbf{y}_{i-1}, \ldots \mathbf{y}_0)} \\
&= \frac{p(\mathbf{y}_i | \mathbf{x}_i) p(\mathbf{x}_i | \mathbf{y}_{i-1}, \ldots \mathbf{y}_0)}{p(\mathbf{y}_i | \mathbf{y}_{i-1}, \ldots \mathbf{y}_0)},
\end{aligned}
\tag{111}
$$

where we have used the Markov property to obtain the second line. The filtering procedure thus comprises iteratively solving the process between measurements and performing the Bayesian measurement update in equation (111) and yields the filtering distribution $p(\mathbf{x}_t | \mathbf{y}_{i \leqslant t})$.

The second term in equation (110) can be obtained by the Backward algorithm, a recursion procedure similar to filtering (but running backward in time) [225]. Furthermore, a modification of the algorithm, the so called *forward filtering/backward sampling* algorithm [227], can be used to draw sample trajectories from the posterior process, which can then be used in MCMC approaches for joint state and parameter inference.

6.3. Parameter inference

Suppose we are not interested in state inference but only in the parameters of the system. In this case we do not need to compute the posterior marginal in equation (110). Rather, it is sufficient to either run the forward algorithm (filtering) or the backward algorithm, since each of them independently deliver the marginal likelihood $p(\mathbf{y})$. The backward algorithm computes the likelihood $p(\mathbf{y}_{i>t}|\mathbf{x}_t)$ of future data conditioned on the current state, and $p(\mathbf{y})$ is simply the end result of the recursion algorithm at time 0.

To see that the forward algorithm also allows to compute the likelihood $p(\mathbf{y})$, note that due to the Markov property the latter can be written as

$$p(\mathbf{y}) = p(\mathbf{y}_0) \prod_{i=1}^{T} p(\mathbf{y}_i|\mathbf{y}_{i-1}, ..., \mathbf{y}_0). \tag{112}$$

We find the factors on the rhs of equation (112) are just the normalisations factors of the Bayesian updates of the filtering procedure in equation (112).

Optimising $p(\mathbf{y}) = p(\mathbf{y}|\theta)$ with respect to the parameters yields asymptotically consistent parameter estimates, also called *maximum likelihood estimate*. In a Bayesian framework, one would combine the likelihood with a parameter prior $p(\theta)$ to give the posterior over the parameters according to Bayes' law

$$p(\theta|\mathbf{y}) \propto p(\mathbf{y}|\theta)p(\theta), \tag{113}$$

which also allows to quantify uncertainty of the inferred parameter values.

6.4. Computational methods for Bayesian inference in stochastic chemical reaction networks

6.4.1. Methods for general networks. The primary difficulty in applying the forward–backward approach to inference in chemical reaction networks is the requirement for forward integrability of the system dynamics: calculating the transition probabilities $p(\mathbf{x}_i|\mathbf{x}_{i-1})$ requires solving the CME, which is generally not possible analytically. Some approaches resort to numerical integration of the equations, including the variational approach of [228] and the uniformisation sampler of [229]. These approaches can be effective for closed systems with low molecular numbers; however their application to open systems invariably requires an artificial truncation of the state space, introducing a bias which is hard to quantify. Truncations are also used in the recent work of [230]; however, here a random truncation scheme guarantees unbiasedness of the results, as well as leading to substantial computational savings. Other approaches that can handle open systems either introduce additional latent auxiliary variables (such as the number of reactions in the time interval as in [231]), or resort to a sequential Monte-Carlo scheme which relies on multiple simulations from different initial conditions (particle filtering, [232]). Both such schemes incur potentially large computational overheads.

The computationally intensive nature of inference methodologies adopting a microscopic system description constitutes a formidable obstacle to inference in large-scale reaction networks. This has justified a considerable interest in the use of mesoscopic approximations for

inference. One of the earliest attempts [233] relied on the chemical Langevin equation approximation to the CME (section 4.1); this provides a more efficient inference scheme compared to the auxiliary variable approach of [231], however the computational costs remain high due to the need to compute transition probabilities for non-linear diffusion processes. In this light, the LNA provides a more promising avenue, since the sufficient statistics of the (Gaussian) single-time marginals can be efficiently computed by integrating a system of ordinary differential equations. Several authors have therefore proposed inference schemes which integrate the LNA approximation [234–236]. Moment closure approximations provide an alternative approximation scheme with similar computational complexity to the LNA, however they do not generally compute a marginal distribution, rather only a few moments of a generally unknown distribution. Their use for time series inference is therefore limited to second-order normal moment closure schemes, where a Gaussian approximation is taken. This approach has been proposed in [237], where it has been shown to yield accurate results with modest computational overheads. In [238] the second-order normal moment closure has been combined with the chemical Langevin equation and integrated into an expectation-propagation algorithm for intractable likelihoods and continuous-time constraints. Another interesting opportunity for moment-based inference is offered by flow-cytometry data: here, simultaneous measurements of millions of cells enable an empirical characterisation of the marginal moments directly (albeit potentially corrupted by extrinsic noise), which can then be fitted to a moment-approximation of the CME [239].

6.4.2. Inference for gene expression data. The inference methods mentioned above do not assume any knowledge about a given system. While this makes them in principle applicable to any type of reaction network, more efficient and/or accurate methods can often be employed by including *a priori* knowledge in a model. Gene expression systems constitute a particularly important example where this is often the case. Cells typically possess only one or very few copies of a gene. Proteins and mRNA molecules on the other hand often occur at copy numbers that are orders of magnitude larger. Such systems are therefore often suitable for hybrid methods (see, section 4.6.4) that model some of the species as discrete variables (e.g. the genes) and others as continuous variables (e.g. the mRNA and proteins).

While it is often not straightforward to integrate hybrid methods into inference schemes, significant progress has recently been made in this respect. In [240], for instance, the different promoter states of a gene are modelled as a change-point process which drives a linear SDE representing the protein dynamics, and an efficient MCMC inference method is developed. The same model has been integrated into a variational inference method which additionally allows transcriptional feedback in [241, 242]. In [243], a particle MCMC scheme is developed based on a hybrid method that approximates the continuous variables by the linear noise approximation (see section 4.2.2). More recently a different hybrid approach combining several types of approximations has been integrated into an efficient Bayesian inference scheme in [244].

6.5. Summary

This bird eye survey of inference for stochastic chemical reaction networks highlights the diversity of statistical research in the area. Such diversity can appear baffling to the outsider, and a major problem for the greater diffusion of these ideas is the lack of standard software tools. Inference tools for stochastic chemical reaction networks often form a small subsection of software tools for parameter estimation of deterministic methods [245, 246], and there

haven't been systematic comparisons of various inference schemes that investigate the relative merits of the different algorithms on a number of relevant examples. Furthermore, inference approaches inevitably construct an approximation of a posterior distribution; while much statistics research investigates the convergence properties of these approximations, it remains entirely unclear how inference errors combine with approximation errors when inference schemes are deployed on approximate dynamics. Due to these reasons, we have chosen not to include an explicit numerical comparison between inference methods in this tutorial, as we feel this would deserve a separate review on its own.

7. Conclusions

Recent years have seen an explosion of experimental studies revealing the crucial role that stochastic fluctuations in chemical reaction networks play for living cells. Driven by these discoveries a plethora of methods for the mathematical and statistical analysis of such systems has evolved. The goal of this review is to give a self-contained introduction to the field of modelling for stochastic chemical kinetics. Moreover, it introduces key approximation and inference methods for this field and gives an overview of recent developments.

The chemical master equation (CME) constitutes the accepted non-spatial description of stochastic chemical networks. Recent years have seen a burst of analysis and approximation methods based on the CME. We gave here an introduction to the CME modelling framework and discussed stochastic simulation and analytic solution methods. Next, we introduced various approximation methods with particular focus on the chemical Langevin equation, the system size expansion and moment closure methods. We also gave an introduction to time-scale separation based approximations, as well as hybrid methods and reviewed the existing literature. Finally, we gave an introduction to the problem of statistical inference from experimental data in a Bayesian framework, and reviewed existing methods. The presentation is aimed to be a self-contained introduction for scientists from different disciplines.

In a numerical case study we compared the chemical Langevin equation, the zeroth order system size expansion (linear noise approximation, LNA), the first-order corrections of the system size expansion to mean and covariance (SSE-1) and the second-order normal moment closure (2MA) with exact results obtained using stochastic simulations. In terms of moments, we found that a naive real-valued implementation of the CLE (CLE-R) enforcing positive concentrations was less accurate than all the other methods. A complex version of the CLE (CLE-C), in contrast, was found to be the most accurate of all methods. The SSE-1 and 2MA performed similar two each other and significantly better than the LNA. In terms of steady-state distributions, we compared the CLE-R, the CLE-C and the LNA with exact results obtained using the SSA. We found that the CLE-R and the LNA were not able to accurately capture the distributions, the LNA being more accurate than the CLE-R, however. The CLE-C, in contrast, gave accurate approximations even for highly skewed distributions.

The CME is a valid description of systems that are well-mixed and sufficiently dilute, i.e. the diffusion of particles constitutes the fastest time scale of the system and the total volume of all molecules in the model is much smaller than the system volume. While these assumptions are valid in some cases, it turns out that they are not met by many biological systems. Whenever this is the case, models need to be employed that take spatial positions and diffusion of particles into account. The main family of such models goes under the name of *stochastic reaction–diffusion processes* (SRDPs), Markovian models where independent particles diffuse in space and react whenever they come in contact (or sufficiently close). The evolution equation for the marginal probabilities of an SRDP, the spatial analogue of

the CME, is complicated by the fact that the number of particles varies in time, and needs to be defined on an infinite-dimensional Fock space [247, 248]. Solving, or even approximating, such an equation is essentially impossible, and hence SRDPs are mainly analysed in an algorithmic way: each particle is simulated performing Brownian motion in continuous space and chemical reactions between particles happen stochastically under certain rules [249]. This is computationally extremely expensive and significant effort has been spent in improved simulation methods [250, 251]. An alternative modelling framework is given by the *reaction–diffusion master equation* (RDME) which coarse-grains an SRDP by assuming a compartmentalisation of space and locally homogeneous conditions within each compartment [252]. While simulations in this framework are generally more efficient than in the continuous case, they are typically still expensive [253], generally significantly more expensive than in the non-spatial CME case. More importantly, the RDME is not a systematic discretisation of an SRDP, in the sense that its continuum limit generally does not lead to the original SRDP in two or more spatial dimensions [252], because bimolecular reactions become vanishingly infrequent in the continuum limit.

Due to these reasons, analysis methodologies for spatial models are much less developed than in the non-spatial CME case. Some studies investigating spatial stochastic phenomena and comparing the SRDP and RDME approaches include [249, 254–259]. In contrast to the CME case, very few studies have attempted analytical approximations for SRDPs. However, some progress has been made in recent years in this respect. In [260] and [261], for example, the linear noise approximation was extended to spatial systems, and in [262] higher orders of the system size expansion for effective one-species systems. Even fewer studies have addressed inference for SRDPs. Only a handful of studies have approached the issue of inference working directly with the SRDP or RDME framework, using likelihood-free sampling methods [263] or variational approximations [264]. In [265], the linear noise approximation has been integrated into an inference scheme for SRDPs. This method is however limited to certain classes of one-dimensional systems. More recently, in [82] it was shown that SRDPs can be approximated by spatio-temporal point processes, a popular class of models from statistics [266]. This was used to derive an efficent inference algorithm for general SRDPs.

In summary, we have presented an introduction to modelling, approximation and inference methods for stochastic chemical kinetics based on the chemical master equation. We hope that this review will help scientists from other disciplines to dive into this exciting field, and that it will stimulate research in the presented areas.

Acknowledgments

References

[1] Eldar A and Elowitz M B 2010 Functional roles for noise in genetic circuits *Nature* **467** 167–73
[2] Elowitz M B, Levine A J, Siggia E D and Swain P S 2002 Stochastic gene expression in a single cell *Science* **297** 1183–6
[3] Gillespie D T 1992 A rigorous derivation of the chemical master equation *Physica* A **188** 404–25

[4] Gillespie D T 1976 A general method for numerically simulating the stochastic time evolution of coupled chemical reactions *J. Comput. Phys.* **22** 403–34

[5] Gillespie D T 1977 Exact stochastic simulation of coupled chemical reactions *J. Phys. Chem.* **81** 2340–61

[6] Turner T E, Schnell S and Burrage K 2004 Stochastic approaches for modelling *in vivo* reactions *Comput. Biol. Chem.* **28** 165–78

[7] El Samad H, Khammash M, Petzold L and Gillespie D 2005 Stochastic modelling of gene regulatory networks *Int. J. Robust Nonlinear Control* **15** 691–711

[8] Li H, Cao Y, Petzold L R and Gillespie D T 2008 Algorithms and software for stochastic simulation of biochemical reacting systems *Biotechnol. Prog.* **24** 56–61

[9] Pahle J 2009 Biochemical simulations: stochastic, approximate stochastic and hybrid approaches *Briefings Bioinf.* **10** 53–64

[10] Mauch S and Stalzer M 2011 Efficient formulations for exact stochastic simulation of chemical systems *IEEE/ACM Trans. Comput. Biol. Bioinf.* **8** 27–35

[11] Gillespie D T, Hellander A and Petzold L R 2013 Perspective: stochastic algorithms for chemical kinetics *J. Chem. Phys.* **138** 170901

[12] Szekely T and Burrage K 2014 Stochastic simulation in systems biology *Comput. Struct. Biotechnol. J.* **12** 14–25

[13] Nicolis G and Prigogine I 1977 *Self-Organization in Nonequilibrium Systems* vol 191977 (New York: Wiley)

[14] van Kampen N G 1992 *Stochastic Processes in Physics and Chemistry* vol 1 (Amsrerdam: Elsevier)

[15] Gardiner C W 1985 *Handbook of Stochastic Methods* vol 3 (Berlin: Springer)

[16] Goutsias J and Jenkinson G 2013 Markovian dynamics on complex reaction networks *Phys. Rep.* **529** 199–264

[17] Weber M F and Frey E 2016 Master equations and the theory of stochastic path integrals (arXiv:1609.02849)

[18] Blythe R A and McKane A J 2007 Stochastic models of evolution in genetics, ecology and linguistics *J. Stat. Mech.* P07018

[19] Datta S, Delius G W and Law R 2010 A jump-growth model for predator–prey dynamics: derivation and application to marine ecosystems *Bull. Math. Biol.* **72** 1361–82

[20] Black A J and McKane A J 2012 Stochastic formulation of ecological models and their applications *Trends Ecol. Evol.* **27** 337–45

[21] Bartlett M S 1949 Some evolutionary stochastic processes *J. R. Stat. Soc.* B **11** 211–29

[22] Bailey N T J 1950 A simple stochastic epidemic *Biometrika* pp 193–202

[23] Keeling M J and Ross J V 2008 On methods for studying stochastic disease dynamics *J. R. Soc. Interface* **5** 171–81

[24] Black A J and McKane A J 2010 Stochasticity in staged models of epidemics: quantifying the dynamics of whooping cough *J. R. Soc. Interface* **7** 1219–27

[25] Jenkinson G and Goutsias J 2012 Numerical integration of the master equation in some models of stochastic epidemiology *PloS One* **7** e36160

[26] Pastor-Satorras R, Castellano C, Van Mieghem P and Vespignani A 2015 Epidemic processes in complex networks *Rev. Mod. Phys.* **87** 925

[27] Weidlich W 1991 Physics and social science? The approach of synergetics *Phys. Rep.* **204** 1–163

[28] Weidlich W 2002 Sociodynamics—a systematic approach to mathematical modelling in the social sciences *Nonlinear Phenom. Complex Syst.—MINSK* **5** 479–87

[29] Bonabeau E 2002 Agent-based modeling: methods and techniques for simulating human systems *Proc. Natl Acad. Sci.* **99** 7280–7

[30] Ohira T and Cowan J D 1993 Master-equation approach to stochastic neurodynamics *Phys. Rev.* E **48** 2259

[31] El Boustani S and Destexhe A 2009 A master equation formalism for macroscopic modeling of asynchronous irregular activity states *Neural Comput.* **21** 46–100

[32] Benayoun M, Cowan J D, van Drongelen W and Wallace E 2010 Avalanches in a stochastic model of spiking neurons *PLoS Comput. Biol.* **6** e1000846

[33] Buice M A, Cowan J D and Chow C C 2010 Systematic fluctuation expansion for neural network activity equations *Neural Comput.* **22** 377–426

[34] Wallace E, Benayoun M, Van Drongelen W and Cowan J D 2011 Emergent oscillations in networks of stochastic spiking neurons *Plos One* **6** e14804

[35] Leen T K and Friel R 2012 Stochastic perturbation methods for spike-timing-dependent plasticity *Neural Comput.* **24** 1109–46

[36] Goychuk I and Goychuk A 2015 Stochastic Wilson–Cowan models of neuronal network dynamics with memory and delay *New J. Phys.* **17** 045029

[37] Alberts B, Bray D, Lewis J, Raff M, Roberts K, Watson J D and Grimstone A V 1995 Molecular biology of the cell (3rd edn) *Trends Biochem. Sci.* **20** 210–0

[38] Swain P S, Elowitz M B and Siggia E D 2002 Intrinsic and extrinsic contributions to stochasticity in gene expression *Proc. Natl Acad. Sci.* **99** 12795–800

[39] Austin D W, Allen M S, McCollum J M, Dar R D, Wilgus J R, Sayler G S, Samatova N F, Cox C D and Simpson M L 2006 Gene network shaping of inherent noise spectra *Nature* **439** 608–11

[40] Ozbudak E M, Thattai M, Kurtser I, Grossman A D and van Oudenaarden A 2002 Regulation of noise in the expression of a single gene *Nat. Genet.* **31** 69–73

[41] Colman-Lerner A, Gordon A, Serra E, Chin T, Resnekov O, Endy D, Gustavo Pesce C and Brent R 2005 Regulated cell-to-cell variation in a cell-fate decision system *Nature* **437** 699–706

[42] Rosenfeld N, Young J W, Alon U, Swain P S and Elowitz M B 2005 Gene regulation at the single-cell level *Science* **307** 1962–5

[43] Bar-Even A, Paulsson J, Maheshri N, Carmi M, O'Shea E, Pilpel Y and Barkai N 2006 Noise in protein expression scales with natural protein abundance *Nat. Genet.* **38** 636–43

[44] Taniguchi Y, Choi P J, Li G-W, Chen H, Babu M, Hearn J, Emili A and Sunney Xie X 2010 Quantifying *E. coli* proteome and transcriptome with single-molecule sensitivity in single cells *Science* **329** 533–8

[45] Brehm-Stecher B F and Johnson E A 2004 Single-cell microbiology: tools, technologies, and applications *Microbiol. Mol. Biol. Rev.* **68** 538–59

[46] Maheshri N and O'Shea E K 2007 Living with noisy genes: how cells function reliably with inherent variability in gene expression *Ann. Rev. Biophys. Biomol. Struct.* **36** 413–34

[47] Balázsi G, van Oudenaarden A and Collins J J 2011 Cellular decision making and biological noise: from microbes to mammals *Cell* **144** 910–25

[48] Guldberg C M and Waage P 1886 Studies concerning affinity *CM Forh.: Vidensk.-Selsk. Christiana* **35**

[49] Waage P 1864 Experiments for determining the affinity law *Forh.: Vidensk.-Selsk. Christiana* **92**

[50] Guldberg C M 1864 Concerning the laws of chemical affinity *CM Forh.: Vidensk.-Selsk. Christiana* **111**

[51] Guldberg C M and Waage P 1879 Concerning chemical affinity *Erdmann's J. Prakt. Chem.* **127** 69–114

[52] Klipp E, Herwig R, Kowald A, Wierling C and Lehrach H 2008 *Systems Biology in Practice: Concepts, Implementation and Application* (New York: Wiley) (doi: 10.1002/3527603603)

[53] Delbrück M 1940 Statistical fluctuations in autocatalytic reactions *J. Chem. Phys.* **8** 120–4

[54] Bartholomay A F 1958 Stochastic models for chemical reactions: I. Theory of the unimolecular reaction process *Bull. Math. Biophys.* **20** 175–90

[55] McQuarrie D A 1963 Kinetics of small systems. I *J. Chem. Phys.* **38** 433–6

[56] Ishida K 1964 Stochastic model for bimolecular reaction *J. Chem. Phys.* **41** 2472–8

[57] McQuarrie D A, Jachimowski C J and Russell M E 1964 Kinetics of small systems. II *J. Chem. Phys.* **40** 2914–21

[58] Gillespie D T 2009 A diffusional bimolecular propensity function *J. Chem. Phys.* **131** 164109

[59] Grima R, Schmidt D R and Newman T J 2012 Steady-state fluctuations of a genetic feedback loop: an exact solution *J. Chem. Phys.* **137** 035104

[60] Maria Guerriero L, Pokhilko A, Fernández A P, Halliday K J, Millar A J and Hillston J 2011 Stochastic properties of the plant circadian clock *J. R. Soc. Interface* **9** 744–56

[61] Anderson D F 2007 A modified next reaction method for simulating chemical systems with time dependent propensities and delays *J. Chem. Phys.* **127** 214107

[62] Shahrezaei V, Ollivier J F and Swain P S 2008 Colored extrinsic fluctuations and stochastic gene expression *Mol. Syst. Biol.* **4** 196

[63] Caravagna G, Mauri G and d'Onofrio A 2013 The interplay of intrinsic and extrinsic bounded noises in biomolecular networks *PLoS One* **8** e51174

[64] Voliotis M, Thomas P, Grima R and Bowsher C G 2016 Stochastic simulation of biomolecular networks in dynamic environments *PLoS Comput. Biol.* **12** e1004923

[65] Al-Mohy A H and Higham N J 2011 Computing the action of the matrix exponential, with an application to exponential integrators *SIAM J. Sci. Comput.* **33** 488–511

[66] Moler C and Van Loan C 2003 Nineteen dubious ways to compute the exponential of a matrix, twenty-five years later *SIAM Rev.* **45** 3–49

[67] Darvey I G and Staff P J 1966 Stochastic approach to first-order chemical reaction kinetics *J. Chem. Phys.* **44** 990–7

[68] Gardiner C W and Chaturvedi S 1977 The poisson representation. I. A new technique for chemical master equations *J. Stat. Phys.* **17** 429–68

[69] Heuett W J and Qian H 2006 Grand canonical markov model: a stochastic theory for open nonequilibrium biochemical networks *J. Chem. Phys.* **124** 044110

[70] Jahnke T and Huisinga W 2007 Solving the chemical master equation for monomolecular reaction systems analytically *J. Math. Biol.* **54** 1–26

[71] Whittle P 1986 *Systems in Stochastic Equilibrium* (New York: Wiley)

[72] Van Kampen N G 1976 The equilibrium distribution of a chemical mixture *Phys. Lett.* A **59** 333–4

[73] Anderson D F, Craciun G and Kurtz T G 2010 Product-form stationary distributions for deficiency zero chemical reaction networks *Bull. Math. Biol.* **72** 1947–70

[74] Horn F and Jackson R 1972 General mass action kinetics *Arch. Ration. Mech. Anal.* **47** 81–116

[75] Feinberg M 1995 The existence and uniqueness of steady states for a class of chemical reaction networks *Arch. Ration. Mech. Anal.* **132** 311–70

[76] Voituriez R, Moreau M and Oshanin G 2005 Corrections to the law of mass action and properties of the asymptotic $t = \infty$ state for reversible diffusion-limited reactions *J. Chem. Phys.* **122** 1

[77] Gerstung M, Timmer J and Fleck C 2009 Noisy signaling through promoter logic gates *Phys. Rev.* E **79** 011923

[78] Thomas P, Straube A V and Grima R 2010 Stochastic theory of large-scale enzyme-reaction networks: finite copy number corrections to rate equation models *J. Chem. Phys.* **133** 195101

[79] Petrosyan K G and Hu C-K 2014 Nonequilibrium lyapunov function and a fluctuation relation for stochastic systems: Poisson-representation approach *Phys. Rev.* E **89** 042132

[80] Iyer-Biswas S and Jayaprakash C 2014 Mixed poisson distributions in exact solutions of stochastic autoregulation models *Phys. Rev.* E **90** 052712

[81] Sugár I and Simon I 2014 Self-regulating genes. Exact steady state solution by using poisson representation *Open Phys.* **12** 615–27

[82] Schnoerr D, Grima R and Sanguinetti G 2016 Cox process representation and inference for stochastic reaction–diffusion processes *Nat. Commun.* **7** 11729

[83] Laurenzi I J 2000 An analytical solution of the stochastic master equation for reversible bimolecular reaction kinetics *J. Chem. Phys.* **113** 3315–22

[84] Kumar N, Platini T and Kulkarni R V 2014 Exact distributions for stochastic gene expression models with bursting and feedback *Phys. Rev. Lett.* **113** 268105

[85] Arányi P and Tóth J 1976 A full stochastic description of the Michaelis–Menten reaction for small systems *Acta Biochim. Biophys. Acad. Sci. Hung.* **12** 375–88

[86] Staff P J 1970 A stochastic development of the reversible Michaelis–Menten mechanism *J. Theor. Biol.* **27** 221–32

[87] Schnoerr D, Sanguinetti G and Grima R 2014 The complex chemical Langevin equation *J. Chem. Phys.* **141** 024103

[88] Mather W H, Cookson N A, Hasty J, Tsimring L S and Williams R J 2010 Correlation resonance generated by coupled enzymatic processing *Biophys. J.* **99** 3172–81

[89] Kramers H A 1940 Brownian motion in a field of force and the diffusion model of chemical reactions *Physica* **7** 284–304

[90] Moyal J E 1949 Stochastic processes and statistical physics *J. R. Stat. Soc.* B **11** 150–210

[91] Gillespie D T 2000 The chemical Langevin equation *J. Chem. Phys.* **113** 297–306

[92] Melykuti B, Burrage K and Zygalakis K C 2010 Fast stochastic simulation of biochemical reaction systems by alternative formulations of the chemical Langevin equation *J. Chem. Phys.* **132** 164109

[93] Kloeden P E and Platen E 1992 *Numerical Solution of Stochastic Differential Equations* vol 1 (Berlin: Springer) (doi: 10.1007/978-3-662-12616-5)

[94] Burrage K, Burrage P and Mitsui T 2000 Numerical solutions of stochastic differential equations–implementation and stability issues *J. Comput. Appl. Math.* **125** 171–82

[95] Higham D J 2001 An algorithmic introduction to numerical simulation of stochastic differential equations *SIAM Rev.* **43** 525–46

[96] Burrage K, Burrage P M and Tian T 2004 Numerical methods for strong solutions of stochastic differential equations: an overview *Proc. R. Soc.* A **460** 373–402

[97] Kurtz T G 1976 Limit theorems and diffusion approximations for density dependent Markov chains *Stochastic Systems: Modeling, Identification and Optimization,* I (Berlin: Springer) pp 67–78

[98] Dana S and Raha S 2011 Physically consistent simulation of mesoscale chemical kinetics: the non-negative FIS-α method *J. Comput. Phys.* **230** 8813–34

[99] Wilkie J and Wong Y M 2008 Positivity preserving chemical Langevin equations *Chem. Phys.* **353** 132–8

[100] Szpruch L and Higham D J 2010 Comparing hitting time behavior of Markov jump processes and their diffusion approximations *Multiscale Model. Simul.* **8** 605–21

[101] Liao S, Vejchodský T and Erban R 2015 Tensor methods for parameter estimation and bifurcation analysis of stochastic reaction networks *J. R. Soc. Interface* **12** 20150233

[102] Biancalani T, Dyson L and McKane A J 2014 Noise-induced bistable states and their mean switching time in foraging colonies *Phys. Rev. Lett.* **112** 038101

[103] Duncan A, Liao S, Vejchodský T, Erban R and Grima R 2015 Noise-induced multistability in chemical systems: discrete versus continuum modeling *Phys. Rev.* E **91** 042111

[104] Beccuti M, Bibbona E, Horvath A, Sirovich R, Angius A and Balbo G 2014 Analysis of petri net models through stochastic differential equations *Int. Conf. on Applications and Theory of Petri Nets and Concurrency* (Berlin: Springer) pp 273–93

[105] van Kampen N G 1976 The expansion of the master equation *Adv. Chem. Phys.* **34** 245–309

[106] Thomas P, Matuschek H and Grima R 2012 Computation of biochemical pathway fluctuations beyond the linear noise approximation using iNA *IEEE Int. Conf. on Bioinformatics and Biomedicine* pp 1–5

[107] Grima R 2010 An effective rate equation approach to reaction kinetics in small volumes: theory and application to biochemical reactions in nonequilibrium steady-state conditions *J. Chem. Phys.* **133** 035101

[108] Thomas P and Grima R 2015 Approximate probability distributions of the master equation *Phys. Rev.* E **92** 012120

[109] McKane A J, Nagy J D, Newman T J and Stefanini M O 2007 Amplified biochemical oscillations in cellular systems *J. Stat. Phys.* **128** 165–91

[110] Pahle J, Challenger J D, Mendes P and McKane A J 2012 Biochemical fluctuations, optimisation and the linear noise approximation *BMC Syst. Biol.* **6** 1

[111] Challenger J D and McKane A J 2013 Synchronization of stochastic oscillators in biochemical systems *Phys. Rev.* E **88** 012107

[112] Bianca C and Lemarchand A 2015 Evaluation of reaction fluxes in stationary and oscillating far-from-equilibrium biological systems *Physica* A **438** 1–16

[113] Hufton P G, Lin Y T, Galla T and McKane A J 2016 Intrinsic noise in systems with switching environments *Phys. Rev.* E **93** 052119

[114] Thomas P, Matuschek H and Grima R 2013 How reliable is the linear noise approximation of gene regulatory networks? *BMC Genom.* **14** S5

[115] Dauxois T, Di Patti F, Fanelli D and McKane A J 2009 Enhanced stochastic oscillations in autocatalytic reactions *Phys. Rev.* E **79** 036112

[116] Cianci C, Di Patti F, Fanelli D and Barletti L 2012 Analytical study of non Gaussian fluctuations in a stochastic scheme of autocatalytic reactions *Eur. Phys. J. Spec. Top.* **212** 5–22

[117] Scott M 2012 Non-linear corrections to the time-covariance function derived from a multi-state chemical master equation *IET Syst. Biol.* **6** 116–24

[118] Thomas P, Straube A V, Timmer J, Fleck C and Grima R 2013 Signatures of nonlinearity in single cell noise-induced oscillations *J. Theor. Biol.* **335** 222–34

[119] Pawula R F 1967 Approximation of the linear Boltzmann equation by the Fokker–Planck equation *Phys. Rev.* **162** 186

[120] Leen T K and Friel R 2011 Perturbation theory for stochastic learning dynamics *The 2011 Int. Joint Conf. on Neural Networks* pp 2031–8

[121] Goodman L A 1953 Population growth of the sexes *Biometrics* **9** 212–25

[122] Whittle P 1957 On the use of the normal approximation in the treatment of stochastic processes *J. R. Stat. Soc.* B **19** 268–81

[123] Gomez-Uribe C A and Verghese G C 2007 Mass fluctuation kinetics: capturing stochastic effects in systems of chemical reactions through coupled mean-variance computations *J. Chem. Phys.* **126** 024109

[124] Nåsell I 2003 An extension of the moment closure method *Theor. Popul. Biol.* **64** 233–9

[125] Keeling M J 2000 Multiplicative moments and measures of persistence in ecology *J. Theor. Biol.* **205** 269–81

[126] Crow E L and Shimizu K 1988 *Lognormal Distributions: Theory and Applications* vol 88 (New York: M Dekker)

[127] Hespanha J 2008 Moment closure for biochemical networks *3rd Int. Symp. on Communications, Control and Signal Processing* pp 142–7

[128] Singh A and Hespanha J P 2006 Lognormal moment closures for biochemical reactions *IEEE Conf. on Decision and Control* pp 2063–8

[129] Grima R 2012 A study of the accuracy of moment-closure approximations for stochastic chemical kinetics *J. Chem. Phys.* **136** 154105

[130] Ale A, Kirk P and Stumpf M P H 2013 A general moment expansion method for stochastic kinetic models *J. Chem. Phys.* **138** 174101

[131] Schnoerr D, Sanguinetti G and Grima R 2014 Validity conditions for moment closure approximations in stochastic chemical kinetics *J. Chem. Phys.* **141** 084103

[132] Schnoerr D, Sanguinetti G and Grima R 2015 Comparison of different moment-closure approximations for stochastic chemical kinetics *J. Chem. Phys.* **143** 185101

[133] Singh A and Hespanha J P 2011 Approximate moment dynamics for chemically reacting systems *IEEE Trans. Autom. Control* **56** 414–8

[134] Lakatos E, Ale A, Kirk P D W and Stumpf M P H 2015 Multivariate moment closure techniques for stochastic kinetic models *J. Chem. Phys.* **143** 094107

[135] Shannon C E 2001 A mathematical theory of communication *ACM SIGMOBILE Mobile Comput. Commun. Rev.* **5** 3–55

[136] Andreychenko A, Mikeev L and Wolf V 2015 Model reconstruction for moment-based stochastic chemical kinetics *ACM Trans. Model. Comput. Simul.* **25** 12

[137] Andreychenko A, Bortolussi L, Grima R, Thomas P and Wolf V 2015 Distribution approximations for the chemical master equation: comparison of the method of moments and the system size expansion (arXiv:1509.09104)

[138] Smadbeck P and Kaznessis Y N 2013 A closure scheme for chemical master equations *Proc. Natl Acad. Sci.* **110** 14261–5

[139] Ramsey S, Orrell D and Bolouri H 2005 Dizzy: stochastic simulation of large-scale genetic regulatory networks *J. Bioinf. Comput. Biol.* **3** 415–36

[140] Hoops S, Sahle S, Gauges R, Lee C, Pahle J, Simus N, Singhal M, Xu L, Mendes P and Kummer U 2006 COPASI—a complex pathway simulator *Bioinformatics* **22** 3067–74

[141] Sanft K R, Wu S, Roh M, Fu J, Lim R K and Petzold L R 2011 StochKit2: software for discrete stochastic simulation of biochemical systems with events *Bioinformatics* **27** 2457–8

[142] Maarleveld T R, Olivier B G and Bruggeman F J 2013 StochPy: a comprehensive, user-friendly tool for simulating stochastic biological processes *PloS One* **8** e79345

[143] Thomas P, Matuschek H and Grima R 2012 Intrinsic noise analyzer: a software package for the exploration of stochastic biochemical kinetics using the system size expansion *PloS One* **7** e38518

[144] Hespanha J P 2006 StochDynTools—a MATLAB toolbox to compute moment dynamics for stochastic networks of bio-chemical reactions (http://www.ece.ucsb.edu/~hespanha/software/stochdyntool.html)

[145] Gillespie C S 2009 Moment-closure approximations for mass-action models *Syst. Biol. IET* **3** 52–8

[146] Fan S, Geissmann Q, Lakatos E, Lukauskas S, Ale A, Babtie A C, Kirk P D W and Stumpf M P H 2016 MEANS: python package for Moment Expansion Approximation, iNference and Simulation *Bioinformatics* **32** 2863–5

[147] Kazeroonian A, Fröhlich F, Raue A, Theis F J and Jan Hasenauer 2016 CERENA: chemical reaction network analyzer—a toolbox for the simulation and analysis of stochastic chemical kinetics *PloS One* **11** e0146732

[148] Munsky B and Khammash M 2006 The finite state projection algorithm for the solution of the chemical master equation *J. Chem. Phys.* **124** 044104

[149] Dinh K N and Sidje R B 2016 Understanding the finite state projection and related methods for solving the chemical master equation *Phys. Biol.* **13** 035003

[150] Gillespie D T 2001 Approximate accelerated stochastic simulation of chemically reacting systems *J. Chem. Phys.* **115** 1716–33

[151] Rathinam M, Petzold L R, Cao Y and Gillespie D T 2003 Stiffness in stochastic chemically reacting systems: the implicit tau-leaping method *J. Chem. Phys.* **119** 12784–94

[152] Cao Y, Gillespie D T and Petzold L R 2007 Adaptive explicit–implicit tau-leaping method with automatic tau selection *J. Chem. Phys.* **126** 224101

[153] Tian T and Burrage K 2004 Binomial leap methods for simulating stochastic chemical kinetics *J. Chem. Phys.* **121** 10356–64

[154] Chatterjee A, Vlachos D G and Katsoulakis M A 2005 Binomial distribution based τ-leap accelerated stochastic simulation *J. Chem. Phys.* **122** 024112

[155] Peng X, Zhou W and Wang Y 2007 Efficient binomial leap method for simulating chemical kinetics *J. Chem. Phys.* **126** 224109

[156] Leier A, Marquez-Lago T T and Burrage K 2008 Generalized binomial τ-leap method for biochemical kinetics incorporating both delay and intrinsic noise *J. Chem. Phys.* **128** 205107

[157] Pettigrew M F and Resat H 2007 Multinomial tau-leaping method for stochastic kinetic simulations *J. Chem. Phys.* **126** 084101

[158] Cai X and Xu Z 2007 K-leap method for accelerating stochastic simulation of coupled chemical reactions *J. Chem. Phys.* **126** 74102–2

[159] Anderson D F 2008 Incorporating postleap checks in tau-leaping *J. Chem. Phys.* **128** 054103

[160] Gillespie D T 2008 Simulation methods in systems biology *Formal Methods for Computational Systems Biology* vol 5016 (Berlin: Springer) pp 125–67

[161] Bowen J R, Acrivos A and Oppenheim A K 1963 Singular perturbation refinement to quasi-steady state approximation in chemical kinetics *Chem. Eng. Sci.* **18** 177–88

[162] Janssen J A M 1989 The elimination of fast variables in complex chemical reactions. II. Mesoscopic level (reducible case) *J. Stat. Phys.* **57** 171–85

[163] Heinrich R and Schuster S 2012 *The Regulation of Cellular Systems* (Berlin: Springer) (doi: 10.1007/978-1-4613-1161-4)

[164] Michaelis L and Menten M L 1913 Die Kinetik der Invertinwirkung *Biochem. Z.* **49** 352

[165] Segel L A and Slemrod M 1989 The quasi-steady-state assumption: a case study in perturbation *SIAM Rev.* **31** 446–77

[166] Briggs G E and Haldane B S J 1925 A note on the kinetics of enzyme action *Biochem. J.* **19** 338

[167] Goutsias J 2005 Quasiequilibrium approximation of fast reaction kinetics in stochastic biochemical systems *J. Chem. Phys.* **122** 184102

[168] Janssen J A M 1989 The elimination of fast variables in complex chemical reactions. III. Mesoscopic level (irreducible case) *J. Stat. Phys.* **57** 187–98

[169] Gonze D, Halloy J and Goldbeter A 2002 Deterministic versus stochastic models for circadian rhythms *J. Biol. Phys.* **28** 637–53

[170] Rao C V and Arkin A P 2003 Stochastic chemical kinetics and the quasi-steady-state assumption: application to the gillespie algorithm *J. Chem. Phys.* **118** 4999–5010

[171] Sanft K R, Gillespie D T and Petzold L R 2011 Legitimacy of the stochastic Michaelis–Menten approximation *IET Syst. Biol.* **5** 58–69

[172] Thomas P, Straube A V and Grima R 2011 Communication: limitations of the stochastic quasi-steady-state approximation in open biochemical reaction networks *J. Chem. Phys.* **135** 181103

[173] Kim J K, Josić K and Bennett M R 2015 The relationship between stochastic and deterministic quasi-steady state approximations *BMC Syst. Biol.* **9** 87

[174] Thomas P, Straube A V and Grima R 2012 The slow-scale linear noise approximation: an accurate, reduced stochastic description of biochemical networks under timescale separation conditions *BMC Syst. Biol.* **6** 39

[175] Cao Y, Gillespie D T and Petzold L R 2005 The slow-scale stochastic simulation algorithm *J. Chem. Phys.* **122** 014116

[176] Salis H and Kaznessis Y N 2005 An equation-free probabilistic steady-state approximation: dynamic application to the stochastic simulation of biochemical reaction networks *J. Chem. Phys.* **123** 214106

[177] Weinan E, Liu D and Vanden-Eijnden E 2005 Nested stochastic simulation algorithm for chemical kinetic systems with disparate rates *J. Chem. Phys.* **123** 194107

[178] Samant A and Vlachos D G 2005 Overcoming stiffness in stochastic simulation stemming from partial equilibrium: a multiscale Monte Carlo algorithm *J. Chem. Phys.* **123** 144114

[179] Weinan E, Liu D and Vanden-Eijnden E 2007 Nested stochastic simulation algorithms for chemical kinetic systems with multiple time scales *J. Computat. Phys.* **221** 158–80

[180] Gómez-Uribe C A, Verghese G C and Tzafriri A R 2008 Enhanced identification and exploitation of time scales for model reduction in stochastic chemical kinetics *J. Chem. Phys.* **129** 244112

[181] Pigolotti S and Vulpiani A 2008 Coarse graining of master equations with fast and slow states *J. Chem. Phys.* **128** 154114

J. Phys. A: Math. Theor. **50** (2017) 093001

Topical Review

[182] Chevalier M W and El-Samad H 2009 A rigorous framework for multiscale simulation of stochastic cellular networks *J. Chem. Phys.* **131** 054102

[183] Cotter S L, Zygalakis K C, Kevrekidis I G and Erban R 2011 A constrained approach to multiscale stochastic simulation of chemically reacting systems *J. Chem. Phys.* **135** 094102

[184] Kang H-W *et al* 2013 Separation of time-scales and model reduction for stochastic reaction networks *Ann. Appl. Probab.* **23** 529–83

[185] Bortolussi L, Milios D and Sanguinetti G 2015 Efficient stochastic simulation of systems with multiple time scales via statistical abstraction *Int. Conf. on Computational Methods in Systems Biology* (Berlin: Springer) pp 40–51

[186] Shahrezaei V and Swain P S 2008 Analytical distributions for stochastic gene expression *Proc. Natl Acad. Sci.* **105** 17256–61

[187] Bokes P, King J R, Wood A T A and Loose M 2012 Multiscale stochastic modelling of gene expression *J. Math. Biol.* **65** 493–520

[188] Mélykúti B, Hespanha J P and Khammash M 2014 Equilibrium distributions of simple biochemical reaction systems for time-scale separation in stochastic reaction networks *J. R. Soc. Interface* **11** 20140054

[189] Popović N, Marr C and Swain P S 2016 A geometric analysis of fast–slow models for stochastic gene expression *J. Math. Biol.* **72** 87–122

[190] Haseltine E L and Rawlings J B 2002 Approximate simulation of coupled fast and slow reactions for stochastic chemical kinetics *J. Chem. Phys.* **117** 6959–69

[191] Salis H and Kaznessis Y 2005 Accurate hybrid stochastic simulation of a system of coupled chemical or biochemical reactions *J. Chem. Phys.* **122** 054103

[192] Peleš S, Munsky B and Khammash M 2006 Reduction and solution of the chemical master equation using time scale separation and finite state projection *J. Chem. Phys.* **125** 204104

[193] Puchałka J and Kierzek A M 2004 Bridging the gap between stochastic and deterministic regimes in the kinetic simulations of the biochemical reaction networks *Biophys. J.* **86** 1357–72

[194] Pahlajani C D, Atzberger P J and Khammash M 2011 Stochastic reduction method for biological chemical kinetics using time-scale separation *J. Theor. Biol.* **272** 96–112

[195] Thomas P, Grima R and Straube A V 2012 Rigorous elimination of fast stochastic variables from the linear noise approximation using projection operators *Phys. Rev.* E **86** 041110

[196] Sinitsyn N A, Hengartner N and Nemenman I 2009 Adiabatic coarse-graining and simulations of stochastic biochemical networks *Proc. Natl Acad. Sci.* **106** 10546–51

[197] Burrage K, Tian T and Burrage P 2004 A multi-scaled approach for simulating chemical reaction systems *Prog. Biophys. Mol. Biol.* **85** 217–34

[198] Thomas P, Popović N and Grima R 2014 Phenotypic switching in gene regulatory networks *Proc. Natl Acad. Sci.* **111** 6994–9

[199] Vasudeva K and Bhalla U S 2004 Adaptive stochastic-deterministic chemical kinetic simulations *Bioinformatics* **20** 78–84

[200] Alfonsi A, Cances E, Turinici G, Di Ventura B and Huisinga W 2004 Exact simulation of hybrid stochastic and deterministic models for biochemical systems *Rr-5435, INRIA-Rocquencourt*

[201] Takahashi K, Kaizu K, Hu B and Tomita M 2004 A multi-algorithm, multi-timescale method for cell simulation *Bioinformatics* **20** 538–46

[202] Neogi N A 2004 Dynamic partitioning of large discrete event biological systems for hybrid simulation and analysis *Hybrid Systems: Computation and Control* (Berlin: Springer) pp 463–76

[203] Bentele M and Eils R 2004 General stochastic hybrid method for the simulation of chemical reaction processes in cells *Computational Methods in Systems Biology* (Berlin: Springer) pp 248–51

[204] Harris L A and Clancy P 2006 A 'partitioned leaping' approach for multiscale modeling of chemical reaction dynamics *J. Chem. Phys.* **125** 144107

[205] Hellander A and Lötstedt P 2007 Hybrid method for the chemical master equation *J. Comput. Phys.* **227** 100–22

[206] Crudu A, Debussche A and Radulescu O 2009 Hybrid stochastic simplifications for multiscale gene networks *BMC Syst. Biol.* **3** 1

[207] Menz S, Latorre J C, Schutte C and Huisinga W 2012 Hybrid stochastic—deterministic solution of the chemical master equation *Multiscale Model. Simul.* **10** 1232–62

[208] Ganguly A, Altintan D and Koeppl H 2015 Jump-diffusion approximation of stochastic reaction dynamics: error bounds and algorithms *Multiscale Model. Simul.* **13** 1390–419

[209] Duncan A, Erban R and Zygalakis K 2016 Hybrid framework for the simulation of stochastic chemical kinetics *J. Comput. Phys.* **326** 398–419

[210] Safta C, Sargsyan K, Debusschere B and Najm H N 2015 Hybrid discrete/continuum algorithms for stochastic reaction networks *J. Comput. Phys.* **281** 177–98

[211] Hepp B, Gupta A and Khammash M 2015 Adaptive hybrid simulations for multiscale stochastic reaction networks *J. Chem. Phys.* **142** 034118

[212] Angius A, Balbo G, Beccuti M, Bibbona E, Horvath A and Sirovich R 2015 Approximate analysis of biological systems by hybrid switching jump diffusion *Theor. Comput. Sci.* **587** 49–72

[213] Jahnke T 2011 On reduced models for the chemical master equation *Multiscale Model. Simul.* **9** 1646–76

[214] Jahnke T and Kreim M 2012 Error bound for piecewise deterministic processes modeling stochastic reaction systems *Multiscale Model. Simul.* **10** 1119–47

[215] Crudu A *et al* 2012 Convergence of stochastic gene networks to hybrid piecewise deterministic processes *Ann. Appl. Probab.* **22** 1822–59

[216] Cotter S L and Erban R 2016 Error analysis of diffusion approximation methods for multiscale systems in reaction kinetics *SIAM J. Sci. Comput.* **38** B144–63

[217] Hasenauer J, Wolf V, Kazeroonian A and Theis F J 2014 Method of conditional moments (MCM) for the chemical master equation *J. Math. Biol.* **69** 687–735

[218] Soltani M, Vargas-Garcia C A and Singh A 2015 Conditional moment closure schemes for studying stochastic dynamics of genetic circuits *IEEE Trans. Biomed. Circuits Syst.* **9** 518–26

[219] Smith S, Cianci C and Grima R 2015 Model reduction for stochastic chemical systems with abundant species *J. Chem. Phys.* **143** 214105

[220] Thattai M and Van Oudenaarden A 2001 Intrinsic noise in gene regulatory networks *Proc. Natl Acad. Sci.* **98** 8614–9

[221] Bortolussi L and Sanguinetti G 2013 Learning and designing stochastic processes from logical constraints *QEST* vol 8054 (Berlin: Springer) pp 89–105

[222] Cseke B, Opper M and Sanguinetti G 2013 Approximate inference in latent Gaussian–Markov models from continuous time observations *Advances in Neural Information Processing Systems* (Cambridge, MA: MIT Press) pp 971–9

[223] Zechner C, Unger M, Pelet S, Peter M and Koeppl H 2014 Scalable inference of heterogeneous reaction kinetics from pooled single-cell recordings *Nat. Methods* **11** 197–202

[224] Bishop C M 2006 Pattern recognition and machine learning *Machine Learning* (New York: Springer)

[225] Rabiner L and Juang B 1986 An introduction to hidden Markov models *IEEE ASSP Mag.* **3** 4–16

[226] Kalman R E 1960 A new approach to linear filtering and prediction problems *J. Basic Eng.* **82** 35–45

[227] Frühwirth-Schnatter S 1994 Data augmentation and dynamic linear models *J. Time Ser. Anal.* **15** 183–202

[228] Opper M and Sanguinetti G 2008 Variational inference for Markov jump processes In *Advances in Neural Information Processing Systems* (Cambridge, MA: MIT Press) pp 1105–12

[229] Rao V and Teh Y W 2013 Fast MCMC sampling for Markov jump processes and extensions *J. Mach. Learn. Res.* **14** 3295–320

[230] Georgoulas A, Hillston J and Sanguinetti G 2016 Unbiased Bayesian inference for population Markov jump processes via random truncations *Stat. Comput.* 1–12

[231] Boys R J, Wilkinson D J and Kirkwood T B L 2008 Bayesian inference for a discretely observed stochastic kinetic model *Stat. Comput.* **18** 125–35

[232] Golightly A and Wilkinson D J 2011 Bayesian parameter inference for stochastic biochemical network models using particle Markov chain Monte Carlo *Interface Focus* **1** 807–20

[233] Golightly A and Wilkinson D J 2005 Bayesian inference for stochastic kinetic models using a diffusion approximation *Biometrics* **61** 781–8

[234] Ruttor A and Opper M 2009 Efficient statistical inference for stochastic reaction processes *Phys. Rev. Lett.* **103** 230601

[235] Stathopoulos V and Girolami M A 2013 Markov chain Monte Carlo inference for Markov jump processes via the linear noise approximation *Phil. Trans. R. Soc.* A **371** 20110541

[236] Fearnhead P, Giagos V and Sherlock C 2014 Inference for reaction networks using the linear noise approximation *Biometrics* **70** 457–66

[237] Milner P, Gillespie C S and Wilkinson D J 2013 Moment closure based parameter inference of stochastic kinetic models *Stat. Comput.* **23** 287–95

[238] Cseke B, Schnoerr D, Opper M and Sanguinetti G 2016 Expectation propagation for continuous time stochastic processes *J. Phys. A: Math. Theor.* **49** 494002

[239] Zechner C, Ruess J, Krenn P, Pelet S, Peter M, Lygeros J and Koeppl H 2012 Moment-based inference predicts bimodality in transient gene expression *Proc. Natl Acad. Sci.* **109** 8340–5

[240] Stimberg F, Opper M, Sanguinetti G and Ruttor A 2011 Inference in continuous-time change-point models *Advances in Neural Information Processing Systems* (Cambridge, MA: MIT Press) pp 2717–25

[241] Ocone A and Sanguinetti G 2011 Reconstructing transcription factor activities in hierarchical transcription network motifs *Bioinformatics* **27** 2873–9

[242] Ocone A, Millar A J and Sanguinetti G 2013 Hybrid regulatory models: a statistically tractable approach to model regulatory network dynamics *Bioinformatics* **29** 910–6

[243] Sherlock C, Golightly A and Gillespie C S 2014 Bayesian inference for hybrid discrete-continuous stochastic kinetic models *Inverse Problems* **30** 114005

[244] Hey K L, Momiji H, Featherstone K, Davis J R E, White M R H, Rand D A and Finkenstädt B 2015 A stochastic transcriptional switch model for single cell imaging data *Biostatistics* **16** 655–69

[245] Liepe J, Barnes C, Cule E, Erguler K, Kirk P, Toni T and Stumpf M P H 2010 ABC-SysBio–approximate bayesian computation in python with GPU support *Bioinformatics* **26** 1797–9

[246] Chen Y, Lawless C, Gillespie C S, Wu J, Boys R J and Wilkinson D J 2010 CaliBayes and BASIS: integrated tools for the calibration, simulation and storage of biological simulation models *Briefings Bioinf.* **11** 278–89

[247] Doi M 1976 Second quantization representation for classical many-particle system *J. Phys. A: Math. Gen.* **9** 1465

[248] Doi M 1976 Stochastic theory of diffusion-controlled reaction *J. Phys. A: Math. Gen.* **9** 1479

[249] Erban R and Chapman J S 2009 Stochastic modelling of reaction–diffusion processes: algorithms for bimolecular reactions *Phys. Biol.* **6** 046001

[250] van Zon J S and Ten Wolde P R 2005 Simulating biochemical networks at the particle level and in time and space: Green's function reaction dynamics *Phys. Rev. Lett.* **94** 128103

[251] Donev A, Bulatov V V, Oppelstrup T, Gilmer G H, Sadigh B and Kalos M H 2010 A first-passage kinetic Monte Carlo algorithm for complex diffusion-reaction systems *J. Comput. Phys.* **229** 3214–36

[252] Isaacson S A 2008 Relationship between the reaction–diffusion master equation and particle tracking models *J. Phys. A: Math. Theor.* **41** 065003

[253] Fu J, Wu S, Li H and Petzold L R 2014 The time dependent propensity function for acceleration of spatial stochastic simulation of reaction–diffusion systems *J. Comput. Phys.* **274** 524–49

[254] Isaacson S A 2009 The reaction–diffusion master equation as an asymptotic approximation of diffusion to a small target *SIAM J. Appl. Math.* **70** 77–111

[255] Lipková J, Zygalakis K C, Jonathan Chapman S and Erban R 2011 Analysis of brownian dynamics simulations of reversible bimolecular reactions *SIAM J. Appl. Math.* **71** 714–30

[256] Hellander S, Hellander A and Petzold L 2012 Reaction-diffusion master equation in the microscopic limit *Phys. Rev.* E **85** 042901

[257] Mahmutovic A, Fange D, Berg O G and Elf J 2012 Lost in presumption: stochastic reactions in spatial models *Nat. Methods* **9** 1163–6

[258] Isaacson S A 2013 A convergent reaction–diffusion master equation *J. Chem. Phys.* **139** 054101

[259] Taylor P R, Baker R E and Yates C A 2014 Deriving appropriate boundary conditions, and accelerating position-jump simulations, of diffusion using non-local jumping *Phys. Biol.* **12** 016006

[260] Scott M, Poulin F J and Tang H 2010 Approximating intrinsic noise in continuous multispecies models *Proc. R. Soc.* A **467** 718–37

[261] Asllani M, Biancalani T, Fanelli D and McKane A J 2013 The linear noise approximation for reaction–diffusion systems on networks *Eur. Phys. J.* B **86** 1–10

[262] Smith S, Cianci C and Grima R 2016 Analytical approximations for spatial stochastic gene expression in single cells and tissues *J. R. Soc. Interface* **13** 20151051

[263] Holmes G R, Anderson S R, Dixon G, Robertson A L, Reyes-Aldasoro C C, Billings S A, Renshaw S A and Kadirkamanathan V 2012 Repelled from the wound, or randomly dispersed? Reverse migration behaviour of neutrophils characterized by dynamic modelling *J. R. Soc. Interface* **9** 3229–39

[264] Dewar M A, Kadirkamanathan V, Opper M and Sanguinetti G 2010 Parameter estimation and inference for stochastic reaction–diffusion systems: application to morphogenesis in D. melanogaster *BMC Syst. Biol.* **4** 21

[265] Ruttor A and Opper M 2010 Approximate parameter inference in a stochastic reaction–diffusion model *Int. Conf. on Artificial Intelligence and Statistics* vol 9 pp 669–76

[266] Cressie N and Wikle C K 2015 *Statistics for Spatio-Temporal Data* (New York: Wiley)