



Article submitted to journal

Subject Areas:

systems theory, computational biology

Keywords:

stochastic reaction networks, reaction rate equations, mean approximations, master equation, continuous-time Markov chains

Author for correspondence:

Mirco Tribastone

e-mail: mirco.tribastone@imtlucca.it

Improved estimations of stochastic chemical kinetics by finite state expansion

Tabea Waizmann¹, Luca Bortolussi²,
Andrea Vandin^{3,4}, and Mirco Tribastone¹

¹IMT School for Advanced Studies, Lucca, 55100, Italy,

²Department of Mathematics and Geosciences,
University of Trieste, 34127, Italy,

³Sant'Anna School of Advanced Studies, Pisa, 56127,
Italy, and

⁴DTU Technical University of Denmark, Kongens
Lyngby, 2800, Denmark.

Stochastic reaction networks are a fundamental model to describe interactions between species where random fluctuations are relevant. The master equation provides the evolution of the probability distribution across the discrete state space consisting of vectors of population counts for each species. However, since its exact solution is often elusive, several analytical approximations have been proposed. The deterministic rate equation (DRE) gives a macroscopic approximation as a compact system of differential equations that estimate the average populations for each species, but it may be inaccurate in the case of nonlinear interaction dynamics. Here we propose finite state expansion (FSE), an analytical method mediating between the microscopic and the macroscopic interpretations of a stochastic reaction network by coupling the master equation dynamics of a chosen subset of the discrete state space with the mean population dynamics of the DRE. An algorithm translates a network into an expanded one where each discrete state is represented as a further distinct species. This translation exactly preserves the stochastic dynamics, but the DRE of the expanded network can be interpreted as a correction to the original one. The effectiveness of FSE is demonstrated in models that challenge state-of-the-art techniques due to intrinsic noise, multi-scale populations, and multi-stability.

1. Introduction

Stochastic reaction networks are a fundamental model to analyze species that interact stochastically through reaction channels according to dynamics governed by the well-known master equation [1]. This provides a microscopic description in terms of a set of coupled linear differential equations, each defining the time course of a discrete state of the system as a vector of population counts of the species involved. It is widely understood, however, that the master equation is intractable in general, since analytical solutions are available only in special cases and direct numerical integration is hindered by the combinatorial growth of the state space as a function of the abundances of the species. Alternatively, it is possible to analyze the network by means of stochastic simulation, e.g. [2]. However, this may preclude other important studies such as stability, perturbation analysis, bifurcation, and parameter inference [3,4]. Thus, it is often useful to consider analytical approximations that trade off precision with computational cost [5].

The deterministic rate equation (DRE) provides a macroscopic dynamical view by associating one ordinary differential equation with each species representing its mean population. This is exact if each reaction's propensity function is linear, for instance in monomolecular chemical reaction networks [2], and for certain classes of networks with bimolecular reactions [6]. In general, however, with nonlinear propensities functions, the DRE does give the true expectations only in the thermodynamic limit under mild conditions [7]. In this case, away from this asymptotic regime the DREs are only an approximation to the true mean dynamics. This occurs, for instance, in models of cell regulation that depend on low-abundance species (in the order of a few units) to describe the behavior of genes [8]. Processes such as activation and deactivation that vary with time as a result of various interactions may introduce significant variability in gene expression [9], caused by inherent stochasticity in the bio-molecular processes involved [10,11]. Since such forms of noise are not accounted for in the DRE, approximation errors may be large.

Here we present finite state expansion (FSE), an analytical method that keeps track discretely of only a user-defined subset of the state space while collapsing the rest as a continuous approximation. In particular, the state space to be tracked discretely is determined by a parameter that specifies the maximum allowed population level for each species. FSE is a systematic translation of a stochastic reaction network with arbitrary propensity functions into an expanded one with additional species and modified reactions. Specifically, each tracked discrete state is represented as a new auxiliary species; the original set of reactions is transformed such that the dynamics of the auxiliary species are coupled with those of the original species, whose role is to buffer the probability mass that falls out of the state space that is tracked.

FSE enjoys two useful properties. The first concerns its *soundness*, in the sense that any expanded network is stochastically equivalent to the original one: the state space of the expanded network can be projected onto a lower-dimensional one that still satisfies the Markov property, according to [12], and which turns out to correspond to the original network. Such correspondence does not carry over to the respective DREs of the original and of the expanded networks. Indeed, any expansion arising from a strict subset of the discrete state space will lead to a DRE with more equations, which can be interpreted as refining terms for the mean estimates. Our second theoretical contribution is a result of *asymptotic correctness*, stating that if every discrete state is tracked then the DRE of the expanded network corresponds to the master equation.

There are several analytical approaches that can be used to improve the accuracy of the DRE. These include moment-closure approximations [13], the effective mesoscopic rate equation, which adds correction terms to van Kampen's well-known system size expansion [14,15], and hybrid techniques [16]; [3] offers an up-to-date review. However, they are applicable under certain assumptions such as smoothness of the propensity functions [14,17,18], mass-action kinetics [19–23], specific network structures, e.g., to describe gene regulatory systems [24,25], and species that can be partitioned into low-abundance and high-abundance classes [16,26,27]. FSE, instead, can be applied to networks with propensity functions of arbitrary form. Using selected

case studies from the literature we show how FSE can provide accurate mean estimates in model instances that challenge state-of-the-art methods.

2. Background theory

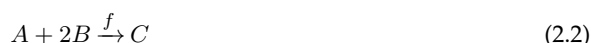
We briefly review here the preliminary definitions and notation on stochastic reaction networks used in the paper.

Consider a set of species \mathcal{S} . Then, $\mathbb{N}^{\mathcal{S}}$ and $\mathbb{R}^{\mathcal{S}}$ are the sets of all integer and real-valued vectors, respectively, with coordinates represented by the elements in \mathcal{S} . For a given vector $\sigma \in \mathbb{R}^{\mathcal{S}}$ (or $\sigma \in \mathbb{N}^{\mathcal{S}}$), we denote by σ_S the value of the component corresponding to species $S \in \mathcal{S}$. We generalize binary operations to the case where operands σ and μ are such that $\sigma \in \mathbb{R}^{\mathcal{S}_1}$ and $\mu \in \mathbb{R}^{\mathcal{S}_2}$, with $\mathcal{S}_1 \neq \mathcal{S}_2$: each binary operation treats them as elements of $\mathbb{R}^{\mathcal{S}_1 \cup \mathcal{S}_2}$.

Formally, we denote a reaction network as a pair $(\mathcal{S}, \mathcal{R})$, where \mathcal{R} is a set of reactions. Each reaction is provided as a triple in the form



where $\rho \in \mathbb{N}^{\mathcal{S}}$ are the *reactants*, $\pi \in \mathbb{N}^{\mathcal{S}}$ are the *products*, and f is the propensity function, $f: \mathbb{R}^{\mathcal{S}} \rightarrow \mathbb{R}_0^+$, with arbitrary form. We will use the standard notation for reactants and products whereby only the nonzero components are written out, separated by the plus sign. For instance, given the species $\mathcal{S} = \{A, B, C\}$, Eq. 2.1 corresponds to the reaction



when $\rho_A = 1$, $\rho_B = 2$, $\rho_C = 0$ and $\pi_A = \pi_B = 0$, $\pi_C = 1$. A discrete state of a reaction network is described by a vector $\sigma \in \mathbb{N}^{\mathcal{S}}$, where σ_S denotes the population of species S in that state. Then, $f(\sigma)$ is the parameter of the exponential distribution of the firing time of that reaction. Upon firing, the system may transition from state σ to $\sigma + \pi - \rho$, thus defining a stochastic behavior in terms of a Markov jump process. Its dynamics is defined by the master equation. It gives the probability $P_\sigma(t)$ of finding the Markov chain in state σ at time t :

$$\frac{dP_\sigma(t)}{dt} = \sum_{\rho \xrightarrow{f} \pi} -f(\sigma)P_\sigma(t) + f(\sigma + \rho - \pi)P_{\sigma + \rho - \pi}(t).$$

Formally, the master equation may be defined for all $\sigma \in \mathbb{N}^{\mathcal{S}}$. However, its solution will be nonzero only for those states that are reachable from the states that have nonzero probability at time $t = 0$. The reachable set of states, also called the *state space*, can be defined as the smallest set such that the following hold:

- (i) σ is in the reachable set if $P_\sigma(0) > 0$;
- (ii) σ is in the reachable set if σ' is in the reachable set and there exists a reaction $\rho \xrightarrow{f} \pi$ such that $\sigma' + \pi - \rho = \sigma$.

For simplicity (and without loss of generality) we will consider networks where the initial probability distribution $P_\sigma(0)$ is concentrated in one state only, which is called the *initial state*. Additionally we shall restrict to *well-defined* reaction networks where each propensity function evaluates to zero for all multisets that do not have the minimum population counts described by the reactants. Formally, a reaction network is well-defined if every reaction $\rho \xrightarrow{f} \pi$ is such that $f(\sigma) = 0$ if $\rho > \sigma$ (the inequalities shall be intended component-wise from now on). This guarantees that the Markov chain does not reach states with negative population counts.

The state space, hence the number of equations required for stochastic analysis, may be finite or infinite depending on the network stoichiometries. Even in the case of finite state spaces, its size may grow combinatorially large with the population counts of the initial state. This may practically preclude exact analysis in most models of interest. The DRE provides a compact model

with $|\mathcal{S}|$ variables. Each variable approximates the expected population level of each species at time t , denoted by the vector $X(t) \in \mathbb{R}^{\mathcal{S}}$, as the solution of the system:

$$\frac{dX(t)}{dt} = \sum_{\rho \xrightarrow{f} \pi} f(X(t))(\pi - \rho).$$

The true expected population counts, denoted by $\mathbb{E}[Y]$, are known to satisfy

$$\frac{d\mathbb{E}[Y(t)]}{dt} = \sum_{\rho \xrightarrow{f} \pi} \mathbb{E}[f(Y(t))](\pi - \rho).$$

However this system is not self-consistent because there are no equations for the expected values of the propensity functions appearing in the right-hand sides. The DRE *closes* the true equations for the expected values by replacing $\mathbb{E}[f(Y(t))]$ with $f(\mathbb{E}[Y(t)])$, introducing an approximation error if the propensity functions are not linear. Under mild conditions such error is known to vanish asymptotically when the initial population levels go to infinity and the DRE is understood as a system of re-scaled equations for the concentrations of species, rather than absolute population counts [7].

3. Methodology

(a) Finite state expansion

With FSE, the original set of species \mathcal{S} of a reaction network is meant to represent the continuous dynamics. This is expanded with a set of auxiliary species, each tracking a specific discrete state. The auxiliary species are defined by the user through an upper bound to the population count to be tracked discretely for each species. Thus, in effect FSE yields a lattice of expansions depending on the choice of the upper bounds.

Let us denote by $\bar{O} \in \mathbb{N}^{\mathcal{S}}$ such upper bound, where each component \bar{O}_S gives the maximum abundance to be tracked for species S . For each discrete state $o \leq \bar{O}$, we denote by $\llbracket o \rrbracket$ the corresponding auxiliary species that is considered in the expansion. Thus we may define $\mathcal{S}_{\bar{O}}$ to be the set of species in the expanded network as $\mathcal{S}_{\bar{O}} = \mathcal{S} \cup \{\llbracket o \rrbracket \mid o \leq \bar{O}\}$. For example, in a network with the single reaction as in Eq. 2.2, let us choose $O_A = O_B = O_C = 1$. Then, the expanded network will have auxiliary species $\llbracket A \rrbracket$, $\llbracket B \rrbracket$, $\llbracket C \rrbracket$, $\llbracket A + B \rrbracket$, $\llbracket A + C \rrbracket$, $\llbracket B + C \rrbracket$, $\llbracket A + B + C \rrbracket$, and $\llbracket \mathbf{0} \rrbracket$, where the last species denotes the zero vector being tracked. We remark that, similarly to the definition of the master equation, it is convenient to consider all states within the upper bound when describing the theory. However, also in this case, not all discrete states may be reached depending on the stoichiometry of the reaction network, hence they can be removed in practice during the analysis.

FSE replaces each reaction $\rho \xrightarrow{f} \pi$ in the original network with a set of reactions for each tracked state $\llbracket o \rrbracket$ as follows:

$$\llbracket o \rrbracket + \eta \xrightarrow{f_o} \llbracket o' \rrbracket + \psi, \quad \text{for } o \leq \bar{O}, \quad (3.1)$$

where η and ψ are vectors of original species, $\llbracket o' \rrbracket$ is a target auxiliary species, and f_o is a modified propensity function that accounts for the fact that the reactant auxiliary species is $\llbracket o' \rrbracket$. Specifically, we have:

$$\begin{aligned} \eta &= \max(\mathbf{0}, \rho - o) \\ \psi &= \max(\mathbf{0}, \max(\mathbf{0}, o - \rho) + \pi - \bar{O}) \\ o' &= \min(\bar{O}, \max(\mathbf{0}, o - \rho) + \pi). \end{aligned} \quad (3.2)$$

Intuitively, for each original reaction as in Eq. 2.1, Eq. 3.1 conditions its dynamics with respect to $\llbracket o \rrbracket$ being the discrete state being tracked. Any expanded reaction maintains the same overall counts of reactants and products as the originating reaction, with a product tracked state $\llbracket o' \rrbracket$ that

results from the addition of products and removal of reactants within the upper bound \bar{O} ; η and ψ act as *buffer species* for populations that are not explicitly tracked. Finally, the propensity function f_o is derived from the original one f as

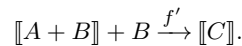
$$f_o: \mathbb{R}^{\mathcal{S}_{\bar{O}}} \rightarrow \mathbb{R}_0^+, \quad \text{with} \quad f_o(x) = x_{\llbracket o \rrbracket} f(o + x_{\setminus \llbracket o \rrbracket}). \quad (3.3)$$

This modification accounts for the fact that the tracked species $\llbracket o \rrbracket$ encodes additional population counts, as given by o .

For example, let us consider an expansion for the reaction in Eq. 2.2 assuming that it evolves with mass-action kinetics. In general, for a reaction with reagents ρ and kinetic parameter $k > 0$, the propensity function by mass-action kinetics for state σ is given by $f_k(\sigma) = k \prod_S (\rho_S^{\sigma_S})$. Here the propensity function is

$$f(x) = k x_A x_B (x_B - 1)/2.$$

Assuming that the upper bounds are $\bar{O}_A = \bar{O}_B = \bar{O}_C = 1$, the expansion for the tracked state $\llbracket A + B \rrbracket$ is given by



Since the tracked state $\llbracket A + B \rrbracket$ does not have enough copies of B , one further copy is used by the buffer species. The product of this reaction does not involve any of the buffer species because $\llbracket C \rrbracket$ is within the chosen bounds. By Eq. 3.3, f' in the expanded reaction is

$$\begin{aligned} f'(x) &= k x_{\llbracket A+B \rrbracket} (1 + x_A)(1 + x_B)((1 + x_B) - 1)/2 \\ &= k x_{\llbracket A+B \rrbracket} (1 + x_A)(1 + x_B)x_B/2. \end{aligned}$$

We denote by $\mathcal{R}_{\bar{O}}$ the full set of reactions in the expanded network.

Every expansion is stochastically equivalent to the original network, in the sense that there is a unique marginal probability distribution for the overall population of each species at every time point. This equivalence can be stated in the sense of ordinary lumpability for Markov chains [12]. That is, with the master equation we show that the probability of being in a state in the original reaction network equals the sum of the probabilities across all states in the expanded network with the same overall abundances for each species. This relation holds at all time points, provided that it is satisfied for the respective probability distributions at time 0.¹

Theorem 3.1. *Let P and \hat{P} denote the solutions of the master equation in the original and expanded network, respectively. Then, for all t ,*

$$\sum_{o+\xi=\sigma} \hat{P}_{\llbracket o \rrbracket+\xi}(0) = P_{\sigma}(0) \implies \sum_{o+\xi=\sigma} \hat{P}_{\llbracket o \rrbracket+\xi}(t) = P_{\sigma}(t).$$

By construction, if $\bar{O} = \mathbf{0}$ then the original and expanded networks coincide. The other limit case, when the auxiliary set of species contains all discrete states, corresponds to a fully expanded reaction network. In this case, the DRE of the expanded network corresponds to the master equation of the original network, hence no approximation occurs.

Theorem 3.2. *Consider a well-defined reaction network $(\mathcal{S}, \mathcal{R})$ and let $(\mathcal{S}_{\bar{O}}, \mathcal{R}_{\bar{O}})$ be its expansion. Let $X(t)$ be the DRE solution of the expanded network and $P(t)$ the solution of the master equation of the original network at time t . Then, the following hold:*

- (i) if $X_S(0) = 0$ then $X_S(t) = 0$ for all t and $S \in \mathcal{S}$;
- (ii) if $X_{\llbracket o \rrbracket}(0) = P_o(0)$ then $X_{\llbracket o \rrbracket}(t) = P_o(t)$, for all t and $o \in \mathbb{N}^{\mathcal{S}}$.

¹The proofs of the forthcoming theorems are available in the enclosed Appendix, Section A.

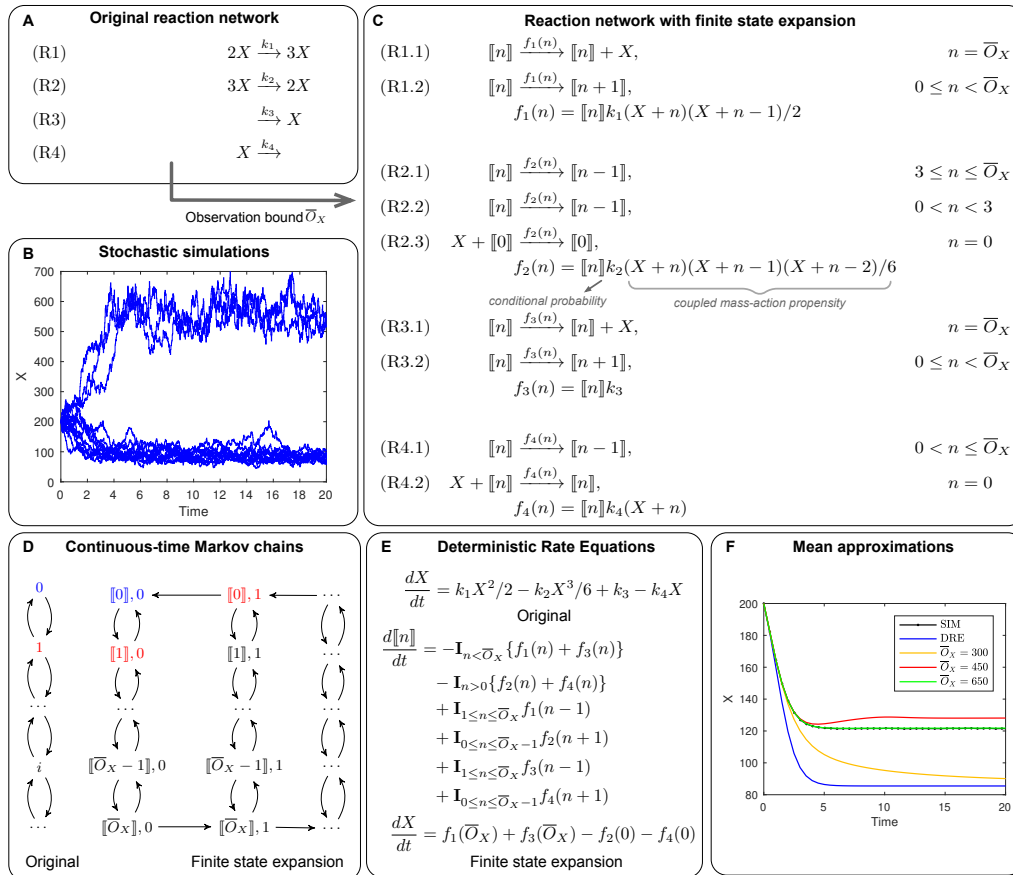


Figure 1: The FSE method applied to the Schlögl system [28]. (A) Mass-action reactions with kinetic parameters taken from ref. [29]: $k_1 = 0.03$, $k_2 = 0.0001$, $k_3 = 200$, $k_4 = 3.5$. (B) Stochastic simulations show the well-known bimodality of the steady-state probability distribution of species X . (C) For a given upper bound \bar{O}_X on the population of species X to be tracked explicitly, FSE yields the auxiliary species denoted by $\llbracket 0 \rrbracket, \llbracket 1 \rrbracket, \dots, \llbracket \bar{O}_X \rrbracket$. The original species X acts as buffer that collects untracked populations levels. For example, reaction R1.1 derives from reaction R1 when the autocatalytic formation of a new molecule occurs when the system tracks the discrete state $\llbracket \bar{O}_X \rrbracket$, thus requiring to increase the buffer species X by one element. Even when the system tracks a discrete state that does not require buffering (R1.2), the propensity function $f_1(n)$ of the reaction effectively considers an overall kinetics of mass-action type, since the factor $k_1(X+n)(X+n-1)/2$ models the total rate due to number of possible collisions between pairs of $X+n$ indistinguishable molecules. Intuitively, the factor $\llbracket n \rrbracket$ conditions these events to the system tracking n discrete molecules. (D) The original state space counts the number of copies of X . The state space in the expanded network consists of the pair tracked discrete state/population level of the buffer species. By Theorem 3.1, the sum of the probabilities across all pairs that have the same overall population matches the corresponding probability in the original Markov chain (as exemplified by matching colors of the states). (E) The single-dimensional DRE of the original Schlögl model is expanded into a DRE with $\bar{O}_X + 1$ variables (where \mathbf{I} denotes the indicator function); an estimate of the total mean population at time t can be computed as $X(t) + \sum_n n \cdot \llbracket n \rrbracket(t)$. (F) Starting from a population of 200 elements of X , the original bi-stable DRE converges to one equilibrium at 85.50 (blue line). FSE achieves excellent agreement with an upper bound $\bar{O}_X = 650$ (with respect to the average computed by stochastic simulation with 100 000 runs).

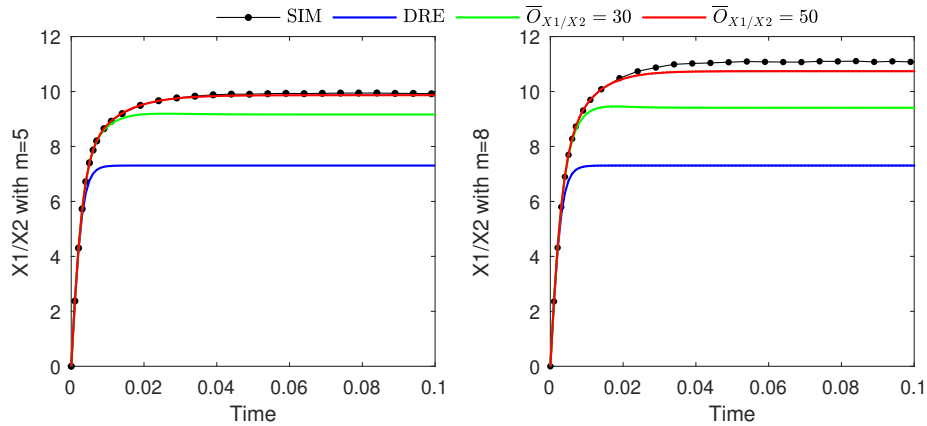


Figure 2: **Heterodimerization model (Eq. 4.1)**. Numerical simulations comparing stochastic simulation (300 000 runs), DRE and FSE for different upper bounds \bar{O}_{X_1/X_2} at burst sizes $m = 5$ (left) and $m = 8$ (right). The DRE approximation is unaffected by the variation of m while the true population averages are increased at higher burst sizes. Corrections for species X_3 are similar with a generally smaller error (cf. Fig. 12c). Kinetic parameters were set as follows: $k_1 = 2500/m$, $k_2 = 40$, $k_3 = 50$. For a burst size of $m = 5$, the parameters are as in Ref. [14]. Initial condition was the zero state.

4. Results

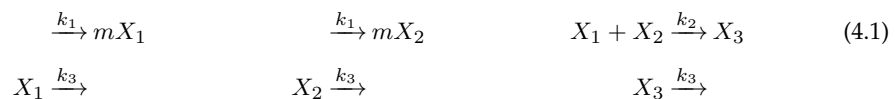
With a number of case studies from the literature, here we show that FSE can refine the accuracy of mean estimates of species populations, even with modest expansions. Ground-truth mean trajectories were computed by stochastic simulation via Gillespie's algorithm [2]. The numerical experiments herein reported were performed with an implementation of FSE within the tool ERODE [30], publicly available at <https://www.erode.eu>.

(a) Schlögl system

The well-known Schlögl system is an autocatalytic process for a single species X [28]. The DRE of the original Schlögl model has two equilibrium points, owing to its strong (cubic) nonlinearity [31], deterministically converging only to one [32]. Its discrepancy with respect to the average mean trajectory computed by stochastic simulation has been observed for a long time [33]. Fig. 1 provides a fully worked application of our FSE as a function of the upper bound for X . The solutions to the DRE of the expanded networks show that larger values of such upper bound increasingly improve the accuracy of the mean estimates.

(b) Heterodimerization model

We now consider a model from Ref. [14], where the main source of noise is the variance caused by the production of the two species X_1 and X_2 undergoing heterodimerization occurs in bursts of size m :



We study two cases with $m = 5$ and $m = 8$; for a better comparison the influx rates were kept constant by setting $k_1 = 2500/m$. Because of the symmetries between X_1 and X_2 we consider equal observation bounds for them. The observation bound on X_3 was set to zero. In this model,

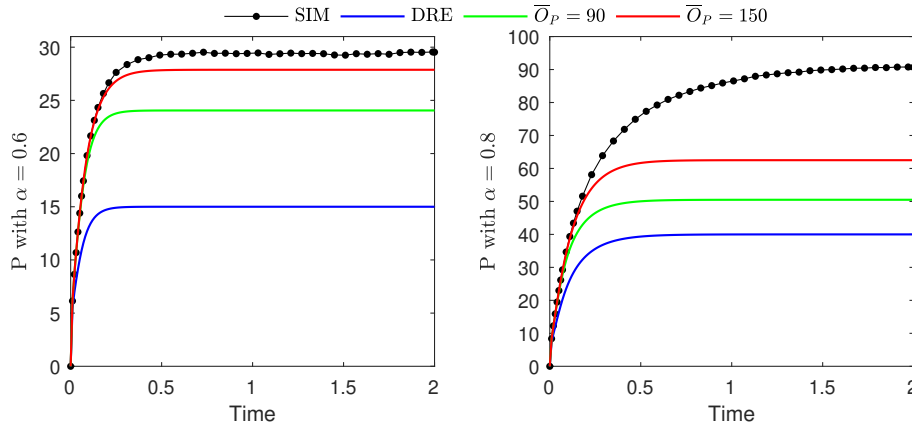
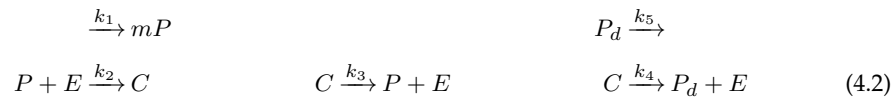


Figure 3: **Protein degradation model (Eq. 4.2)**. Numerical simulations comparing stochastic simulation (200 000 runs), DRE and FSE for different upper bounds \bar{O}_P at enzyme saturation levels $\alpha = 0.6$ (left) and $\alpha = 0.8$ (right). Enzyme observation bounds have been fixed at $\bar{O}_E = \bar{O}_C = 60$ to encompass the total population E_T . Kinetic parameters were set as follows: $k_2 = 4$, $k_3 = 3$, $k_4 = 37$, $k_5 = 10$, $k_1 = \alpha k_4 E_T / m = 74\alpha$. The initial state is $(P, E, C, P_d) = (0, 60, 0, 0)$.

the DRE is insensitive to the choice of m while the stochastic trajectories do depend on the burst size. Since a larger m introduces more noise, larger observation bounds are needed to increase the accuracy of the approximation (Fig. 2).

(c) Protein degradation model

The model of enzyme-catalysed protein degradation from [14] allows us to study the behavior of FSE when more species are to be tracked with nonzero population bounds than in the previous two case studies. Here, protein P is generated in bursts of size m and can bind to catalyst enzyme E , forming the enzyme-substrate complex C . When the protein unbinds from the enzyme, it can be degraded, forming P_d . The total amount E_T of catalyst enzyme in the system always remains constant:



By varying the burst production rate k_1 , different saturation levels of the catalyst enzyme are reached. The closer the ratio $\alpha = mk_1/k_4 E_T$ is to 1, the more saturated the enzyme becomes with substrate.

The parameters given in [14] for this model assume a burst size of $m = 30$ and a total enzyme population of $E_T = 60$. For an accurate approximation, FSE requires nonzero observation bounds for E , C , and P (Fig. 3); similarly to the previous case, no observation bound is required for species P_d . The size of the tracked state space grows proportionally to the product $\bar{O}_E \cdot \bar{O}_C \cdot \bar{O}_P$. Additionally, in contrast to the previous model, burst production rates are not adjusted when increasing the burst size here. The overall higher production rate increases the population additionally to the effect caused by higher variances. This exacerbates the need for higher observation bounds at higher burst sizes. For a fixed choice of the observation bounds, larger values of α tend to worsen the accuracy of the FSE approximation.

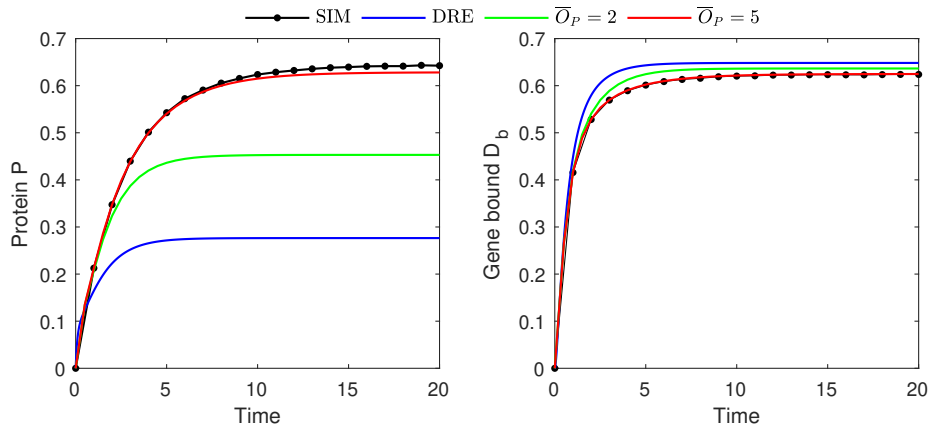
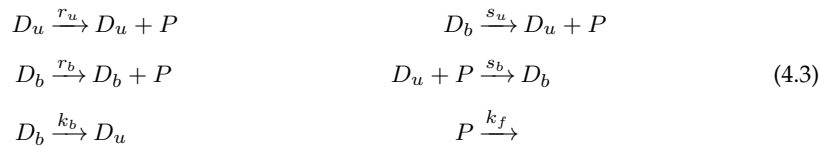


Figure 4: **Genetic feedback switch in Eq. 4.3.** Numerical simulations comparing stochastic simulation (1 000 000 runs), DRE and FSE for fixed $\bar{O}_{D_u} = \bar{O}_{D_b} = 1$ and different upper bounds \bar{O}_P . The resulting DRE from FSE has $2 \cdot \bar{O}_P + 2$ equations. Kinetic parameters were set as follows: $r_u = 1.0$, $r_b = 0.5$, $k_f = 0.1$, $k_b = 1.0$, $s_b = 10.0$, $s_u = 0.5$. The initial state is $(P, D_u, D_b) = (0, 1, 0)$.

(d) Feedback switch

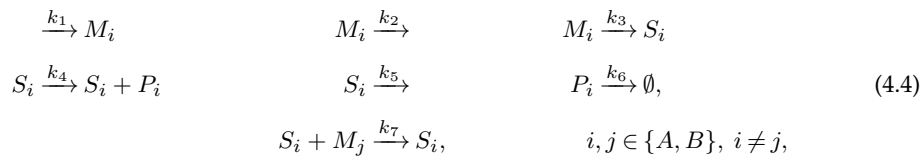
Let us now consider a model of a genetic feedback switch taken from [34] and [35]:



Species D_u and D_b represent the state of a single gene when its promoter region is unbound (respectively, bound) to a protein P . The reaction propensities obey mass-action dynamics through the kinetic parameters on the arrows. This is a basic model for negative autoregulation, a well-known motif appearing in more than 40% of the known transcription factors in *E.coli* [36]. A natural choice of upper bounds for the gene species is $\bar{O}_{D_u} = \bar{O}_{D_b} = 1$, by which the DRE of the expanded network can be interpreted as the solution of the conditional expectation of the protein population based on the gene state. Small values of \bar{O}_P yield a significant correction of the protein levels as well as of the marginal probability distribution of the gene state (Fig. 4).

(e) Toggle switch

The toggle switch network is a fundamental regulatory system of two mutually repressing genes [37]. Its mathematical modeling is challenging because of multimodality [24,38], as well as stochastic noise due to the species such as mRNA present in low molecular abundances [39]. Here we study the reaction scheme analyzed in [40], consisting of a mass-action variant from [37]:



where M_i and S_i denote the precursor mRNA and the mRNA for target protein P_i . The last two reactions model mutual inhibition by means of a precursor of one protein repressing the mRNA of the other.

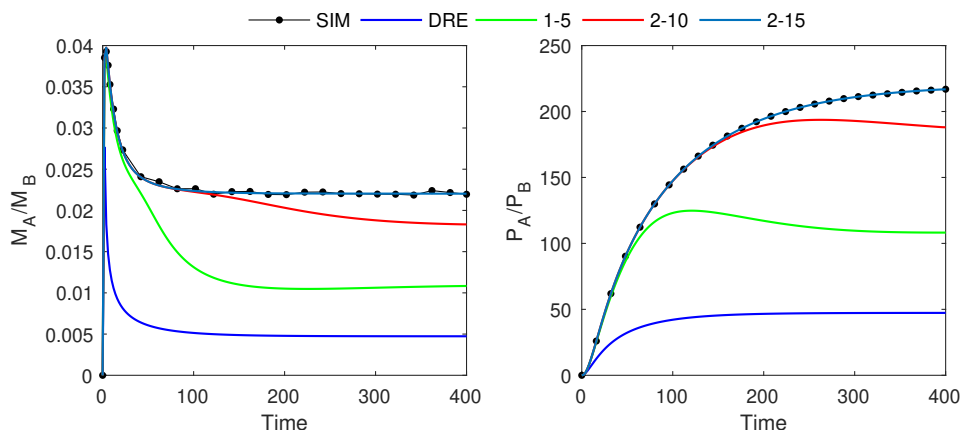


Figure 5: **Toggle switch (Eq. 4.4)**. Numerical simulations comparing stochastic simulation (500000 runs), DRE and FSE by fixing $\bar{O}_{P_A} = \bar{O}_{P_B} = 0$ while using different upper bounds \bar{O}_M and \bar{O}_S ($\bar{O}_M - \bar{O}_S$ in short) for the number of copies of M_A/M_B and S_A/S_B (as indicated in the legend), respectively. Initial condition was the zero state. The size of the DRE for the tested choices of upper bounds is equal to $(\bar{O}_M + 1)^2 \cdot (\bar{O}_S + 1)^2 + 6$ (corresponding to 150, 1095 and 2310 equations for $\bar{O}_M - \bar{O}_S = 1-5$, $\bar{O}_M - \bar{O}_S = 2-10$, and $\bar{O}_M - \bar{O}_S = 2-15$, respectively). Kinetic parameters were chosen as follows: $k_1 = 0.05$, $k_2 = 0.1$, $k_3 = 1.0$, $k_4 = 10.0$, $k_5 = 0.01$, $k_6 = 0.1$, $k_7 = 20.0$. Protein production (right plot) is controlled by a low population of precursor mRNA (left plot), which causes significant underestimation errors with DRE. Increasing the upper bounds of FSE improves the accuracy of the mean estimate. Corrections for species S_A and S_B , not reported here, are similar.

When protein production is controlled by low populations of precursor mRNA, the stochastic fluctuations are not adequately approximated by the DRE. By explicitly observing few copies of mRNA (up to tens), FSE provides precise estimates of the time courses of the mean populations (Fig. 5). The resulting systems of equations, of size at most 2310, can be analyzed effectively, as opposed to time-consuming simulations using hybrid stochastic-deterministic approaches such as in [40].

5. Comparison with related work

Using the models presented in the previous section, here we compare FSE with state-of-the-art analytical techniques that can be used to obtain approximate estimates of mean population levels in stochastic reaction networks. Specifically, we considered the following methods:

- Moment-closure approximation (MCA). We considered the second-order low-dispersion moment closure [41,42], in which variance and covariance are the highest observed moments and all higher-order central moments are set to zero; in all models considered in this paper, computing approximations with higher-order moments did not improve the quality of the approximation.
- The effective mesoscopic rate equation (EMRE), which adds mean-correction terms to the linear-noise approximation under the assumption of an underlying Gaussian process [14].
- The method of conditional moments (MCM), a hybrid analytical technique combining a discrete representation of low-abundance species and a moment-based approximation of high-abundance ones [16].

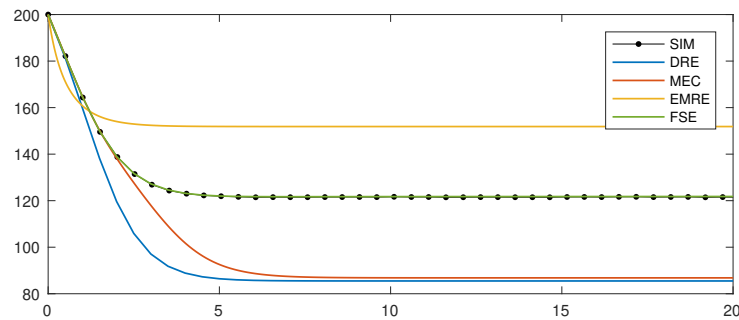


Figure 6: **Comparison with related techniques on the Schlögl model.** The average population of X computed by stochastic simulation (100000 runs) is compared against DRE, MCA, EMRE and FSE with $\bar{O}_X = 650$.

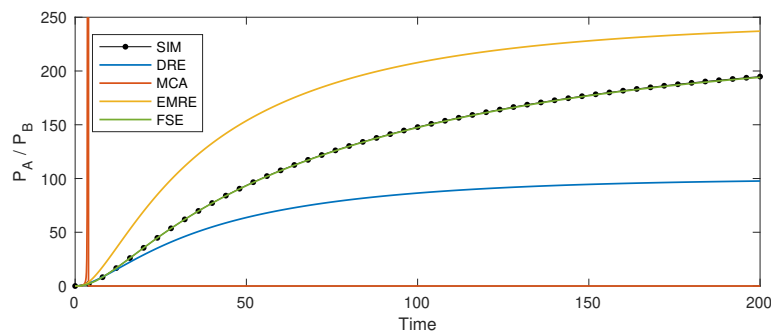


Figure 7: **Comparison with related techniques on the toggle switch model.** Stochastic simulation to compute the average populations of species P_A/P_B (500 000 runs) is compared against DRE, MCA, EMRE, and FSE. FSE is run with upper bounds $\bar{O}_P = 0$, $\bar{O}_M = 2$ and $\bar{O}_S = 10$. MCA estimates population levels approaching 75 000 (out of scale in this plot to improve readability) before dropping to zero.

- Finite state projection (FSP), a well-known method to obtain a finite-dimensional master equation through a truncation of the state space by redirecting transitions toward unobserved states into an absorbing state with provable bounds [43].

For this study, we used an implementation of these techniques as available on the software tool CERENA [41].

The Schlögl system is known to stress MCA because of their reported difficulties with multimodal distributions [44,45]. Fig. 6 shows that MCA behaves similarly to DRE in this case, while EMRE tends to overestimate the mean population of species X at longer time horizons. Similar results were obtained on the toggle switch network (Fig. 7). Here we confirm physically meaningless moment-closure estimates due to the presence of low-abundance species, as already reported in [22]. Furthermore, MCM could not be tested on this model since its implementation returned with an error.

Both MCA and EMRE provide accurate estimates in the heterodimerization model from Eq. 4.1 (Fig. 8). This model also allows examining the effects of burst size variation without also changing the total production rate. We used this model for a closer comparison to EMRE in this regard. Figure 9 shows that, similarly to FSE (with fixed observation bounds) and the DRE

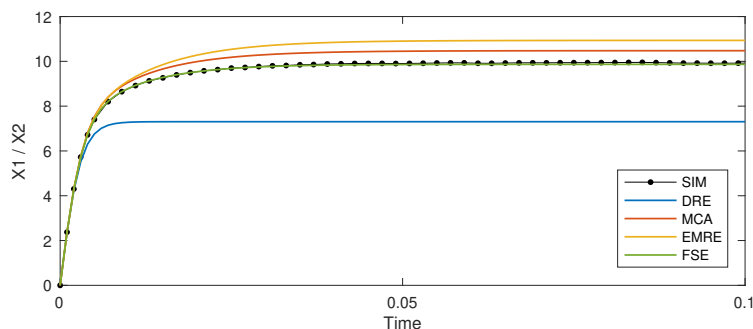


Figure 8: **Comparison with related techniques on the heterodimerization model with burst size $m = 5$.** Stochastic simulation to compute the average populations of species X_1/X_2 (300 000 runs) is compared against DRE, MCA, EMRE, FSE using upper bounds $\bar{O}_{X_1} = \bar{O}_{X_2} = 50$ and $\bar{O}_{X_3} = 0$.

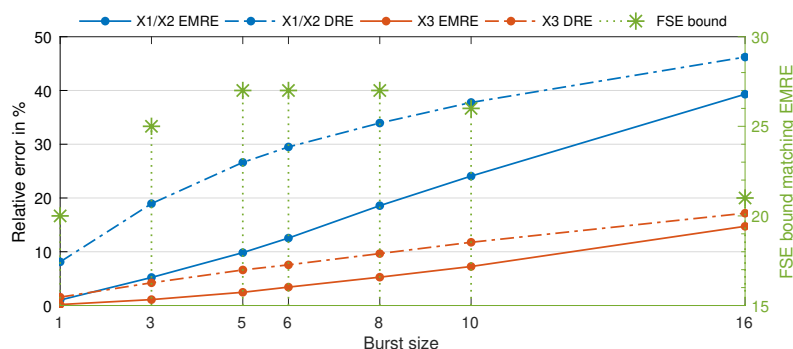


Figure 9: **Direct comparison against EMRE on the heterodimerization model across different burst sizes.** Although it is consistently better than the DRE (broken line, shown for reference) for all tested values, the relative error of EMRE to the stochastic simulation rises on all species with increasing burst size. In green, the graph shows for each burst size the FSE bound \bar{O}_{X_1/X_2} that is necessary to match the accuracy of EMRE.

approximation, the approximation error of EMRE increases with the burst size m . For each tested burst size, Figure 9 also marks (secondary y -axis) the minimum FSE observation bound \bar{O}_{X_1/X_2} for which the accuracy of EMRE can be matched. In this study, the highest observation bound $\bar{O}_{X_1/X_2} = 27$ is enough to match EMRE's accuracy in the range $5 \leq m \leq 8$. For burst sizes larger than 8, the loss of accuracy in EMRE outpaces the analogous effect in FSE, to the degree that the bound to match EMRE becomes small. The accuracy of FSE can be increased further by raising the bound above the marked values (Fig. 12c).

The main challenge of the protein degradation model in Eq. 4.2 for FSE is the rapid growth of the state space as a function of the observation bounds. Using the observation bounds as in Fig. 3, FSE is outperformed by both EMRE and MCA for $\alpha = 0.8$. For $\alpha = 0.6$, FSE is more accurate than EMRE; however, it uses a significantly more complex system with 9214 differential equations.

In the genetic feedback switch model, species D_u and D_b describe the distinct binary states of a single gene. Hence they represent the natural candidates of the low-abundance class when applying MCM. On this model, however, the method could not return valid results as early as time point 0.36. We further tested a gene regulatory model with an inhibition feedback loop taken from [46], where it was studied using a hybrid stochastic/deterministic method based on piecewise deterministic Markov processes. Here, MCM showed similar difficulties that confirm

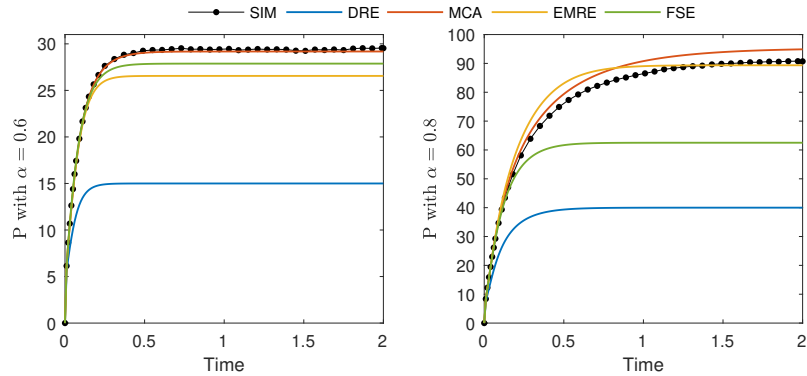


Figure 10: **Comparison with related techniques on the protein degradation model.** Stochastic simulation to compute the average populations of species P (200 000 runs) is compared against DRE, MCA, EMRE and FSE for two different catalyst saturation levels α . FSE is run with upper bounds $\bar{O}_P = 150, \bar{O}_E = \bar{O}_C = 60$ and $\bar{O}_{P_d} = 0$.

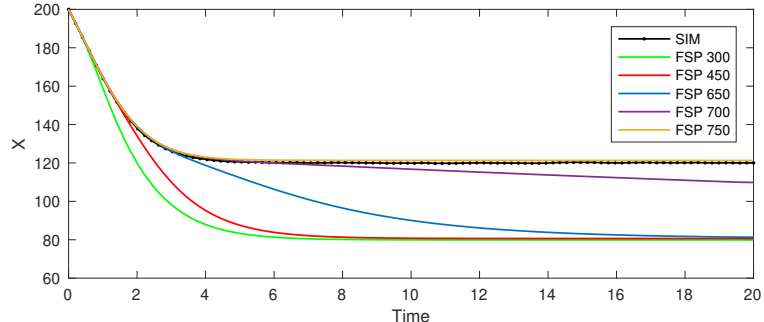


Figure 11: **Analysis by FSP of the Schlögl system.** With parameter settings as in Fig. 1, for state spaces of equal size as in Fig. 1F (450 and 650), FSE estimates the average population of X more precisely than finite state projection (stochastic simulation performed with 100 000 runs). In particular, while FSE requires a state space with 650 states to accurately match stochastic simulations, FSP needs 750 states.

already reported numerical issues [3]. (The numerical results of this analysis, not shown here, are replicable using the supporting data of this article.) Stochastic models of gene expression such as these are amenable to domain-specific techniques (see, e.g., [25] and references therein), which may prove more effective. For example, in the enclosed Appendix B we compare FSE against the linear mapping approximation method from [25], showing that FSE requires larger systems of differential equations to obtain a similar level of accuracy.

Defining incoming and outgoing transitions with respect to the buffer species maintained in the expanded network represents a crucial difference with FSP, where transitions toward unobserved state are collapsed into a sink state that absorbs the probability mass. Experimentally, this results in increased accuracy of mean estimates by FSE when tracking the same subset of the state space in both methods. For example, Figure 11 shows this effect on the Schlögl model; a similar behavior could be found in the other case studies (not reported here). More recently, variants of FSP have been proposed to cope with the decay of the probability mass, e.g., to obtain solutions that are able to estimate stationary distributions and passage times [47–49]. By means of a comparison against the most recent improvement by [49], in the enclosed Appendix (Section C) we show that FSE may provide accurate estimates using systems of fewer equations.

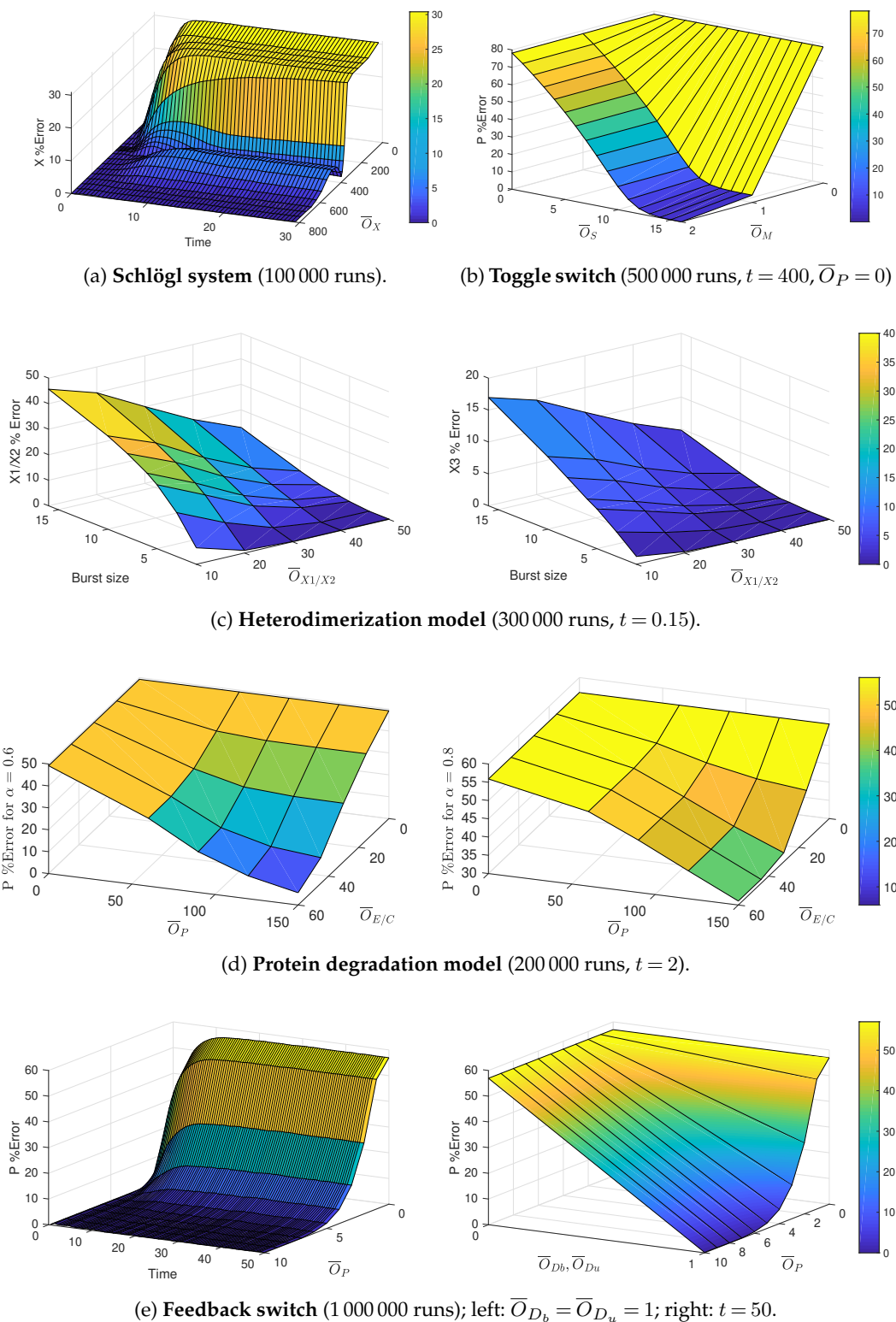


Figure 12: Sensitivity analysis. Approximation error is measured as percentage error between the FSE estimate and the average by stochastic simulation. The reported time points are representative of steady state.

The solution by the aforementioned state-space truncation methods is a lower bound on the true probability distribution, and increasing the set of observed states tightens that bound. Instead, although FSE ensures that the expansion coincides with the master equation when the whole state space is tracked, it does not give theoretical guarantees on the degree of accuracy, nor does it guarantee monotonically increasing accuracy with larger observation bounds. Indeed, experimentally we confirmed that monotonicity of the error is model dependent. For instance, the relative percentage error between the mean population predicted by FSE and the estimated mean by stochastic simulation is not monotonic in the Schlögl system (Fig. 12a), while it is monotonic for the other case studies (Fig. 12b-e).

6. Conclusion

We have presented finite state expansion (FSE) as a novel analytical method that offers a trade off between the exactness of the solution of the master equation and the approximation errors introduced by the deterministic rate equation (DRE) for stochastic reaction networks. FSE maintains a user-defined subset of the discrete state space and couples this with whole-population continuous dynamics to account for the behavior of states that are not explicitly tracked. By an algorithmic translation of a network into an expanded one with auxiliary species and modified reactions, FSE leads to equations that can be interpreted as a refinement of the original DRE. A theoretical result of asymptotic correctness increases the confidence as to the effectiveness of the method, since it shows that the DRE of the expanded network corresponds to the original master equation.

The performance of FSE in correcting the original DRE when tracking a strict subset of the discrete state space has been shown numerically in models that turn out to be challenging for related state-of-the-art techniques. The effective mesoscopic rate equation relies on perturbation arguments around the linear-noise approximation, hence it inherently assumes a limiting regime, unlike FSE. Experimentally, we found that this resulted in less accurate mean estimates than FSE. With respect to analytical approximations of the master equation based on moment closure, the case studies proved difficult since the analyses returned unphysical results or exhibited numerical issues, as also reported in the literature. Unlike state-space truncation methods based on finite state projection, FSE does not provide error bounds with respect to the original master equation. Yet, we found excellent accuracy when the observed state space is large enough. Overall, these findings make FSE a useful tool to study chemical reaction networks for which exact stochastic analysis through the master equation is not accessible. Currently, the FSE translation provides a refinement of mean estimates. It is an interesting line of future work to extend the method to compute improved approximations of higher-order moments.

Despite the encouraging results herein reported, the applicability of FSE may not always be feasible. Since it is based on an enumeration of the discrete state space, FSE may too suffer from combinatorial complexity, such that the number of equations can grow rapidly large as a function of the observation bound. For larger networks in particular, this may require relatively small tracked state spaces to keep reasonable computational costs. On the other hand, if significant probability mass falls outside the tracked state space then the performance of FSE may not be adequate, as the Schlögl system shows when small enough bounds are used.

There are a number of methods that are worth investigating in the future in order to tackle these challenges. Model-reduction techniques could help relieve the computational cost of the analysis of the DRE by providing a lower-order approximation that preserves the dynamics of interest [50,51]. It would be particularly beneficial to study either bounds or monotonicity properties on the FSE approximation error in order to develop adaptive strategies to find optimal values for the observation bounds. In principle, there might be other expansions than the one presented here, which give rise to different correction behavior of the DRE while still preserving the stochastic dynamics. A further line of improvement might consist in devising variants of FSE where the tracked state space can be arbitrarily fixed, instead of being dependent on an upper bound for the population counts. This might allow the fine-tuning of the choice of the

discrete region where the probability mass is mostly concentrated. For such expansions, smaller observation bounds (hence lower computational cost) may suffice to obtain the same degree of accuracy as in this paper, thus potentially extending the practical applicability of FSE to more complex models.

Ethics. This article does not present research with ethical considerations.

Data Accessibility. Software implementation and all material to replicate the experiments herein reported are accessible from <https://zenodo.org/record/4739329>.

Authors' Contributions. TW participated in the design of the study, carried out research, and drafted the manuscript; LB participated in the design of the study, helped draft and critically revise the manuscript; AV helped carry out the research and critically revised the manuscript; MT conceived the study, designed the study, coordinated the study and helped draft the manuscript. All authors gave final approval for publication and agree to be held accountable for the work performed therein.

Competing Interests. We declare we have no competing interests.

Funding. This work has been partially supported by the Italian Ministry for Research (PRIN project "SEDUCE", no. 2017TWRCN).

References

1. Van Kampen NG. 2007 *Stochastic Processes in Physics and Chemistry*. Elsevier.
2. Gillespie D. 1977 Exact Stochastic Simulation of Coupled Chemical Reactions. *Journal of Physical Chemistry* **81**, 2340–2361.
3. Schnoerr D, Sanguinetti G, Grima R. 2017 Approximation and inference methods for stochastic biochemical kinetics—a tutorial review. *Journal of Physics A: Mathematical and Theoretical* **50**, 093001.
4. Fröhlich F, Thomas P, Kazeroonian A, Theis FJ, Grima R, Hasenauer J. 2016 Inference for Stochastic Chemical Kinetics Using Moment Equations and System Size Expansion. *PLoS Computational Biology* **12**, 1–28.
5. MacNamara S, Burrage K. 2009 Krylov and steady-state techniques for the solution of the chemical master equation for the mitogen-activated protein kinase cascade. *Numerical Algorithms* **51**, 281–307.
6. Grima R. 2015 Linear-noise approximation and the chemical master equation agree up to second-order moments for a class of chemical systems. *Phys. Rev. E* **92**, 042124.
7. Kurtz TG. 1972 The Relationship between Stochastic and Deterministic Models for Chemical Reactions. *The Journal of Chemical Physics* **57**, 2976–2978.
8. Guptasarma P. 1995 Does replication-induced transcription regulate synthesis of the myriad low copy number proteins of *Escherichia coli*?. *BioEssays* **17**, 987–997.
9. Elowitz MB, Levine AJ, Siggia ED, Swain PS. 2002 Stochastic Gene Expression in a Single Cell. *Science* **297**, 1183–1186.
10. Paulsson J. 2005 Models of stochastic gene expression. *Physics of Life Reviews* **2**, 157–175.
11. Swain PS, Elowitz MB, Siggia ED. 2002 Intrinsic and extrinsic contributions to stochasticity in gene expression. *Proceedings of the National Academy of Sciences* **99**, 12795–12800.
12. Buchholz P. 1994 Exact and Ordinary Lumpability in Finite Markov Chains. *Journal of Applied Probability* **31**, 59–75.
13. Kuehn C. 2016 Moment Closure—A Brief Review. In Schöll E, Klapp SHL, Hövel P, editors, *Control of Self-Organizing Nonlinear Systems* pp. 253–271. Springer International Publishing.
14. 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. *The Journal of Chemical Physics* **133**, 035101.
15. Thomas P, Matuschek H, Grima R. 2012 Computation of biochemical pathway fluctuations beyond the linear noise approximation using iNA. In *IEEE International Conference on Bioinformatics and Biomedicine* pp. 1–5.
16. Hasenauer J, Wolf V, Kazeroonian A, Theis FJ. 2014 Method of conditional moments (MCM) for the Chemical Master Equation. *Journal of Mathematical Biology* **69**, 687–735.
17. Gómez-Urbe CA, Verghese GC. 2007 Mass fluctuation kinetics: Capturing stochastic effects in systems of chemical reactions through coupled mean-variance computations. *The Journal of Chemical Physics* **126**, 024109.

18. Ale A, Kirk P, Stumpf MPH. 2013 A general moment expansion method for stochastic kinetic models. *The Journal of Chemical Physics* **138**, 174101.
19. Lee CH, Kim KH, Kim P. 2009 A moment closure method for stochastic reaction networks. *The Journal of Chemical Physics* **130**, 134107.
20. Sotiropoulos V, Kaznessis YN. 2011 Analytical derivation of moment equations in stochastic chemical kinetics. *Chemical Engineering Science* **66**, 268 – 277.
21. Gillespie C. 2009 Moment-closure approximations for mass-action models. *IET Systems Biology* **3**, 52–58.
22. Singh A, Hespanha JP. 2011 Approximate Moment Dynamics for Chemically Reacting Systems. *IEEE Transactions on Automatic Control* **56**, 414–418.
23. Smadbeck P, Kaznessis YN. 2013 A closure scheme for chemical master equations. *Proceedings of the National Academy of Sciences* **110**, 14261–14265.
24. Thomas P, Popović N, Grima R. 2014 Phenotypic switching in gene regulatory networks. *Proceedings of the National Academy of Sciences* **111**, 6994–6999.
25. Cao Z, Grima R. 2018 Linear mapping approximation of gene regulatory networks with stochastic dynamics. *Nature Communications* **9**, 3305.
26. Jahnke T. 2011 On Reduced Models for the Chemical Master Equation. *Multiscale Modeling & Simulation* **9**, 1646–1676.
27. Menz S, Latorre J, Schütte C, Huisinga W. 2012 Hybrid Stochastic–Deterministic Solution of the Chemical Master Equation. *SIAM Interdisciplinary Journal Multiscale Modeling and Simulation* **10**, 1232–1262.
28. Schlögl F. 1972 Chemical reaction models for non-equilibrium phase transitions. *Zeitschrift für Physik* **253**, 147–161.
29. Li H, Cao Y, Petzold LR, Gillespie DT. 2008 Algorithms and Software for Stochastic Simulation of Biochemical Reacting Systems. *Biotechnology Progress* **24**, 56–61.
30. Cardelli L, Tribastone M, Vandin A, Tschaikowski M. 2017 ERODE: A Tool for the Evaluation and Reduction of Ordinary Differential Equations. In *Tools and Algorithms for the Construction and Analysis of Systems — 23rd International Conference, TACAS*.
31. Bishop LM, Qian H. 2010 Stochastic Bistability and Bifurcation in a Mesoscopic Signaling System with Autocatalytic Kinase. *Biophysical Journal* **98**, 1–11.
32. Vellela M, Qian H. 2009 Stochastic dynamics and non-equilibrium thermodynamics of a bistable chemical system: the Schlögl model revisited. *Journal of The Royal Society Interface* **6**, 925–940.
33. Zheng Q, Ross J. 1991 Comparison of deterministic and stochastic kinetics for nonlinear systems. *The Journal of Chemical Physics* **94**, 3644–3648.
34. Hornos JEM, Schultz D, Innocentini GCP, Wang J, Walczak AM, Onuchic JN, Wolynes PG. 2005 Self-regulating gene: An exact solution. *Phys. Rev. E* **72**, 051907.
35. Grima R, Schmidt DR, Newman TJ. 2012 Steady-state fluctuations of a genetic feedback loop: An exact solution. *The Journal of Chemical Physics* **137**, 035104.
36. Shen-Orr SS, Milo R, Mangan S, Alon U. 2002 Network motifs in the transcriptional regulation network of *Escherichia coli*. *Nature Genetics* **31**, 64–68.
37. Gardner TS, Cantor CR, Collins JJ. 2000 Construction of a genetic toggle switch in *Escherichia coli*. *Nature* **403**, 339–342.
38. Tian T, Burrage K. 2006 Stochastic models for regulatory networks of the genetic toggle switch. *Proceedings of the National Academy of Sciences* **103**, 8372–8377.
39. Kærn M, Elston TC, Blake WJ, Collins JJ. 2005 Stochasticity in gene expression: from theories to phenotypes. *Nature Reviews Genetics* **6**, 451 EP –.
40. Hepp B, Gupta A, Khammash M. 2015 Adaptive hybrid simulations for multiscale stochastic reaction networks. *The Journal of Chemical Physics* **142**, 034118.
41. Kazerooni A, Fröhlich F, Raue A, Theis FJ, Hasenauer J. 2016 CERENA: ChEmical REaction Network Analyzer—A Toolbox for the Simulation and Analysis of Stochastic Chemical Kinetics. *PLOS ONE* **11**, 1–15.
42. Schnoerr D, Sanguinetti G, Grima R. 2015 Comparison of different moment-closure approximations for stochastic chemical kinetics. *The Journal of Chemical Physics* **143**, 185101.
43. Munsky B, Khammash M. 2006 The finite state projection algorithm for the solution of the chemical master equation. *The Journal of Chemical Physics* **124**, 044104.
44. Lakatos E, Ale A, Kirk PDW, Stumpf MPH. 2015 Multivariate moment closure techniques for stochastic kinetic models. *The Journal of Chemical Physics* **143**, 094107.

45. Schnoerr D, Sanguinetti G, Grima R. 2014 Validity conditions for moment closure approximations in stochastic chemical kinetics. *The Journal of Chemical Physics* **141**, 084103.
46. Winkelmann S, Schütte C. 2017 Hybrid models for chemical reaction networks: Multiscale theory and application to gene regulatory systems. *The Journal of Chemical Physics* **147**, 114115.
47. Cao Y, Terebus A, Liang J. 2016 Accurate Chemical Master Equation Solution Using Multi-Finite Buffers. *Multiscale Modeling & Simulation* **14**, 923–963.
48. Gupta A, Mikelson J, Khammash M. 2017 A finite state projection algorithm for the stationary solution of the chemical master equation. *The Journal of Chemical Physics* **147**, 154101.
49. Kim J, Dark J, Enciso G, Sindi S. 2020 Slack reactants: A state-space truncation framework to estimate quantitative behavior of the chemical master equation. *The Journal of Chemical Physics* **153**, 054117.
50. Cardelli L, Tribastone M, Tschaikowski M, Vandin A. 2017 Maximal aggregation of polynomial dynamical systems. *Proceedings of the National Academy of Sciences* **114**, 10029–10034.
51. Kim JK, Sontag ED. 2017 Reduction of multiscale stochastic biochemical reaction networks using exact moment derivation. *PLOS Computational Biology* **13**, 1–24.

A. Proofs of statements

This appendix provides accompanying material for the main text. This section shows the proofs of Theorems 3.1 and 3.2 in the main text. Section C presents the comparison of finite state expansion (FSE) against the method of slack reactants [49].

Notation. We define the following operations for any two $\sigma, \mu \in \mathbb{R}^{\mathcal{S}}$.

- Minimum: $\sigma \wedge \mu$ is such that $(\sigma \wedge \mu)_S = \min(\sigma_S, \mu_S)$ for all $S \in \mathcal{S}$.
- Saturated subtraction: $\sigma \ominus \mu$ is such that $(\sigma \ominus \mu)_S = \max(0, \sigma_S - \mu_S)$ for all $S \in \mathcal{S}$.
- Projection: Given $\mathcal{P} \subseteq \mathcal{S}$, $\sigma_{|\mathcal{P}} \in \mathbb{R}^{\mathcal{P}}$ is such that $(\sigma_{|\mathcal{P}})_P = \sigma_P$ for all $P \in \mathcal{P}$.
- Mapping: Given $\mathcal{P} \subseteq \mathcal{S}$, and a function $m: \mathcal{S} \rightarrow \mathcal{P}$, $\sigma^m \in \mathbb{R}^{\mathcal{P}}$ is such that $\sigma_P^m = \sum_{m(S)=P} \sigma_S$, for all $P \in \mathcal{P}$.

With these, the quantities in Eq. 3.2 in the main text can be rewritten as follows:

$$\eta = \rho \ominus o, \quad \psi = ((o \ominus \rho) + \pi) \ominus \bar{O}, \quad o' = \bar{O} \wedge ((o \ominus \rho) + \pi).$$

(a) Proof of Theorem 3.1

Let P and \hat{P} denote the solutions of the master equation in the original and expanded network, respectively. Then

$$\sum_{o+\xi=\sigma} \hat{P}_{[o]+\xi}(0) = P_{\sigma}(0) \implies \sum_{o+\xi=\sigma} \hat{P}_{[o]+\xi}(t) = P_{\sigma}(t),$$

for all t .

Proof. We prove the following equivalence for the derivatives of the solutions of the respective master equations

$$\sum_{o+\xi=\sigma} \frac{d\hat{P}_{[o]+\xi}}{dt} = \frac{dP_{\sigma}}{dt} \quad \text{for all } \sigma \in \mathbb{N}^{\mathcal{S}},$$

from which the statement holds under the assumption of consistent initial conditions.

$$\begin{aligned}
& \sum_{o+\xi=\sigma} \frac{d\hat{P}_{\llbracket o \rrbracket + \xi}}{dt} = \\
&= \sum_{o+\xi=\sigma} \sum_{(\llbracket \epsilon \rrbracket + \eta) \xrightarrow{f_\epsilon} (\llbracket o' \rrbracket + \psi) \in \mathcal{R}_{\overline{\sigma}}} \left(f_\epsilon(\llbracket o \rrbracket + \xi + \llbracket \epsilon \rrbracket + \eta - \llbracket o' \rrbracket - \psi) \hat{P}_{\llbracket o \rrbracket + \xi + \llbracket \epsilon \rrbracket + \eta - \llbracket o' \rrbracket - \psi} - f_\epsilon(\llbracket o \rrbracket + \xi) \cdot \hat{P}_{\llbracket o \rrbracket + \xi} \right) \\
&= \sum_{o+\xi=\sigma} \left(\sum_{(\llbracket \epsilon \rrbracket + \eta) \xrightarrow{f_\epsilon} (\llbracket o \rrbracket + \psi) \in \mathcal{R}_{\overline{\sigma}}} f_\epsilon(\llbracket \epsilon \rrbracket + \xi + \eta - \psi) \cdot \hat{P}_{\llbracket \epsilon \rrbracket + \xi + \eta - \psi} - \sum_{(\llbracket o \rrbracket + \eta) \xrightarrow{f_o} (\llbracket o' \rrbracket + \psi) \in \mathcal{R}_{\overline{\sigma}}} f_o(\llbracket o \rrbracket + \xi) \cdot \hat{P}_{\llbracket o \rrbracket + \xi} \right) \\
&= \sum_{\substack{(\llbracket \epsilon \rrbracket + \eta) \xrightarrow{f_\epsilon} (\llbracket o \rrbracket + \psi) \in \mathcal{R}_{\overline{\sigma}} \\ o+\xi=\sigma}} f_\epsilon(\llbracket \epsilon \rrbracket + \xi + \eta - \psi) \cdot \hat{P}_{\llbracket \epsilon \rrbracket + \xi + \eta - \psi} - \sum_{\substack{(\llbracket o \rrbracket + \eta) \xrightarrow{f_o} (\llbracket o' \rrbracket + \psi) \in \mathcal{R}_{\overline{\sigma}} \\ o+\xi=\sigma}} f_o(\llbracket o \rrbracket + \xi) \cdot \hat{P}_{\llbracket o \rrbracket + \xi} \\
&= \sum_{\substack{(\llbracket \epsilon \rrbracket + \eta) \xrightarrow{f_\epsilon} (\llbracket o \rrbracket + \psi) \in \mathcal{R}_{\overline{\sigma}} \\ \epsilon + \xi + \eta - \psi = \sigma - (o + \psi) + (\epsilon + \eta)}} f(\epsilon + \xi + \eta - \psi) \cdot \underbrace{(\llbracket \epsilon \rrbracket + \xi + \eta - \psi)_{\llbracket \epsilon \rrbracket}}_{=1} \cdot \hat{P}_{\llbracket \epsilon \rrbracket + \xi + \eta - \psi} + \\
&\quad - \sum_{\substack{(\llbracket o \rrbracket + \eta) \xrightarrow{f_o} (\llbracket o' \rrbracket + \psi) \in \mathcal{R}_{\overline{\sigma}} \\ o+\xi=\sigma}} f(o + \xi) \cdot \underbrace{(\llbracket o \rrbracket + \xi)_{\llbracket o \rrbracket}}_{=1} \cdot \hat{P}_{\llbracket o \rrbracket + \xi} \\
&= \sum_{\substack{\rho \xrightarrow{f} \pi \in \mathcal{R} \\ \epsilon + \xi + \eta - \psi = \sigma - \pi + \rho}} f(\epsilon + \xi + \eta - \psi) \cdot \hat{P}_{\llbracket \epsilon \rrbracket + \xi + \eta - \psi} - \sum_{\substack{\rho \xrightarrow{f} \pi \in \mathcal{R} \\ o+\xi=\sigma}} f(\sigma) \cdot \hat{P}_{\llbracket o \rrbracket + \xi} \\
&= \sum_{\rho \xrightarrow{f} \pi \in \mathcal{R}} f(\sigma - \pi + \rho) \cdot P_{\sigma - \pi + \rho} - \sum_{\rho \xrightarrow{f} \pi \in \mathcal{R}} f(\sigma) \cdot P_\sigma \\
&= \sum_{\rho \xrightarrow{f} \pi \in \mathcal{R}} \left(f(\sigma - \pi + \rho) \cdot P_{\sigma - \pi + \rho} - f(\sigma) \cdot P_\sigma \right) \\
&= \frac{dP_\sigma}{dt}.
\end{aligned}$$

□

Now we consider the other limit case, namely when the auxiliary set of species contains all discrete states, corresponding to a fully expanded reaction network. In this case, the DRE of the expanded network corresponds to the master equation of the original network, hence no approximation occurs.

(b) Proof of Theorem 3.2

In order to prove Theorem 3.2 in the main text, we prove two preliminary results stated as lemmata.

Lemma A.1

Lemma A.1. *The expansion of a well-defined reaction network is well-defined.*

Proof. Let $\rho \xrightarrow{f} \pi$ be a reaction of a well-defined network and $\llbracket o \rrbracket + \eta \xrightarrow{f_o} \llbracket o' \rrbracket + \psi$ its expansion according to the main text. Let us take $z \in \mathbb{N}^{\mathcal{S}_{\overline{\sigma}}}$ such that $(\llbracket o \rrbracket + \eta) \not\leq z$ and separately consider the two cases for which this holds. If $\llbracket o \rrbracket \notin z$, then propensity function f_o in the expanded reaction is equal to zero by definition, keeping with the requirement for the reaction being well-defined.

If $\eta \not\subseteq z$, then we have that $(\rho \ominus o) \not\subseteq z|_{\mathcal{S}}$ as ρ and o are both members of $\mathbb{N}^{\mathcal{S}}$. This implies that $\rho \not\subseteq (o + z|_{\mathcal{S}})$, and since, the reaction is well-defined we have that $f(o + z|_{\mathcal{S}}) = 0$, from which $f_o(z) = 0$. \square

Lemma A.2

The following lemma proves that the expansion preserves the overall population jumps. That is, for each original reaction, every expanded reaction is such that each species is subject to the same change of its abundance level.

Lemma A.2. *Let $\rho \xrightarrow{f} \pi$ be a reaction and $[[o]] + \eta \xrightarrow{f \circ} [[o']] + \psi$ its expansion according to The main text. Then it holds that:*

- (i) $(o + \eta) \ominus (o' + \psi) = \rho \ominus \pi$;
- (ii) $(o' + \psi) \ominus (o + \eta) = \pi \ominus \rho$;
- (iii) $\sigma \ominus (o + \eta) + (o' + \psi) = \sigma \ominus \rho + \pi$, for all $\sigma \in \mathbb{N}^{\mathcal{S}}$ such that $(o + \eta) \subseteq \sigma$.

Proof. For case (1):

$$\begin{aligned} (o + \eta) \ominus (o' + \psi) &= (o + (\rho \ominus o)) \ominus ((\overline{O} \wedge (o \ominus \rho + \pi)) \\ &\quad + ((o \ominus \rho + \pi) \ominus \overline{O})) \\ &= (o + (\rho \ominus o)) \ominus (o \ominus \rho + \pi) \\ &= (\rho + (o \ominus \rho)) \ominus ((o \ominus \rho) + \pi) \\ &= \rho \ominus \pi. \end{aligned}$$

For case (2):

$$\begin{aligned} (o' + \psi) \ominus (o + \eta) &= ((\overline{O} \wedge (o \ominus \rho + \pi)) + ((o \ominus \rho + \pi) \ominus \overline{O})) \\ &\quad \ominus (o + (\rho \ominus o)) \\ &= (o \ominus \rho + \pi) \ominus (o + (\rho \ominus o)) \\ &= ((o \ominus \rho) + \pi) \ominus (\rho + (o \ominus \rho)) \\ &= \pi \ominus \rho. \end{aligned}$$

For case (3):

$$\begin{aligned} \sigma \ominus \rho + \pi &= \sigma \ominus (\rho \ominus \pi) + (\pi \ominus \rho) \\ &= \sigma \ominus ((o + \eta) \ominus (o' + \psi)) + (o' + \psi) \ominus (o + \eta) \\ &= \sigma \ominus ((o + \eta) \ominus ((o + \eta) \wedge (o' + \psi))) \\ &\quad + (o' + \psi) \ominus ((o + \eta) \wedge (o' + \psi)) \tag{A 1} \\ &= \sigma \ominus (o + \eta) + ((o + \eta) \wedge (o' + \psi)) \\ &\quad + (o' + \psi) \ominus ((o + \eta) \wedge (o' + \psi)) \tag{A 2} \\ &= \sigma \ominus (o + \eta) + (o' + \psi) + ((o + \eta) \wedge (o' + \psi)) \\ &\quad \ominus ((o + \eta) \wedge (o' + \psi)) \\ &= \sigma \ominus (o + \eta) + (o' + \psi), \end{aligned}$$

where Eq. (A 2) follows from Eq. (A 1) because of the relations:

$$o + \eta \leq \sigma \text{ and } (o + \eta) \geq (o + \eta) \wedge (o' + \psi) \leq (o' + \psi).$$

\square

Proof of Theorem 3.2

Consider a well-defined reaction network $(\mathcal{S}, \mathcal{R})$ and let $(\mathcal{S}_{\overline{O}}, \mathcal{R}_{\overline{O}})$ be its expansion where

$$\mathcal{S}_{\overline{O}} = \mathcal{S} \cup \{ \llbracket o \rrbracket \mid o \in \mathbb{N}^{\mathcal{S}} \}.$$

Let $X(t)$ be the DRE solution of the expanded network and $P(t)$ the solution of the master equation of the original network at time t . Then

- (i) if $X_S(0) = 0$ then $X_S(t) = 0$ for all t and $S \in \mathcal{S}$;
- (ii) if $X_{\llbracket o \rrbracket}(0) = P_o(0)$ then $X_{\llbracket o \rrbracket}(t) = P_o(t)$, for all t and $o \in \mathbb{N}^{\mathcal{S}}$.

Proof. Case i). This statement holds if, whenever $X_S(t) = 0$, then $\frac{dX_S(t)}{dt} = 0$ for all $S \in \mathcal{S}$. The DRE for the expanded reaction network can be written as follows:

$$\frac{dX_S}{dt} = \sum_{\rho \xrightarrow{f} \pi \in \mathcal{R}_{\overline{O}}} (\pi_S - \rho_S) \cdot f_o(X), \quad \text{for all } S \in \mathcal{S}. \quad (\text{A } 3)$$

Since $o \in \mathbb{N}^{\mathcal{S}}$, every expanded reaction $\llbracket o \rrbracket + \eta \xrightarrow{f_o} \llbracket o' \rrbracket + \psi$ will be such that $\psi_S = 0$ for each $S \in \mathcal{S}$, hence $\pi_S = 0$ in Eq. (A 3). Let us now assume toward a contradiction that $\frac{dX_S(t)}{dt} \neq 0$ for $X_S(t) = 0$. This must hold only if both $\rho_S \neq 0$ and $f_o(X(t)) \neq 0$ for a reaction $\rho \xrightarrow{f} \pi \in \mathcal{R}_{\overline{O}}$ expanded from $\rho' \xrightarrow{f} \pi' \in \mathcal{R}$. For a given auxiliary species $\llbracket o \rrbracket$, the propensity function is in the form $f_o(X(t)) = X_{\llbracket o \rrbracket}(t) \cdot f(o + X_{|\mathcal{S}}(t))$. Since $X_S(t) = 0$ for each $S \in \mathcal{S}$ this reduces to $f_o(X(t)) = X_{\llbracket o \rrbracket}(t) \cdot f(o)$. As the reaction network is well-defined, $f(o) > 0$ implies that $\rho' \leq o$. In this case, from Eq 3.2 in the main text it follows that ρ must be in the form $\rho = \llbracket o \rrbracket + \emptyset$, that is, $\rho_S = 0$ for all $S \in \mathcal{S}$, closing this case by contradiction.

Case ii). For each $\llbracket o \rrbracket$, the DRE can be written as:

$$\begin{aligned} \frac{dX_{\llbracket o \rrbracket}}{dt} &= \sum_{(\llbracket \epsilon \rrbracket + \rho) \xrightarrow{f_\epsilon} (\llbracket o \rrbracket + \xi) \in \mathcal{R}_{\overline{O}}} f_\epsilon(X) \\ &\quad - \sum_{(\llbracket o \rrbracket + \xi) \xrightarrow{f_o} (\llbracket o' \rrbracket + \pi) \in \mathcal{R}_{\overline{O}}} f_o(X) \\ &= \sum_{(\llbracket \epsilon \rrbracket + \rho) \xrightarrow{f_\epsilon} (\llbracket o \rrbracket + \xi) \in \mathcal{R}_{\overline{O}}} X_{\llbracket \epsilon \rrbracket} \cdot f(\epsilon + X_{|\mathcal{S}}) \\ &\quad - \sum_{(\llbracket o \rrbracket + \xi) \xrightarrow{f_o} (\llbracket o' \rrbracket + \pi) \in \mathcal{R}_{\overline{O}}} X_{\llbracket o \rrbracket} \cdot f(o + X_{|\mathcal{S}}). \end{aligned}$$

Since $X_S = 0$ and the expanded reaction network is well-defined by Lemma A.1, this simplifies to:

$$\begin{aligned} \frac{dX_{\llbracket o \rrbracket}}{dt} &= \sum_{(\llbracket \epsilon \rrbracket + \emptyset) \xrightarrow{f_\epsilon} (\llbracket o \rrbracket + \emptyset) \in \mathcal{R}_{\overline{O}}} X_{\llbracket \epsilon \rrbracket} \cdot f(\epsilon + \emptyset) \\ &\quad - \sum_{(\llbracket o \rrbracket + \emptyset) \xrightarrow{f_o} (\llbracket o' \rrbracket + \emptyset) \in \mathcal{R}_{\overline{O}}} X_{\llbracket o \rrbracket} \cdot f(o + \emptyset) \\ &= \sum_{(\llbracket \epsilon \rrbracket + \emptyset) \xrightarrow{f_\epsilon} (\llbracket o \rrbracket + \emptyset) \in \mathcal{R}_{\overline{O}}} X_{\llbracket \epsilon \rrbracket} \cdot f(\epsilon) \\ &\quad - \sum_{(\llbracket o \rrbracket + \emptyset) \xrightarrow{f_o} (\llbracket o' \rrbracket + \emptyset) \in \mathcal{R}_{\overline{O}}} X_{\llbracket o \rrbracket} \cdot f(o). \end{aligned}$$

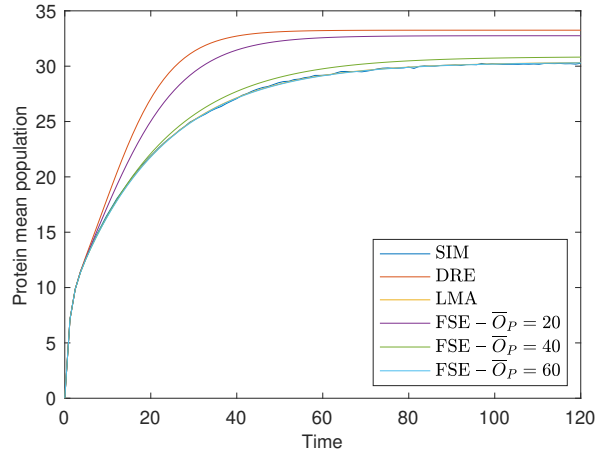


Figure 13: Comparison between FSE and linear mapping approximation (LMA) using the example in [25, Fig. 1] with parameter settings given by ‘Point B’ as therein specified. The model corresponds to the network from Eq. 4.3 with $k_b = 0$. The plot shows the average population of the protein species P (computed by stochastic simulation using 50k runs); the DRE approximation; the LMA approximation; and the approximation by FSE by setting $\bar{O}_P = 20, 40, 60$ and fixing $\bar{O}_{D_u} = \bar{O}_{D_b} = 1$. The trajectories for LMA and FSE with the largest observation bound $\bar{O}_P = 60$ provide an excellent agreement with stochastic simulation; however, FSE requires 65 equations, whereas LMA only defines 3 linear differential equations to provide a refined estimate of the mean protein populations.

The summations in the above equation can be written in terms of reactions of the original network as follows:

$$\begin{aligned}
 \frac{dX_{[[o]]}}{dt} &= \sum_{([\epsilon] + \emptyset) \xrightarrow{f_\epsilon} ([o] + \emptyset) \in \mathcal{R}_{\bar{O}}} X_{[[\epsilon]]} \cdot f(\epsilon) \\
 &\quad - \sum_{([o] + \emptyset) \xrightarrow{f_o} ([o'] + \emptyset) \in \mathcal{R}_{\bar{O}}} X_{[[o]]} \cdot f(o) \\
 &= \sum_{\substack{\rho \xrightarrow{f} \pi \in \mathcal{R} \\ o = \epsilon \ominus \rho + \pi}} X_{[[\epsilon]]} \cdot f(\epsilon) - \sum_{\substack{\rho \xrightarrow{f} \pi \in \mathcal{R} \\ o' = o \ominus \rho + \pi}} X_{[[o]]} \cdot f(o) \\
 &= \sum_{\substack{\rho \xrightarrow{f} \pi \in \mathcal{R} \\ \epsilon = o + \rho - \pi}} X_{[[\epsilon]]} \cdot f(\epsilon) - \sum_{\rho \xrightarrow{f} \pi \in \mathcal{R}} X_{[[o]]} \cdot f(o) \\
 &= \sum_{\rho \xrightarrow{f} \pi \in \mathcal{R}} \left(f(o - \pi + \rho) \cdot X_{[[o - \pi + \rho]]} - f(o) \cdot X_{[[o]]} \right),
 \end{aligned}$$

from which the claim follows by noting that substituting variable name $X_{[[\sigma]]}$ with P_σ , for all $\sigma \in \mathbb{N}^{\mathcal{S}}$, precisely gives the original master equation. \square

B. Comparison against the linear mapping approximation

This section compares FSE against linear mapping approximation [25], an analytical technique to compute probability distributions of species populations in a class of models of stochastic

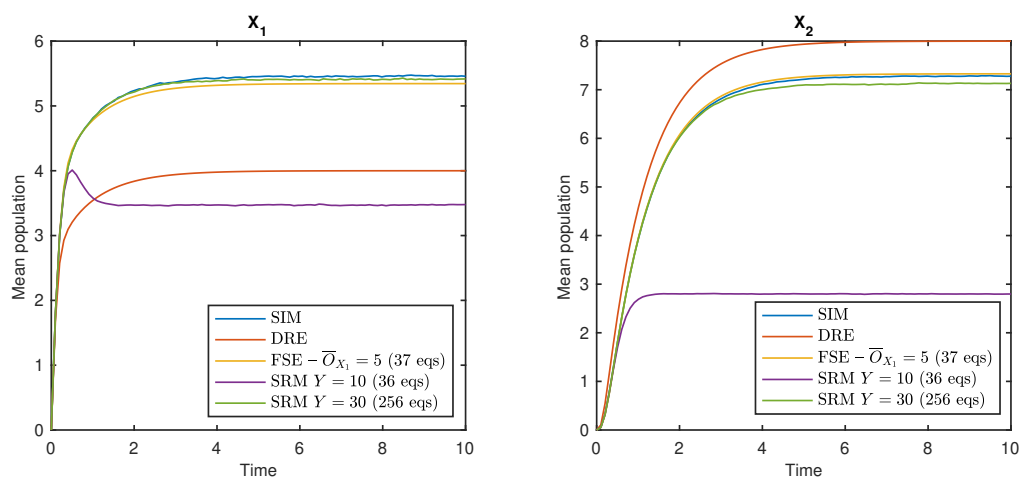


Figure 14: Comparison between FSE and SRM using the example in [49]. The plots show the average populations of species X_1 and X_2 (computed by stochastic simulation); their DRE approximation; the approximation by FSE by setting $\bar{O}_{X_1} = \bar{O}_{X_2} = 5$; and two estimates by SRM by setting the initial population of the buffer species Y equal to 10 and 30, respectively (with the master equation solved by stochastic simulation). The SRM setting $Y = 10$ gives a comparable number of equations than the FSE setting; however, the accuracy of the mean approximation is worse than DRE. On the other hand, the SRM setting $Y = 30$ gives estimates of comparable accuracy to those by FSE, but it gives rise to 256 equations instead of 37 as in FSE.

gene expression. In particular we consider the feedback switch model from [25, Fig. 1], which turns out to correspond to the network from Eq. 4.3 with $k_b = 0$. Figure 13 confirms that FSE can accurately approximate the mean population of the protein if the observation bound for the number of proteins is large enough, i.e., 60 in this case, corresponding to a system of (nonlinear) ODEs with 65 variables. On the other hand, LMA can achieve a similar accuracy, overlapping the simulated mean trajectory, in a more parsimonious fashion with a system of 3 linear ODEs.

C. Comparison against the slack reactants method

This section compares FSE against recent improvements of the finite state projection method (discussed in the main text) which cope with the problem of leaking probability mass into the absorbing state used for the truncation of the original state space. In particular we consider a numerical comparison against the most recently published method, called the *slack reactants* method (SRM) [49], which has been shown to offer the best performance against the other state-space truncation techniques.

The idea of SRM is to add “buffer species” that turn a reaction network with an infinite state space into one with a finite state space, but without absorbing states, while still preserving convergence properties as with the original FSP. Eliminating absorbing states allows for stochastic analysis over long time horizons, such as first passage times and estimations of stationary distributions.

We compare FSE against SRM an example provided by the authors of SRM in [49, Fig. 3A]. The chemical reaction network is reported here for convenience:



SRM modifies these reactions by adding a further species Y that acts as a buffer for the copies of the original species X_1 and X_2 that enter or exit the network. The modified reaction network is as follows:



where f is a modified propensity function of the form $20 \min(Y, 1)$.

The stochastic behavior of this network increasingly corresponds to that of the original one with larger buffer sizes represented by the initial number of copies of Y ; on the other hand, the larger such initial population the larger the number of states in the now finite state space of the Markov chain (hence the larger the number of equations in the master equation). Thus, a comparison between FSE and SRM can be done by measuring the difference between the estimations of the average populations given by the solution of the master equation of SRM and the DRE of FSE when both methods are set such that they give rise to systems of equations of similar size. Fig. 14 shows the results of this comparison, indicating that SRM needs ca. 7 times more equations to estimate the average population of the original species X_1 and X_2 in this model.