Rigorous elimination of fast stochastic variables from the linear noise approximation using projection operators

Citation for published version:

Digital Object Identifier (DOI):
10.1103/PhysRevE.86.041110

Link:
Link to publication record in Edinburgh Research Explorer

Document Version:
Publisher's PDF, also known as Version of record

Published In:
Physical Review E - Statistical, Nonlinear and Soft Matter Physics

Publisher Rights Statement:
RoMEO green

General rights
Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy
The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.
Rigorous elimination of fast stochastic variables from the linear noise approximation using projection operators

Philipp Thomas
Department of Physics, Humboldt University of Berlin, Newtonstr. 15, D-12489 Berlin, Germany and School of Biological Sciences, University of Edinburgh, Edinburgh EH9 3JR, United Kingdom

Ramon Grima
School of Biological Sciences, University of Edinburgh, Edinburgh EH9 3JR, United Kingdom

Arthur V. Straube
Department of Physics, Humboldt University of Berlin, Newtonstr. 15, D-12489 Berlin, Germany

(Received 30 July 2012; published 8 October 2012)

The linear noise approximation (LNA) offers a simple means by which one can study intrinsic noise in monostable biochemical networks. Using simple physical arguments, we have recently introduced the slow-scale LNA (ssLNA), which is a reduced version of the LNA under conditions of timescale separation. In this paper we present the first rigorous derivation of the ssLNA using the projection operator technique and show that the ssLNA follows uniquely from the standard LNA under the same conditions of timescale separation as those required for the deterministic quasi-steady-state approximation. We also show that the large molecule number limit of several common stochastic model reduction techniques under timescale separation conditions constitutes a special case of the ssLNA.

DOI: 10.1103/PhysRevE.86.041110

PACS number(s): 05.40.—a

I. INTRODUCTION

In the study of complex systems, it is common to invoke assumptions under which the dimension (and hence complexity) of the system is reduced; such a strategy often leads to relatively simple theories capable of exact analytical predictions, which offer insights typically lost in numerical approaches. This approach is particularly useful in the study of biochemical pathway dynamics, which typically involves the interaction of hundreds or thousands of different chemical species [1]. Deterministic models of such systems are frequently simplified by invoking the presence of well-separated timescales [1,2], in particular by means of the quasi-steady-state approximation (QSSA) [3,4]. However, it is well appreciated nowadays that a discrete stochastic approach in terms of chemical master equations (CMEs) is more faithful in describing dynamics inside living cells since the number of molecules of many species is in the range of few tens to few thousands [5]. The development of reduced stochastic models consistently coarse-grained over timescales presents a significantly larger challenge than the reduction of deterministic models because of the much larger number of differential equations which need to be solved in the former compared to the latter.

Indeed, it has been shown that it is not possible to formulate a reduced CME for the whole region of parameter space in which the deterministic QSSA is valid but rather only over a subset of this space [6–8]. For example, consider the Michaelis-Menten mechanism in which substrate reversibly binds with enzyme to form an enzyme-substrate complex which then decays into product. We are interested in conditions such that the substrate species decays over a much longer timescale than the enzyme and complex species. For this case, a consistently reduced CME can be written provided only that the complex decay rate into product is much less than its decay rate into substrate and enzyme [8,9]. However, it is possible to write down reduced deterministic rate equations (REs) using the QSSA even when the aforementioned condition is not satisfied [8].

Recently we have shown that a reduced stochastic description with the same range of validity as the deterministic QSSA is possible. This reduction is achieved by first approximating the CME by linear Langevin equations, an approximation called the linear noise approximation (LNA) and then integrating out the fast fluctuations such that one obtains a reduced set of linear Langevin equations for the fluctuations in the slowly varying species only. The latter is the slow-scale LNA (ssLNA) [10]. The advantage of the ssLNA over the reduced CME approach is that the former is valid over the same range of parameter space as the QSSA, a claim which has been numerically verified for a number of biochemical circuits including cooperative and noncooperative enzyme reactions and gene networks with or without negative feedback regulation [10]. The ssLNA, as it currently stands, has been derived by means of simple and intuitively clear physical arguments but is lacking a formal and rigorous derivation. In this article we provide such a derivation and also show the relationship of the ssLNA with other common methods of stochastic model reduction based on timescale separation assumptions.

The paper is divided as follows. In Sec. II we consider the CME of a general system of elementary chemical reactions with two characteristic and clearly separated timescales and reformulate the conventional system size expansion of the CME under such conditions. In Sec. III we use the leading order terms of the expansion to show that the deterministic...
limit of the CME under timescale separation conditions is formally the same as the conventional reduced rate equations obtained from the QSSA. The next to leading order terms of the expansion provide us with a Fokker-Planck equation which we reduce to a simpler Fokker-Planck equation for the slow fluctuations only, by means of the projection operator formalism. The reduced Fokker-Planck equation is obtained in closed-form for all monostable reaction networks and is the same as the ssLNA obtained in Ref. [10]. Finally in Sec. IV we show how the large molecule number limit of conventional stochastic reduction methods leads to a special case of the QSSA. The next to leading order terms formally the same as the conventional reduced rate equations limit of the CME under timescale separation conditions is

and assume that the species can interact via \( j = 1, \ldots, R \) elementary chemical reactions

\[
\sum_{i=1}^{N} s_{ij} X_i \xrightarrow{k_j} \sum_{i=1}^{N} r_{ij} X_i,
\]

where \( s_{ij} \) and \( r_{ij} \) are the stoichiometric coefficients [11] and \( k_j \) is the macroscopic reaction rate. The constraint \( \sum_i s_{ij} \leq 2 \) for all \( j \) implies that the reactions are unimolecular or bimolecular and hence elementary. The total number of species \( N \) is assumed to comprise \( N_s \) slow and \( N_f = N - N_s \) fast species, respectively. For convenience, we stick to the convention that \( X_1 \) to \( X_{N_s} \) denote the slow species, while \( X_{N_s+1} \) to \( X_{N} \) are reserved for the fast species.

In what follows, matrices are denoted by underlined quantities, and all vectors are column vectors. Let \( n_i \) denote the number of molecules of species \( i \), then the probability \( P(\vec{n}, t) \) to find the system in a particular state \( \vec{n} = (n_1, \ldots, n_N)^T \) is determined by the CME [12,13]:

\[
\frac{\partial P(\vec{n}, t)}{\partial t} = \Omega \sum_{j=1}^{N} \left( \prod_{i=1}^{N} E_i^{-S_{ij}} - 1 \right) \hat{f}_j(\vec{n}, \Omega) P(\vec{n}, t),
\]

where \( S_{ij} = (\underline{S})_{ij} = s_{ij} - r_{ij} \) is the stoichiometric matrix and \( E_i^j \) is a step operator, the action of which on some function of the absolute number of molecules results in the same function but with \( n_i \) replaced by \( n_i + z_i \), and \( \hat{f}_j(\vec{n}, \Omega) \) are the entries of the microscopic rate function vector. The product \( \Omega \hat{f}_j(\vec{n}, \Omega) \) \( dt \) represents the propensity function, which has the meaning of the probability that reaction \( j \) takes place in a small time interval \( dt \).

The CME is typically analytically intractable, and hence a systematic approximation method is needed. The system size expansion as developed by van Kampen is one such technique [12]. The heart of the method is the ansatz that the molecular concentration of the CME can be written as

\[
\frac{\vec{n}}{\Omega} = \vec{\phi} + \Omega^{-1/2} \vec{\eta},
\]

where \( \vec{\phi} \) is a vector of concentrations as given by the corresponding REs and the second term on the right-hand side is a stochastic term describing fluctuations about the concentrations of the REs.

We now impose timescale separation conditions, i.e., we assume that the transients in concentrations of a group of species decay much slower than those of the remaining group of species. The first group of species we label as the slow species, while the second are the fast species. The characteristic timescales of these two are \( \tau_s \) and \( \tau_f \), respectively. The vectors of molecular populations, of macroscopic concentrations, and of fluctuations can be divided into subpopulations of slow and fast species: \( \vec{n} = (\vec{n}_s, \vec{n}_f) \), \( \vec{\phi} = (\vec{\phi}_s, \vec{\phi}_f) = (\tau_s \vec{x}_s, \tau_f \vec{x}_f) \) and \( \vec{\eta} = (\vec{\eta}_s, \vec{\eta}_f) = (\tau_s^{-1/2} \vec{\epsilon}_s, \tau_f^{1/2} \vec{\epsilon}_f) \). The ansatz Eq. (3) can then be written as

\[
\frac{1}{\tau_s} \frac{\vec{n}_s}{\tau} = \vec{x}_s + (\Omega \tau_s)^{-1/2} \vec{\epsilon}_s, \quad \frac{1}{\tau_f} \frac{\vec{n}_f}{\tau} = \vec{x}_f + (\Omega \tau_f)^{-1/2} \vec{\epsilon}_f.
\]

We now use this ansatz to derive an expansion of the CME in powers of the small parameter \( \Omega^{-1/2} \) valid in

FIG. 1. Schematic illustrating the derivation of deterministic and stochastic quasi-steady-state approximations and rapid-equilibrium approximations for a reaction network with slow and fast species. The deterministic approach based on rate equations uses Tikhonov’s theorem while the stochastic approach utilizes the projection operator method applied to the system size expansion derived from a modified van Kampen ansatz.
the case of large volumes or, equivalently, for large copy numbers of molecules. For convenience the components of the various vectors will be denoted as follows: \( \bar{x}_i = (x_1, \ldots, x_N)_i \), \( \bar{x}_f = (x_{N+1}, \ldots, x_N)_f \), \( \bar{\epsilon}_i = (\epsilon_1, \ldots, \epsilon_N)_i \) and \( \bar{\epsilon}_f = (\epsilon_{N+1}, \ldots, \epsilon_N)_f \). The probability distribution of molecular populations \( P(\bar{n}, t) \) is henceforth changed into the distribution of slow and fast fluctuations \( \Pi(\bar{\epsilon}_i, \bar{\epsilon}_f, t) \). In what follows we consider how the time derivative, the step operator, and the microscopic rate function vector which compose the CME Eq. (2) transform under the proposed ansatz Eq. (4).

### A. Transformation of the time derivative

Using the change of variable theorem the time derivative can be written as

\[
\frac{\partial}{\partial t} P(\bar{n}, t) = \left( \frac{\partial}{\partial t} + \nabla_f^T \frac{d \bar{\epsilon}_f}{dt} \right) \Pi(\bar{\epsilon}_i, \bar{\epsilon}_f, t),
\]

where \( \nabla_f = (\partial/\partial \epsilon_1, \ldots, \partial/\partial \epsilon_N)^T \) and \( \nabla_f = (\partial/\partial \epsilon_{N+1}, \ldots, \partial/\partial \epsilon_N)^T \). The time derivatives at constant \( \bar{n} \), denoted by \( \Pi_{\bar{n}} \), can be computed from the ansatz Eq. (4), leading to

\[
\frac{\partial}{\partial t} P(\bar{n}, t) = \left( \frac{\partial}{\partial t} - \sqrt{\Omega} \nabla_i^T \frac{d \bar{x}_i}{dt} - \sqrt{\Omega} \nabla_f^T \frac{d \bar{\epsilon}_f}{dt} \right) \times \Pi(\bar{\epsilon}_i, \bar{\epsilon}_f, t).
\]  

### B. Transformation of the step operator

By the definition of the step operator we have

\[
\prod_{i=1}^{N} E_i^{S_i} g(n_1, \ldots, n_N) = g(n_1 - (S_i)_i, 1, \ldots, n_{N_i} - (S_i)_N, j, n_{N+1} - (S_f)_j, \ldots, n_N - (S_f)_N),
\]

where we have partitioned the stoichiometric matrix into two parts: \( (S_i)_i \) for \( 1 \leq i \leq N_i \) (the stoichiometric matrix for the slow species) and \( (S_f)_j \) for \( 1 \leq j \leq N_f \) (the stoichiometric matrix for the fast species). Note that \( g \) denotes some general function of the molecule numbers. Applying the ansatz Eq. (4) it follows that the above equation can be written as

\[
\prod_{i=1}^{N} E_i^{S_i} g(\epsilon_1, \ldots, \epsilon_N) = g(\epsilon_i, (\Omega \tau_s)^{-1/2}(S_i)_i, 1, \ldots, \epsilon_N, (\Omega \tau_s)^{-1/2}(S_i)_N, j, (\Omega \tau_f)^{-1/2}(S_f)_1, \ldots, \epsilon_N, (\Omega \tau_f)^{-1/2}(S_f)_N),
\]

where we have introduced the nondimensional ratio of slow and fast timescales \( \tau = \tau_s/\tau_f \). We substitute Eqs. (9), (9), and (10) in the CME [Eq. (2)] and rescaling time by the slow timescale \( \tau = \tau_s/\tau_f \), we obtain the transformed CME:

\[
\frac{\partial}{\partial \tau} \Pi(\bar{\epsilon}_i, \bar{\epsilon}_f, \tau) - (\Omega \tau_s)^{1/2} \nabla_i^T \left[ \frac{\partial \bar{x}_i}{\partial \tau} - S_i \tilde{f}(\bar{x}_i, \bar{x}_f) \right] \Pi(\bar{\epsilon}_i, \bar{\epsilon}_f, \tau) - (\Omega \tau_f)^{1/2} \nabla_f^T \left[ \frac{\partial \bar{\epsilon}_f}{\partial \tau} - \gamma S_f \tilde{f}(\bar{x}_i, \bar{x}_f) \right] \Pi(\bar{\epsilon}_i, \bar{\epsilon}_f, \tau) = \Omega^0 (\gamma \mathcal{L} + \gamma^{1/2} \mathcal{L}_{int} + \mathcal{L}_d) \Pi(\bar{\epsilon}_i, \bar{\epsilon}_f, \tau) + O(\Omega^{-1/2}),
\]

where we have introduced the nondimensional ratios of slow and fast timescales

\[
\gamma = \frac{\tau_s}{\tau_f}.
\]
Note that the order $\Omega^0$ is defined by the operators
\begin{align}
\mathcal{L}_s &= -\nabla^T_s \tilde{\mathbf{j}} + \frac{1}{2} \nabla^T_s \tilde{D}_s \nabla_s, \\
\mathcal{L}_f &= -\nabla^T_f \tilde{\mathbf{j}} + \frac{1}{2} \nabla^T_f \tilde{D}_f \nabla_f, \\
\mathcal{L}_{\text{int}} &= -\nabla^T_f \tilde{\mathbf{j}}_f + \frac{1}{2} \nabla^T_f \tilde{D}_{\text{int}} \nabla_f
\end{align}
(14a,b,c)
where $\tilde{\mathbf{j}}_s = \mathcal{S}_s (\nabla \tilde{\mathbf{j}}_s)^T$, $\tilde{\mathbf{j}}_f = \mathcal{S}_f (\nabla \tilde{\mathbf{j}}_f)^T$, and $\tilde{\mathbf{j}}_{\text{int}} = \mathcal{S}_f (\nabla \tilde{\mathbf{j}}_{\text{int}})^T$.

The presence of timescale separation is reflected by the large parameter $\gamma$ diminishing the time derivative in Eq. (15b). Such a set of equations present a special case in singular perturbation theory, where Eqs. (15a) and (15b) for the slow and fast variables are typically referred to as the degenerate and adjoined systems, respectively [11]. Tikhonov’s first theorem [16,17] states that a simplification of the above equations under timescale separation conditions is possible whenever certain requirements are met: (1) the solutions of both the degenerate and adjoined systems [Eqs. (15)] are unique and their right-hand sides are continuous functions; (2) the root $\tilde{x}_f = h(\tilde{x}_s, \tau)$ is the stable solution of the adjoined system; and (3) the initial values $\tilde{x}_f (\tau = 0)$ are in the domain of influence of the solution as in (2). Whenever these prerequisites are met, the solution of the full system (15) for $\tilde{x}_s$ tends to the solution of the reduced system
\begin{equation}
\frac{d\tilde{x}_s}{d\tau} = \mathcal{S}_s \tilde{f}(\tilde{x}_s, h(\tilde{x}_s)),
\end{equation}
(16)
in the limit of timescale separation, i.e., $\gamma^{-1} \to 0$. Note that $\tilde{x}_f = h(\tilde{x}_s)$ is the solution of $\mathcal{S}_f \tilde{f}(\tilde{x}_s, \tilde{x}_f) = 0$.

These requirements are typically fulfilled for the biochemical networks of interest. This is since the chemical transformation scheme (1) is formulated for elementary reactions, which are bimolecular or simpler, the right-hand sides of Eqs. (15) are continuous polynomial functions of the second order at most. For monostable networks, the rate equations admit a single steady state which is the same for the full and the reduced REs, i.e., Eqs. (15) and (16). It is therefore clear that all the solutions will tend to this state with time, quicker for fast variables and slower for the slow ones.

III. DERIVATION OF THE SLOW-SCALE LINEAR NOISE APPROXIMATION

A. Deterministic QSSA

The leading order terms, $O(\Omega^{1/2})$, of Eq. (12) describe the dynamics of macroscopic concentrations and is given by the coupled set of REs:
\begin{align}
\frac{d\tilde{x}_s}{d\tau} &= \mathcal{S}_s \tilde{f}(\tilde{x}_s, \tilde{x}_f), \\
\frac{1}{\gamma} \frac{d\tilde{x}_f}{d\tau} &= \mathcal{S}_f \tilde{f}(\tilde{x}_s, \tilde{x}_f).
\end{align}
(15a,b)

The reduction is then carried out defining the operator
\begin{equation}
\mathcal{L}_f \pi(\tilde{\mathbf{e}}_f) = 0.
\end{equation}
(18)
The reduction is then carried out defining the operator
\begin{equation}
(\mathcal{P}_1) \Pi(\tilde{\mathbf{e}}_s, \tilde{\mathbf{e}}_f, \tau) = \pi(\tilde{\mathbf{e}}_f) \int d\tilde{\mathbf{e}}_f \Pi(\tilde{\mathbf{e}}_s, \tilde{\mathbf{e}}_f, \tau) = \pi(\tilde{\mathbf{e}}_f) \Pi(\tilde{\mathbf{e}}_s, \tilde{\mathbf{e}}_f, \tau),
\end{equation}
(19)
projecting the probability distribution $\Pi(\tilde{\mathbf{e}}_s, \tilde{\mathbf{e}}_f, \tau)$ onto the distribution of fast fluctuations evaluated at steady state. Note that the above definition satisfies the relation $\mathcal{P}^2 = \mathcal{P}$ and hence $\mathcal{P}$ is indeed a projector.

In what follows we use the forms we have derived for the operators in the Fokker-Planck Eq. (17), i.e., those given by Eq. (14), to deduce three properties of the projection operator. Given these properties we then show how the projection operator applied to Eq. (17) leads to a reduced Fokker-Planck equation in the slow variables only.

1. Properties of the projection operator

In this subsection we will show that the following properties hold:
\begin{equation}
\mathcal{P} \mathcal{L}_s = \mathcal{L}_s \mathcal{P},
\end{equation}
(20a)
\[ \mathcal{PL}_f = \mathcal{L}_f \mathcal{P} = 0, \quad (20b) \]
\[ \mathcal{PL}_\text{int} \mathcal{P} = 0. \quad (20c) \]

First, property (20a) follows from the fact that the projection operator \( \mathcal{P} \), as defined by Eq. (19), acts only on the fast variables \( \tilde{\epsilon}_f \), whereas \( \mathcal{L}_s \), see Eq. (14a), acts only on the slow variables \( \epsilon_s \), and hence the two operators \( \mathcal{P} \) and \( \mathcal{L}_s \) commute.

Second, we show that both equalities of property (20b) are satisfied. Considering the left-hand side, we obtain
\[ \mathcal{PL}_f \mathcal{P} = (\mathcal{L}_f \pi(\tilde{\epsilon}_f)) \int d\tilde{\epsilon}_f = 0, \quad (22) \]

by the quasi-steady-state condition, \( \mathcal{L}_f \pi = 0 \).

The third property (20c) can be obtained as follows. The first, second, and fourth terms of \( \mathcal{L}_\text{int} \) as given by Eq. (14c) have the form of a divergence in the fast variables and hence by the partial integration lemma, its integral vanishes in the absence of boundary terms. By considering the right-hand side, we have
\[ \mathcal{L}_f \mathcal{P} = (\mathcal{L}_f \pi(\tilde{\epsilon}_f)) \int d\tilde{\epsilon}_f = 0, \quad (21) \]

since \( \mathcal{L}_f \), as given by Eq. (14b), has the form of a divergence in the fast variables and hence by the partial integration lemma, its integral vanishes in the absence of boundary terms. By considering the right-hand side, we have
\[ s \tilde{\epsilon}(s) - v(0) - y^{1/2} \mathcal{P} \mathcal{L}_\text{int} D(s)v(0) = [\mathcal{L}_s + \gamma \mathcal{P} \mathcal{L}_\text{int} D(s)] \tilde{\epsilon}(s), \quad (27) \]

where we have used definition (24) and property (20c) and introduced \( D(s) = (s - \mathcal{L}_s - \gamma \mathcal{L}_f - y^{1/2} \mathcal{Q} \mathcal{L}_\text{int} \mathcal{P} \). From the above equation one can draw the limit \( y \to \infty \) for which \( D(s) \sim -(\gamma \mathcal{L}_f)^{-1} \).

Note that use has been made of the properties (20a) and (20b). Solving for \( \tilde{\epsilon}(s) \), we obtain
\[ \frac{\partial}{\partial \tau} v(\tau) = [\mathcal{L}_s - \mathcal{P} \mathcal{L}_\text{int} \mathcal{L}_f^{-1} \mathcal{L}_\text{int}] v(\tau), \quad (29a) \]

where we have inverted the Laplace transform Eq. (25). Note that due to the vanishing of the third term on the left-hand side of Eq. (27) this asymptotic limit is Markovian and hence does not require the knowledge of the initial distribution \( w(0) \) of the fast fluctuations. Using \( v(\tau) = \pi(\tilde{\epsilon}_f) \Pi(\tilde{\epsilon}_s, \tau) \) and integrating over the fast fluctuations \( \tilde{\epsilon}_f \) we obtain
\[ \frac{\partial}{\partial \tau} \Pi(\tilde{\epsilon}_s, \tau) = \mathcal{L}' \Pi(\tilde{\epsilon}_s, \tau), \quad (29b) \]

where the angled brackets with subscript \( \pi \) in Eq. (29b) denote the statistical average over the steady-state probability distribution \( \pi(\tilde{\epsilon}_f) \) of fast fluctuations.

### 3. Derivation of the slow-scale linear noise approximation

The above equation is a generic form for the Fokker-Planck equation for the slow fluctuations \( \tilde{\epsilon}_s \) under timescale separation conditions. What remains is to explicitly evaluate the average over \( \pi(\tilde{\epsilon}_f) \) such that we obtain a closed-form expression for the reduced Fokker-Planck equation. We now show these evaluation steps in detail.

Using \( \mathcal{L}_\text{int} \) as given by Eq. (14) together with Eq. (29b) we can deduce the form of the reduced Fokker-Planck operator:
\[ \mathcal{L}' = \mathcal{L}_s - \mathcal{P} \mathcal{L}_\text{int} \mathcal{L}_f^{-1} \mathcal{L}_\text{int} \]
\[ \equiv \mathcal{L}_s - \mathcal{P} \mathcal{L}_\text{int} \mathcal{L}_f^{-1} \mathcal{L}_\text{int} \]
\[ = \mathcal{L}_s - \mathcal{P} \mathcal{L}_\text{int} \mathcal{L}_f^{-1} \mathcal{L}_\text{int}, \quad (30) \]

Note that terms which have \( \nabla_f \) to the left do not contribute to the reduced operator and hence are missing from the above equation.

We proceed by evaluating the two distinct correlators appearing in the above expression explicitly. We shall make use of the identity
\[ \int_0^\infty du e^{\mathcal{L}_f u} = \mathcal{L}_f^{-1} e^{\mathcal{L}_f u}|_0^\infty = -\mathcal{L}_f^{-1}(1 - \mathcal{P}), \quad (31) \]

which can be verified from straightforward integration and the fact that \( \mathcal{P} \Pi(\tilde{\epsilon}_s, \tilde{\epsilon}_f, \tau) = \lim_{u \to \infty} e^{\mathcal{L}_f u} \Pi(\tilde{\epsilon}_s, \tilde{\epsilon}_f, \tau) \) [19]. Using
the fact that $\mathcal{P}_{\epsilon f}^T \pi = 0$, we can write

$$\langle \bar{\epsilon}_f L_f^{-1} \bar{\epsilon}_f^T \rangle_\pi = \int d\bar{\epsilon}_f \bar{\epsilon}_f L_f^{-1} \epsilon (1 - \mathcal{P}) \epsilon_f^T \pi \quad \text{where } \epsilon_f = e^{-\sqrt{\phi_s} t} \int \frac{d\phi_s}{\sqrt{2\pi}} e^{-\frac{\phi_s^2}{2}}$$

$$\langle \bar{\epsilon}_f L_f^{-1} \bar{\epsilon}_f^T \rangle_\pi = -\int_0^\infty dt \int d\bar{\epsilon}_f \bar{\epsilon}_f e^{\mathcal{P} \epsilon_u^T} \epsilon_f^T \pi \quad \text{where } \epsilon_u = \frac{\sqrt{2\phi_s}}{\sqrt{\Omega}}$$

$$\langle \bar{\epsilon}_f L_f^{-1} \bar{\epsilon}_f^T \rangle_\pi = -\int_0^\infty dt d\bar{\epsilon}(\bar{\epsilon}_f(0) \bar{\epsilon}_f^T(0))_\pi. \quad (32)$$

Note that in the third step we have taken into account that $\bar{\epsilon}_f^T \pi(\bar{\epsilon}_f, u) = e^{\mathcal{P} \epsilon_u^T} \epsilon_f^T \pi(\bar{\epsilon}_f, 0)$ is a solution to $\epsilon, \pi = L_f \pi$ with the initial condition $\bar{\epsilon}_f^T \pi(0)$. One can utilize the Fourier transform of the autocorrelation matrix $\langle \bar{\epsilon}_f(u) \bar{\epsilon}_f^T(0) \rangle_\pi = \int d\omega(e^{2\pi}\epsilon_f^T \pi).$

Similarly, one can show that

$$\langle \bar{\epsilon}_f L_f^{-1} \bar{\epsilon}_f^T \rangle_\pi = -\int_0^\infty dt d\bar{\epsilon}(\bar{\epsilon}_f(0) \bar{\epsilon}_f^T(0))_\pi. \quad (34)$$

Here $\langle \bar{\epsilon}_f \bar{\epsilon}_f^T \rangle_\pi = -I$ with identity matrix $I$, is evaluated by partial integration. Plugging Eqs. (33) and (34) into Eq. (30) gives a closed-form expression for the Fokker-Planck operator of the reduced probability distribution $\Pi_{\epsilon f}(t)$:

$$\mathcal{L} = -\nabla f^T \mathcal{D}_{ss} \nabla S_f, \quad (35a)$$

$$\mathcal{J} = \mathcal{D}_{ss} + \mathcal{D}_{fs} \epsilon_f, \quad (35b)$$

$$\mathcal{D}_{ss} = \mathcal{D}_{ss} + \frac{1}{2} \left( \frac{1}{2} \mathcal{D}_{ss} \mathcal{D}_{fs} - \frac{1}{2} \mathcal{D}_{fs} \mathcal{D}_{ss} \right). \quad (35c)$$

We note that the slow-scale Jacobian (36b) coincides with the reduced Jacobian as obtained from the macroscopic QSSA equations, i.e., Eqs. (16), as shown in Ref. [10]. It is also important to note that the reduced diffusion matrix $\mathcal{D}_{ss}$ admits the representation

$$\mathcal{D}_{ss} = \Omega^{-1}(\mathcal{A} - \mathcal{B})(\mathcal{A} - \mathcal{B})^T, \quad (37)$$

where $\mathcal{A} = \frac{\mathcal{S}_s}{\sqrt{\mathcal{F}}} \mathcal{F}$ and $\mathcal{B} = \frac{1}{2} \frac{1}{\sqrt{\mathcal{F}}} \frac{1}{\sqrt{\mathcal{F}}} \mathcal{S}_f$. From this representation it can be immediately deduced that the reduced matrix $\mathcal{D}_{ss}$ is symmetric and positive semidefinite, two crucial properties of the diffusion matrices for all Fokker-Planck equations [12]. Using standard methods [19] one can also obtain the Langevin equations corresponding to the slow-scale Fokker-Planck equation (36a). These are given by

$$\frac{d}{dt} \hat{\epsilon}_s(t) = \frac{1}{\sqrt{\eta}} \epsilon_s(t) + \frac{1}{\sqrt{2}} \sqrt{\eta} \mathcal{D}_{ss} \epsilon_s, \quad \epsilon_{ss} = -\frac{1}{2} \int d\epsilon_f \mathcal{D}_{ff} \epsilon_f - \frac{1}{2} \int d\epsilon_f \mathcal{D}_{fs} \epsilon_f + \frac{1}{2} \int d\epsilon_f \mathcal{D}_{fs} \epsilon_f + \frac{1}{2} \int d\epsilon_f \mathcal{D}_{ss} \epsilon_f. \quad (36a)$$

$$\frac{d}{dt} \epsilon_{ss}(t) = -\frac{1}{\sqrt{\eta}} \epsilon_{ss}(t) + \frac{1}{\sqrt{2}} \sqrt{\eta} \mathcal{D}_{ss} \epsilon_{ss}, \quad \mathcal{D}_{ss} = \mathcal{D}_{ss} + \frac{1}{2} \left( \frac{1}{2} \mathcal{D}_{ss} \mathcal{D}_{fs} - \frac{1}{2} \mathcal{D}_{fs} \mathcal{D}_{ss} \right). \quad (36c)$$

where $\frac{d}{dt} \hat{\epsilon}_s(t) = \frac{1}{\sqrt{\eta}} \epsilon_s(t) + \frac{1}{\sqrt{2}} \sqrt{\eta} \mathcal{D}_{ss} \epsilon_s$ and $\frac{d}{dt} \epsilon_{ss}(t) = -\frac{1}{\sqrt{\eta}} \epsilon_{ss}(t) + \frac{1}{\sqrt{2}} \sqrt{\eta} \mathcal{D}_{ss} \epsilon_{ss}$ as well as $
abla_s = \Omega^{-1} \mathcal{S}_s \mathcal{F} \mathcal{F}^T$, $\mathcal{F} = \frac{1}{2} \mathcal{D}_{ss} \mathcal{D}_{fs}$ as well as $\mathcal{F} = \frac{1}{2} \mathcal{D}_{ss} \mathcal{D}_{fs}$ and $\mathcal{F} = \text{diag}(\hat{f})$.

Using the size parameter $\mu$ we can write the vector of macroscopic rate functions as

$$\tilde{f}(\hat{x}_s, \hat{x}_f) = (f_1, f_2, \ldots, f_R) \mu \tilde{f}(\hat{x}_s, \hat{x}_f), \quad (40)$$

where $\tilde{f}_j(\hat{x}_s, \hat{x}_f) = (f_1, f_2, \ldots, f_R)$ are the rates of the slow reactions and $\tilde{f}_j(\hat{x}_s, \hat{x}_f) = \mu^{-1}(f_{R+1}, f_{R+2}, \ldots, f_R)$ are the rates of the fast reactions rescaled by the size parameter $\mu$. Note that here $\mu$ is determined by the infinite volume limit of
Eq. (39). We can now partition the stoichiometric matrix into block matrices
\[
\mathbf{S} = \begin{bmatrix}
\mathbf{S}_s \\
\mathbf{S}_f
\end{bmatrix} = \begin{bmatrix}
\mathbf{S}_s^{(i)} & \mathbf{S}_s^{(f)} \\
\mathbf{S}_f^{(i)} & \mathbf{S}_f^{(f)}
\end{bmatrix},
\tag{41}
\]
discriminating slow and fast reactions (superscript) as well as slow and fast species (subscript). The matrices \(\mathbf{S}_s^{(i)}\) and \(\mathbf{S}_s^{(f)}\) denote the stoichiometries of the slow and fast species in the slow and fast reactions, respectively, while \(\mathbf{S}_f^{(i)}\) and \(\mathbf{S}_f^{(f)}\) represent the stoichiometry of slow species in fast reactions and the stoichiometry of fast species in slow reactions.

### A. Deterministic rapid-equilibrium approximation

The macroscopic elimination starts from the conventional REs:
\[
\frac{d\mathbf{x}_s}{d\tau} = \mu \mathbf{S}_s^{(f)} \mathbf{f}_f + \mathbf{S}_s^{(i)} \mathbf{f}_i,
\tag{42a}
\]
\[
\frac{1}{\gamma} \frac{d\mathbf{x}_f}{d\tau} = \mu \mathbf{S}_f^{(f)} \mathbf{f}_f + \mathbf{S}_f^{(i)} \mathbf{f}_i,
\tag{42b}
\]
which are similar to Eqs. (15) but discriminate slow and fast reactions by the size parameter \(\mu\). It is clear that \(\mu\) and \(\gamma\) must be of the same order since slow and fast timescales are determined by the size of the rate functions.

Further, we observe that for \(\mu \rightarrow \infty\) the above equations are not immediately of the form required by Tikhonov’s theorem [compare Eqs. (15)], and hence the adiabatic elimination is not immediately applicable unless we impose \(\mathbf{S}_s^{(f)} = 0\). This condition implies that the populations of slow species are not changed in fast reactions and is also imposed throughout the literature in reducing stochastic slow-fast reaction networks [24–26,28]. Setting the time derivative of the second equation to zero we can solve \(\mathbf{S}_s^{(f)} \mathbf{f}_f \approx 0\) for \(\mathbf{f}_f = \mathbf{h}(\mathbf{x}_s) + O(\mu^{-1})\) to obtain the reduced system
\[
\frac{d\mathbf{x}_s}{dt} = \mathbf{S}_s^{(i)} \mathbf{f}(\mathbf{x}_s, \mathbf{h}(\mathbf{x}_s)).
\tag{43}
\]
In the case where the equilibrium of the fast reactions \(\mathbf{S}_f^{(f)} \mathbf{f}_f = 0\) is detailed balanced, the above approximation is called the deterministic rapid-equilibrium approximation [11], i.e., the case when the fast reactions are given by a set of reversible reactions for which the forward and backward rates of each reaction cancel each other.

### B. Stochastic rapid-equilibrium approximation

We can now apply the ssLNA to obtain the contribution of the fluctuations using the same conditions as used above for the deterministic rapid-equilibrium approximation. First, we make use of the condition \(\mathbf{S}_s^{(f)} = 0\) to obtain the coefficients
\[
\begin{align*}
\mathbf{I}_s &= \mathbf{S}_s^{(i)}(\mathbf{v}_{\phi_s} \mathbf{f}_s)^T, \\
\mathbf{I}_f &= \mu \mathbf{S}_s^{(f)}(\mathbf{v}_{\phi_s} \mathbf{f}_s)^T + \mathbf{S}_s^{(i)}(\mathbf{v}_{\phi_s} \mathbf{f}_s)^T, \\
\mathbf{I}_s f &= \mathbf{S}_s^{(f)}(\mathbf{v}_{\phi_s} \mathbf{f}_s)^T, \\
\mathbf{I}_s f s &= \mathbf{S}_s^{(i)}(\mathbf{v}_{\phi_s} \mathbf{f}_s)^T + \mu \mathbf{S}_s^{(f)}(\mathbf{v}_{\phi_s} \mathbf{f}_s)^T,
\end{align*}
\]
which distinguish contributions from slow and fast reactions. Second, we can use these together within Eq. (36b) to obtain the reduced Jacobian by taking the limit \(\mu \rightarrow \infty\):
\[
\mathbf{I} = \mathbf{S}_s^{(i)}(\mathbf{v}_{\phi_s} \mathbf{f}_s)^T - \mathbf{S}_s^{(i)}(\mathbf{v}_{\phi_s} \mathbf{f}_s)^T [\mathbf{S}_s^{(f)}(\mathbf{v}_{\phi_s} \mathbf{f}_s)^T]^{-1} \times \mathbf{S}_s^{(f)}(\mathbf{v}_{\phi_s} \mathbf{f}_s)^T.
\tag{44}
\]
It can be shown that the above expression coincides with the Jacobian obtained using Eq. (43). Third, we calculate the coefficients of the noise from Eq. (37) as
\[
\mathbf{A} = \mathbf{S}_s^{(i)} \sqrt{\mathbf{F}_s}, \quad \mathbf{B} = 0,
\tag{45}
\]
where \(\mathbf{F}_s = \text{diag}(\mathbf{f}_s)\) and \(\mathbf{F}_f = \text{diag}(\mathbf{f}_f)\). Note that the second equation follows from inserting \(\mathbf{S}_f \sqrt{\mathbf{F}_s} = \mathbf{S}_f^{(i)} \sqrt{\mathbf{F}_s} + \mu \mathbf{S}_f^{(f)} \sqrt{\mathbf{F}_s} + \mu \mathbf{S}_f^{(f)} \sqrt{\mathbf{F}_f}\) into the definition of \(\mathbf{B}\) after Eq. (37) together with the expression for \(\mathbf{J}_f\) and \(\mathbf{J}_s\) and taking the limit \(\mu \rightarrow \infty\). This implies that \(\mathbf{B}\) is of order \(\mu^{-1/2}\), and hence the noise stemming from the fast reactions can be neglected in the limit \(\mu \rightarrow \infty\). Finally, we can formulate the Langevin equations:
\[
\frac{d}{dt} \mathbf{h}_s(t) = \mu \mathbf{h}_s(t) + \mathbf{A} \mathbf{h}_s(t) + \Omega_s^{(i)} \mathbf{S}_s^{(i)} \sqrt{\mathbf{F}_s} \mathbf{h}_s(t).
\tag{46}
\]
Note that these equations are consistent with those obtained using stoichiometry \(\mathbf{S}_s^{(i)}\) and propensity vector \(\mathbf{f}(\mathbf{x}_s, \mathbf{h}(\mathbf{x}_s))\) of the reduced macroscopic equations (43). Note that while in the ssLNA [Eq. (38)] (which is consistent with QSSA conditions) both the noise in the fast and slow reactions contribute to the noise of the slow species, in the stochastic rapid-equilibrium approximation [Eq. (46)], the noise in the slow reactions alone determines that in the slow species.

The latter Langevin equation has also been obtained by Pahalajani et al. [28] starting from the decomposition of reactions into slow and fast categories. However, our derivation is the first to clearly show that this Langevin equation is a partial case of the ssLNA and hence is valid only over a subset of the parameter space over which the QSSA holds.

### V. Discussion

In this article we have shown how to rigorously reduce the linear noise approximation of the CME by using the projection operator formalism to eliminate the fast fluctuation variables. The resulting Langevin equation, Eq. (38), is in agreement with the ssLNA as has been previously deduced from intuitive arguments only [10]. The present derivation provides a rigorous basis for deriving the LNA from the system size expansion using a modified van Kampen ansatz, Eq. (4), which is applicable under conditions of timescale separation. The resulting REs, Eq. (15), and the Fokker-Planck equation (17) obtained by this approach are of a particular form which allows direct application of Tikhonov’s theorem and the projection operator method, respectively. Hence by this procedure it is guaranteed that the mesoscopic elimination of the fast fluctuations is valid in exactly the same limit as the macroscopic elimination of the concentrations of the fast species by the deterministic QSSA.

For reaction networks composed of slow and fast reactions, conditions considered by the majority of available stochastic model reduction methods employing timescale separation, we have shown that in the limit of large volumes or molecule
numbers the CME can be approximated by a Langevin equation, Eq. (46), which is a special case of the ssLNA. The main advantage of using the ssLNA over the various aforementioned methods [23–28] is that the ssLNA is valid over a larger parameter range than the latter methods. The path integral approach developed in Ref. [29] also enjoys this property. However, the ssLNA enjoys the further advantage that it is available in closed form for any monostable reaction network and hence can be readily constructed from knowledge of the stoichiometric matrix and deterministic rate equations. The disadvantage of the ssLNA is that for pathways composed of some second-order reactions, the ssLNA (and the LNA on which it is based) is valid only for large enough molecule numbers [30,31]. This limitation can be lifted by consideration of higher order terms in the system size expansion; such calculations present more formidable analytical challenges than encountered in the derivation of the ssLNA and are under current investigation.

For realistic biochemical networks there may be particular parameter ranges for which the stability of the dynamics is either monostable or bistable or even oscillatory states can occur. The applicability of the method therefore depends on the type of stability realized for a particular network under consideration. Oscillatory states are found for 10% of the transcriptome and 20% of the proteome in mouse liver [32,33] and similar fractions in the human metabolome [34]. Although bimodal distributions have been observed experimentally, as, for instance, in the lac operon of E. coli [35], a recent proteome-wide study suggests that such probability distributions (which potentially indicate bistability) are quite rare [36] and similarly for the human transcriptome [37]. Despite the fact that bistable and oscillatory properties are important for specific cellular functions, it appears that monostable networks for which the present theory has been developed are common in living cells.

We note that the LNA has been applied also to networks with limit cycles [38]. The resulting equation is still a linear Fokker-Planck equation, and hence the elimination of fast variables can be performed along the same lines as in the present derivation. However, the analysis of the resulting equation has to be carried out by means of Floquet theory due to the inherent phase diffusion in these systems [39]. For bistable systems, the underlying distribution cannot be captured by a linear Fokker-Planck equation, and hence the LNA is not applicable in this case [12]. A commonly employed procedure to eliminate fast variables in such systems is the stochastic QSSA. A recent numerical study reported considerable discrepancies in the probability distributions of full and QSSA-reduced bistable systems [40]. We expect these discrepancies to be similar to the difference in the ssLNA and rapid-equilibrium approximations discussed in this article for the monostable case. Therefore, developing a technique to rigorously reduce the CME of bistable networks remains still an open question. Since our method is devised for monostable systems and can be extended to oscillatory ones under timescale separation conditions we expect it to be of broad applicability for the study of intracellular reaction networks.

ACKNOWLEDGMENTS

R.G. acknowledges support by SULSA (Scottish Universities Life Science Alliance).