AIP | The Journal of Chemical Physics

Stochastic theory of large-scale enzyme-reaction networks: Finite copy number corrections to rate equation models

Philipp Thomas, Arthur V. Straube, and Ramon Grima

Citation: The Journal of Chemical Physics **133**, 195101 (2010); doi: 10.1063/1.3505552
View online: http://dx.doi.org/10.1063/1.3505552
View Table of Contents: http://scitation.aip.org/content/aip/journal/jcp/133/19?ver=pdfcov
Published by the AIP Publishing

Stochastic theory of large-scale enzyme-reaction networks: Finite copy number corrections to rate equation models

Philipp Thomas,[1] Arthur V. Straube,[1] and Ramon Grima[2,a)]
[1]*Department of Physics, Humboldt University of Berlin, Newtonstr. 15, D-12489 Berlin, Germany*
[2]*School of Biological Sciences, University of Edinburgh, Edinburgh EH9 3JR, United Kingdom*

(Received 5 August 2010; accepted 5 October 2010; published online 18 November 2010)

Chemical reactions inside cells occur in compartment volumes in the range of atto- to femtoliters. Physiological concentrations realized in such small volumes imply low copy numbers of interacting molecules with the consequence of considerable fluctuations in the concentrations. In contrast, rate equation models are based on the implicit assumption of infinitely large numbers of interacting molecules, or equivalently, that reactions occur in infinite volumes at constant macroscopic concentrations. In this article we compute the finite-volume corrections (or equivalently the finite copy number corrections) to the solutions of the rate equations for chemical reaction networks composed of arbitrarily large numbers of enzyme-catalyzed reactions which are confined inside a small subcellular compartment. This is achieved by applying a mesoscopic version of the quasisteady-state assumption to the exact Fokker–Planck equation associated with the Poisson representation of the chemical master equation. The procedure yields impressively simple and compact expressions for the finite-volume corrections. We prove that the predictions of the rate equations will always underestimate the actual steady-state substrate concentrations for an enzyme-reaction network confined in a small volume. In particular we show that the finite-volume corrections increase with decreasing subcellular volume, decreasing Michaelis–Menten constants, and increasing enzyme saturation. The magnitude of the corrections depends sensitively on the topology of the network. The predictions of the theory are shown to be in excellent agreement with stochastic simulations for two types of networks typically associated with protein methylation and metabolism. © *2010 American Institute of Physics*.
[doi:10.1063/1.3505552]

I. INTRODUCTION

Recent years have seen a distinctive surge in the formulation and application of stochastic models of biochemical reaction kinetics. This trend has resulted from a deeper, ongoing appreciation of the conditions characteristic of the intracellular environment[1] and of their dissimilarity from *in vitro* conditions. Typical *in vivo* concentrations are in the range of nanomolar to millimolar; such concentrations realized in a macroscopic volume imply very large copy numbers of interacting molecules whereas the same concentrations in the small volume of a cell frequently imply copy numbers ranging from few tens to at most few thousands (for a detailed experimental protein abundance study see for example Ref. 2). Reaction kinetics is inherently a stochastic process;[3] this noisiness is not apparent in macroscopic conditions due to an implicit averaging over a very large number of molecules but cannot be overlooked when we are studying the kinetics of a system in which the copy number of at least one species is small. This is frequently the case of intracellular kinetics.

The introduction of the stochastic simulation algorithm by Gillespie[4] has popularized the numerical study of stochastic reaction kinetics. However to-date the analytical study of the properties of such systems has received comparatively very little attention principally because the mathematical formalism of stochastic kinetics [i.e., chemical master equations, (CMEs)] is very different than that of deterministic kinetics [i.e., rate equations (REs) which are based on ordinary differential equations] and is less amenable to analysis. This problem is augmented by the fact that many biological networks of interest are considerably large.

One of the main analytical methods for systematically exploring the stochastic properties of these networks has been the linear-noise approximation.[5,6] The advantage of this method is the relative ease with which one can compute the magnitude of intrinsic noise (i.e., coefficients of variation and Fano factors). The major drawback is that the linear-noise approximation gives only meaningful results provided the copy number of molecules is not small or to be more precise it is correct in the limit of infinitely large reaction volumes, i.e., the same limit in which the REs are valid. Since intracellular reactions occur in the opposite limit of small volumes, it is highly desirable to calculate the finite-volume corrections to the concentrations and moments of intrinsic noise.

Developing a theory of finite-volume corrections presents a considerable analytical challenge. One way to obtain the latter is via the system-size expansion of the CME.[5] The linear-noise approximation comes about by evaluating the first term (of order V^0 where V is the reaction volume) in this

a)Electronic mail: ramon.grima@ed.ac.uk.

expansion. The next term is proportional to $V^{-1/2}$ and hence consideration of this term will necessarily give a finite-volume correction. Grima[7,8] calculated the first such corrections for the mean concentrations of species involved in single-substrate enzyme reactions and recently also for a general chemical reaction network of arbitrary complexity.[9] The general result is that the kinetics of a reaction pathway confined in a volume V can be described by the usual REs plus new terms which are proportional to V^{-1}. These equations are referred to as effective mesoscopic rate equations (EMREs). The differences between the solution of EMREs and the corresponding REs for finite volumes stems from a coupling between the mean concentrations and the fluctuations about them. EMREs can be explicitly solved for pathways characterized by a handful of chemical species, but otherwise one has to resort to numerical solution. In the latter cases, one obtains a solution at the expense of losing the insight which typically comes from analytical results.

In this article we develop an alternative, powerful method of calculating finite-volume corrections to the solutions of the REs. Note that in the context of this article, finite-volume corrections exclusively refer to corrections to the mean concentrations not to the moments of intrinsic noise. The method is based on the Poisson representation of the CME rather than the system-size expansion used in the derivation of EMREs. This method unlike the system-size expansion has been applied to study systems of biochemical or biological relevance in only a handful of cases (see for example Refs. 10 and 11) but as we shall show it is a tool with great potential for this field. We focus on chemical reaction networks which are composed of enzyme-catalyzed reactions, a commonly encountered case in intracellular biochemistry.[12] We show that when the timescales of complex and substrate fluctuations are well-separated, it is possible to obtain explicit and impressively simple equations for the finite-volume corrections. For simple reactions these corrections are shown to be the same as given by the EMRE. The distinct advantage of the new method over the EMRE is that it provides analytically simple results even for complex networks with hundreds or thousands of species. This is inherently possible because of the large reduction in the effective dimensionality of the CME when timescales are well separated.

The paper is organized as follows. In Sec. II, we derive the Poisson representation for a general enzyme-reaction network and use the resulting Fokker–Planck equation (FPE) to obtain an exact Liouville equation encoding all information about deviations from the deterministic solution of the REs. In Sec. III we show that in the limit of well-separated timescales of complex and substrate species, the Liouville equation simplifies to a compact approximate form. This is used in Sec. IV to compute explicit expressions for the finite-volume corrections of a general network. In the latter section we show that the corrections for a simple Michaelis–Menten type reaction agree with those previously derived using the EMRE formalism. More importantly we apply the theoretical results to two common types of large-scale networks and confirm the predictions using simulations. We finish by a discussion in Sec. V.

II. THE POISSON REPRESENTATION FOR THE ENZYME REACTION NETWORK

In this section we use the Poisson representation to derive a general FPE for an enzyme reaction network. The latter while being exactly equivalent to the CME is much more amenable to analysis and hence is a very convenient starting point for detailed calculation purposes.

We consider a generic type of enzyme network composed of two major types of chemical processes: (i) the input of a substrate species A_0 into a subcellular compartment; (ii) the transformation of A_0 into some final product A_N via N consecutive enzyme-catalyzed reactions of the type

$$E_i + A_i \underset{k_{-1}^i}{\overset{k_1}{\rightleftharpoons}} C_i \xrightarrow{k_2^i} E_i + A_{i+1}, \tag{1}$$

where A_i, C_i, and E_i denote the ith substrate, complex, and enzyme species, respectively; the index i takes values from 0 to $N-1$ and the k's denote the relevant macroscopic rate constants. Note that we have assumed here that the bimolecular reaction rate, k_1, is the same for all substrates. Note also that in such types of networks there are N distinct substrate species, an equal number of distinct complex species and a number of enzyme species varying between 1 and N. The freedom in choosing the number of enzyme species comes from the fact that an enzyme can generally bind to more than one type of substrate.

If we assume that we have well-mixed conditions inside the compartment then the instantaneous description of the state of a chemical system at time t is simply given by the vector of the absolute number of molecules of each species, $\boldsymbol{n}=(\{n_{A_i}\},\{n_{E_i}\},\{n_{C_i}\})$. Since the mesoscopic kinetics are stochastic, a full description of the system is necessarily probabilistic and is achieved by defining the probability density function $P=P(\boldsymbol{n},t)$ and its time-evolution equation, which is commonly referred to as the CME (Refs. 5 and 13)

$$\partial_t P = \sum_{i=0}^{N-1} \Pi_i P + \Xi P, \tag{2}$$

$$\Pi_i = \frac{k_1}{V}(\Theta_{A_i}\Theta_{C_i}^{-1} - 1)n_{A_i}n_{E_i} + k_{-1}^i(\Theta_{C_i}\Theta_{A_i}^{-1} - 1)n_{C_i} + k_2^i(\Theta_{C_i}\Theta_{A_{i+1}}^{-1} - 1)n_{C_i}, \tag{3}$$

$$\Xi = k_{\text{in}}V(\Theta_{A_0}^{-1} - 1), \tag{4}$$

where V is the compartment volume, Ξ is the contribution due to the input of substrate species A_0 into the system at a rate k_{in} while Π_i describes the ith catalytic reaction step, Eq. (1), in the reaction network. The CME is compactly expressed using Van Kampen's step operators defined as $\Theta_{X_i}^{\pm 1} g(n_{X_i}) = g(n_{X_i} \pm 1)$.

Substantially, the CME depends only on the set of variables $\{n_{A_i}\}$ and $\{n_{C_i}\}$ since an enzyme molecule can be either in the free state or in the complexed state and hence the variables $\{n_{E_i}\}$ are redundant. We can express this conservation law by writing $n_{E_i} = E_T^i - \Sigma_j G_{ij} n_{C_j}$ where the matrix G_{ij} is defined by construction to be

$$G_{ij} = (\hat{\boldsymbol{e}}_i)_j = \begin{cases} 1 & \text{enzyme } i \text{ binds substrate } j \\ 0 & \text{otherwise.} \end{cases} \tag{5}$$

The N-dimensional vector $\hat{\boldsymbol{e}}_i$ is associated with the enzyme binding substrate in the ith catalytic reaction. Its jth entry is chosen to be equal to one if the enzyme can form a complex with the jth substrate and zero otherwise. Hence the connectivity of the network is explicitly encoded in the form of these vectors.

We define the moment generating function, parameterized by the vector of continuous variables, $z=(\{z_{A_i}\},\{z_{C_i}\})$, as

$$G(z) = \sum_{\boldsymbol{n}} \prod_i z_{A_i}^{n_{A_i}} z_{C_i}^{n_{C_i}} P(\boldsymbol{n}). \tag{6}$$

Multiplying the CME, Eq. (2), by $\Pi_i z_{A_i}^{n_{A_i}} z_{C_i}^{n_{C_i}}$, performing the summation over all values of the variables $\boldsymbol{n}$ and expressing the resulting equation in terms of the moment generating function $G(z)$, we obtain the moment generating function equation

$$\partial_t G(z) = \left(k_1 \sum\nolimits_i \mathcal{R}_i^G + \mathcal{S}^G\right) G, \tag{7}$$

$$\mathcal{S}^G = k_{\text{in}} V (z_{A_0} - 1), \tag{8}$$

$$\mathcal{R}_i^G = (z_{C_i} - z_{A_i})(\hat{\mathcal{E}}_i \partial_z^{A_i} - K_M^i \partial_z^{A_i}) + K_2^i (z_{A_{i+1}} - z_{A_i}) \partial_z^{C_i}, \tag{9}$$

where $\hat{\mathcal{E}}_i$ abbreviates $[E_T^i] - V^{-1} \Sigma z_{C_i} \partial_C^i$ and $[E_T^i] = E_T^i / V$ is the total enzyme concentration associated with the enzyme binding substrate in the ith catalytic reaction step. Furthermore we set $K_1^i = k_{-1}^i / k_1$, $K_2^i = k_2^i / k_1$, $K_M^i = K_1^i + K_2^i$ (commonly referred to as the Michaelis–Menten constant), and $\partial_z^X = \partial / \partial z_X$. Note that in Eq. (9) all derivatives are to the right.

We proceed by making use of Gardiner's Poisson representation. At the heart of this method is the assumption that the probability density function $P(\mathbf{n},t)$ can be expanded as a superposition of multivariate uncorrelated Poissons[13,14]

$$P(\boldsymbol{n},t) = \int \mathrm{d}\boldsymbol{\alpha} \prod_i \frac{e^{-\alpha_{A_i} V} (\alpha_{A_i} V)^{n_{A_i}}}{n_{A_i}!} \frac{e^{-\alpha_{C_i} V} (\alpha_{C_i} V)^{n_{C_i}}}{n_{C_i}!} f(\boldsymbol{\alpha},t), \tag{10}$$

where the function $f(\boldsymbol{\alpha},t)$ is usually referred to as the quasi-probability density function. The vector $\boldsymbol{\alpha}$ is defined to be $(\{\alpha_{A_i}\},\{\alpha_{C_i}\})$. It has been shown that this superposition always exists if the range of $\boldsymbol{\alpha}$ is extended to the complex plane by analytic continuation of the Poisson kernel. We have explicitly introduced V in our superposition definition, which is not customarily done in the Poisson representation. In doing so, we obtain the representation in intensive variables, i.e., in units of concentrations, as those encountered in the theory of REs. The above expansion is equivalent to writing the moment generating function as

$$G(z,t) = \int \mathrm{d}\boldsymbol{\alpha} f(\boldsymbol{\alpha},t) e^{\Sigma(z_{A_i}-1)\alpha_{A_i} V} e^{\Sigma(z_{C_i}-1)\alpha_{C_i} V}. \tag{11}$$

It follows from Eqs. (7) and (11) (see Appendix A for details)

$$-\partial_t f = (\mathcal{R} + \mathcal{S}) f, \tag{12}$$

$$\mathcal{S} = k_{\text{in}} \partial_A^0, \tag{13}$$

$$\mathcal{R} = k_1 \sum_{i=0}^{N-1} \mathcal{R}_i,$$

$$\mathcal{R}_i = (\partial_C^i - \partial_A^i)(\alpha_{A_i} \mathcal{E}_i - K_M^i \alpha_{C_i}) + K_2^i (\partial_A^{i+1} - \partial_A^i) \alpha_{C_i}, \tag{14}$$

where we have utilized the notation $\nu = V^{-1/2}$, $\partial_X^i = \partial / \partial \alpha_{X_i}$. Note that $\alpha_{A_N} = \alpha_P$ is the variable for the product formed after N catalytic steps. Note also that $\mathcal{E}_i$ is not a constant, but is given by the operator

$$\mathcal{E}_i = [E_T^i] - \sum\nolimits_j G_{ij} (1 - \nu^2 \partial_{C_j}) \alpha_{C_j} \tag{15}$$

which is essentially the Poisson representation of the conservation of total enzyme molecules. The above equation generally differs from the corresponding deterministic conservation law in terms of concentrations; the latter is described only by its average, while the former exhibits finite-volume corrections due to the finite copy number of enzyme molecules. Given Eq. (15), we see that the Poisson representation, Eq. (12), yields a FPE in terms of substrate and complex variables. Note also that for the case of $N=1$ we obtain the representation for the single-substrate single-enzyme reaction, which is usually referred to as the Michaelis–Menten reaction.

This completes the derivation of the Poisson representation of our general enzyme-reaction network. Note that this is not the same FPE as that which arises from the system-size expansion method of Van Kampen.[5] In the latter case, the FPE is an approximation to the CME in the limit of large volumes whereas the FPE obtained from the Poisson representation is exactly equivalent to the CME.

A further boon of the Poisson representation is that once we have calculated the moments of the continuous variables, α_{X_i}, using the FPE, we can very easily find the corresponding moments of the copy number of molecules, n_{X_i}, using the following simple relationships:[14]

$$\frac{\langle n_{X_i} \rangle}{V} = \langle \alpha_{X_i} \rangle, \tag{16}$$

$$\frac{\langle n_{X_i} n_{X_j} \rangle}{V^2} = \langle \alpha_{X_i} \alpha_{X_j} \rangle + \frac{1}{V} \delta_{i,j} \langle \alpha_{X_i} \rangle, \tag{17}$$

where the angled brackets imply the statistical average: on the left hand side of the above equations these are given by $\langle .. \rangle = \int \mathrm{d}\boldsymbol{n} \, .. P(\boldsymbol{n},t)$ while those on the right hand side imply $\langle .. \rangle = \int \mathrm{d}\boldsymbol{\alpha} \, .. f(\boldsymbol{\alpha},t)$.

A. The mesoscopic equation

The kinetics becomes deterministic in the macroscopic limit of infinitely large volumes. This can be easily verified by noting that all second-order derivatives in the FPE are multiplied by a factor proportional to the inverse of the volume. To compute the finite-volume corrections we will need to separate the mesoscopic and macroscopic evolution equations. We now show that this can be done by applying a

suitable change of variables to the Poisson representation. The macroscopic corresponds to a shot noise contribution, which agrees on average with the mean-field result i.e., with the solution of the REs for the substrate-enzyme network. The mesoscopic contribution reflects the nonequilibrium properties of the network due to the bimolecular character of the substrate-enzyme interaction, and depends parametrically on the mean-field expectation.

We express the deviation from the deterministic path by the following change of variables:

$$(\boldsymbol{\alpha}_{A_i}, \boldsymbol{\alpha}_{C_i}) \to ([A_i](t) + \nu\epsilon_{A_i}, [C_i](t) + \nu\epsilon_{C_i}). \tag{18}$$

We shall refer to ϵ_{X_i} as the mesoscopic correction to the deterministic, macroscopic concentration $[X_i]$ of species X_i. Note that the above equation has the same apparent form as the Van Kampen (VK) ansatz,[5,6] at the heart of the system-size expansion, but the context of the application is completely different. The VK ansatz is applied to the integer number of particles in the CME leading to an infinite series in powers of the inverse square root of the volume, which has to be truncated; the first term of this expansion (the one proportional to V^0) gives a linear FPE which is an approximation to the CME. In our case the change of variables, Eq. (18), is applied on the FPE arising from the Poisson representation which leads to a finite series and allows for exact analytical treatment; as we shall show now, this divides the exact FPE into a macroscopic term and a term which captures *all* deviations from the macroscopic.

The transformation applied to the FPE transforms the time derivative into

$$\left.\frac{\partial}{\partial t}\right|_{\boldsymbol{\alpha}} f(\boldsymbol{\alpha},t) = \left.\frac{\partial}{\partial t}\right|_{\epsilon} f([\boldsymbol{X}] + \nu\boldsymbol{\epsilon}, t) + \left.\frac{d\boldsymbol{\epsilon}}{dt}\right|_{\boldsymbol{\alpha}} \cdot \nabla_{\epsilon} f([\boldsymbol{X}] + \nu\boldsymbol{\epsilon}, t), \tag{19}$$

where $\partial/\partial_t|_x$ denotes taking the derivative with x held constant. It follows from Eq. (18) that $d\boldsymbol{\epsilon}/dt|_{\boldsymbol{\alpha}} = -\nu^{-1} d[\boldsymbol{X}]/dt$. Finally by expressing the right hand side of Eq. (12) in terms of the new variables $\boldsymbol{\epsilon}$ and equating the result to Eq. (19), we find that the FPE takes the form

$$-\frac{\partial}{\partial t} g(\boldsymbol{\epsilon}) = k_1 \mathcal{L} g(\boldsymbol{\epsilon}) + \nu^{-1}\left(k_1 \mathcal{R}_{\text{macro}} + \mathcal{S}_{\text{macro}} - \frac{\partial[\boldsymbol{A}]}{\partial t} \cdot \nabla_{\epsilon_A} - \frac{\partial[\boldsymbol{C}]}{\partial t} \cdot \nabla_{\epsilon_C}\right) g(\boldsymbol{\epsilon}), \tag{20}$$

where the relevant operators are

$$\mathcal{S}_{\text{macro}} = k_{\text{in}} \frac{\partial}{\partial \epsilon_{A_0}}, \tag{21}$$

$$\mathcal{R}_{\text{macro}} = \sum_i ([E_i][A_i] - K_M^i [C_i]) \frac{\partial}{\partial \epsilon_{C_i}} + \sum_i (K_1^i [C_i] - [E_i][A_i]) \frac{\partial}{\partial \epsilon_{A_i}}, \tag{22}$$

$$\mathcal{L} = \sum_i \{(\partial_C^j - \partial_A^j)([E_i]\epsilon_{A_i} + [A_i]\delta\mathcal{E}_i + \nu\epsilon_{A_i}\delta\mathcal{E}_i - K_M^i \epsilon_{C_i}) + K_2^i(\partial_A^{j+1} - \partial_A^j)\epsilon_{C_i}\}. \tag{23}$$

Note that the new probability density function necessarily satisfies $g(\boldsymbol{\epsilon})\mathrm{d}\boldsymbol{\epsilon} = f(\boldsymbol{\alpha})\mathrm{d}\boldsymbol{\alpha}$. Note also that the notation ∂_X^j now denotes the derivative $\partial/\partial\epsilon_{X_i}$. The quantity $[E_i] = [E_T^i] - \Sigma_j G_{ij}[C_j]$ is the macroscopic concentration of the free enzyme species associated with binding substrate in the ith catalytic reaction step. The operator $\delta\mathcal{E}_i = -([E_i] - \mathcal{E}_i)/\nu = -\Sigma_j G_{ij}(\epsilon_{C_j} - \partial_C^j[C_j] - \nu\partial_C^j\epsilon_{C_j})$ is the contribution of the enzyme-operator Eq. (15).

Note that the resulting form of the FPE is clearly divided into two parts. In the macroscopic limit the terms proportional to ν^{-1} dominate and their sum must equate to zero; this leads to the macroscopic equations. It also then follows that the mesoscopic equation is simply given by

$$\partial_\tau g(\boldsymbol{\epsilon}, t) = -\mathcal{L} g(\boldsymbol{\epsilon}, t), \tag{24}$$

where we have renormalized time to $\tau = k_1 t$. Note also that due to the Poissonian nature of the substrate input process, it only contributes to the macroscopic part, a feature which is unique to the Poisson representation. We emphasize that up until this point, we have made no approximations and hence the resulting mesoscopic equation is exact.

III. ADIABATIC ELIMINATION OF THE COMPLEX SPECIES VARIABLES

In this section we show how to rigorously eliminate the fast variables from our description. This will be done in two steps: on the macroscopic contribution of the FPE and on the mesoscopic contribution including terms up to order ν (i.e., finite-volume corrections). As we shall see later on, the reduced mesoscopic description does depend on the reduced macroscopic description and hence the need to treat the latter first.

A. The reduced macroscopic equations

The macroscopic equations are obtained from Eq. (20) by taking the limit $\nu \to 0$, leading to

$$\begin{aligned} \frac{d[A_i]}{dt} &= k_{-1}^i [C_i] - k_1 [E_i][A_i] + \delta_{i,0} k_{\text{in}}, \\ \frac{d[C_i]}{dt} &= k_1 [E_i][A_i] - (k_2^i + k_{-1}^i)[C_i], \\ \frac{d[A_N]}{dt} &= k_2^{N-1}[C_{N-1}], \\ [E_i] &= [E_T^i] - \sum_j G_{ij}[C_j]. \end{aligned} \tag{25}$$

These agree exactly with those that can be obtained from the RE approach. The elimination of the complex species from the macroscopic equations is a well-known procedure commonly referred to as the quasisteady-state approximation (QSSA).[15,16] Briefly speaking the approximation is tanta-

mount to assuming that the complex equilibrates on a much shorter timescale than the substrate. This is implemented by imposing the approximation $d[C_i]/dt=0$ on the macroscopic equations. The resulting reduced equations (commonly referred to as the Briggs–Haldane equations) are then given by replacing $[C_i]$ in the full time-evolution equations for the substrate concentrations by

$$[C_i] = \frac{[E_i][A_i]}{K_M^i}. \tag{26}$$

B. The reduced mesoscopic equation

Now we are interested in deriving the reduced mesoscopic equation corresponding to the reduced macroscopic equations that we just considered. Timescale separation on the mesoscopic scale is nontrivial because of the inherent correlations between the mesoscopic fluctuations of the various species. Our presentation shall be as follows. First we shall show that the mesoscopic Liouvillian, Eq. (23), can be generally cast into an asymptotic form of the interaction representation which is typically encountered in the theory of adiabatic elimination of fast fluctuating variables.[17,18] The latter yields a particularly simple result for the reduced mesoscopic equation.

We will now show that the Liouvillian can be rewritten in the general form

$$\mathcal{L}(\gamma) = \gamma\mathcal{L}_1 + \gamma^{1/2}\mathcal{L}_2 + \mathcal{L}_3, \tag{27}$$

where γ^{-1} is the characteristic fast timescale of the complex fluctuations which will be specified later on. We now proceed to derive the operators $\mathcal{L}_1$, $\mathcal{L}_2$, and $\mathcal{L}_3$ for the general enzyme-reaction network under study. We start by grouping all terms containing only the pair of complex variables $(\epsilon_{C_i}, \partial_C^i)$ into $\mathcal{L}_1$, terms concerning solely the substrate ones $(\epsilon_{A_i}, \partial_A^i)$ into $\mathcal{L}_3$, while treating the remaining terms as interaction $\mathcal{L}_2$. Thus we have

$$\mathcal{L}_1 = \sum_i \mathcal{L}_1^{(i)}, \quad \mathcal{L}_2 = \sum_i \mathcal{L}_2^{(i)}, \quad \mathcal{L}_3 = \sum_i \mathcal{L}_3^{(i)}, \tag{28}$$

$$\mathcal{L}_1^{(i)} = \partial_C^i([A_i]\delta\mathcal{E}_i - K_M^i\epsilon_{C_i}), \tag{29}$$

$$\mathcal{L}_2^{(i)} = -\partial_A^i([A_i]\delta\mathcal{E}_i - K_M^i\epsilon_{C_i}) + K_2^i(\partial_A^{i+1} - \partial_A^i)\epsilon_{C_i} + \nu(\partial_C^i - \partial_A^i)\epsilon_{A_i}\delta\mathcal{E}_i, \tag{30}$$

$$\mathcal{L}_3^{(i)} = -\partial_A^i[E_i]\epsilon_{A_i}. \tag{31}$$

It is instructive to rescale all complex variables by their characteristic timescale γ

$$z_i = \gamma^{1/2}\epsilon_{C_i}, \quad x_i = \epsilon_{A_i}. \tag{32}$$

The operator $\mathcal{L}_3$ is trivially obtained

$$\mathcal{L}_3^{(i)} = -\partial_x^i[E_i]x_i. \tag{33}$$

Note that ∂_x^i denotes the derivative $\partial/\partial x_i$ whereas ∂_X^i stands for the derivative $\partial/\partial\epsilon_{X_i}$. Next we observe that the enzyme operator transforms as

$$\gamma^{1/2}\delta\mathcal{E}_i = -\sum_j G_{ij}(z_j - \gamma\partial_z^j[C_j] - \nu\gamma^{1/2}\partial_z^j z_j). \tag{34}$$

Plugging this into Eq. (29) and putting $\mathcal{L}_1^{(i)} \to \gamma\mathcal{L}_1^{(i)}$ we find that the last term in Eq. (34) can be asymptotically neglected. Hence the dominant contribution in the limit of large γ is given by

$$\mathcal{L}_1^{(i)} = -\partial_z^i\sum\nolimits_j M_{ij}z_j + \partial_z^i\sum\nolimits_j D_{ij}\partial_z^j, \tag{35}$$

where we have utilized the abbreviations

$$J_{ij} = ([A_i]G_{ij} + K_M^i\delta_{ij}), \quad M_{ij} = \gamma^{-1}J_{ij}, \quad D_{ij} = G_{ij}[A_i][C_j]. \tag{36}$$

Thus the asymptotic form of the complex fluctuations is described by an Ornstein–Uhlenbeck process, centered on the deterministic expectation Eq. (26). It is clear, by virtue of the fluctuation-dissipation theorem, that the matrix $\underline{J}$ (constant matrices, i.e., those independent of the ϵ_{A_i} and ϵ_{C_i} variables and of the associated partial derivatives are underlined throughout the rest of the article) must correspond to the Jacobian of the complex species while the matrix $\underline{D}$ determines the strength of the fluctuations in the Poisson representation. Note that throughout the article by Jacobian of a species we mean the negative of the Jacobian matrix as obtained from the macroscopic rate equations of that species with the time scaling $\tau=k_1t$.

The characteristic timescale γ^{-1} can be inferred from the explicit form of the Jacobian. We choose $\gamma \equiv \mathrm{tr}(\underline{J})$ giving the relaxation rate of the complex vector $\boldsymbol{z}$. Consequently $\underline{M}$ is finite, as required by the stability of $\mathcal{L}_1$, and is given by the Jacobian of unit trace. Equally we observe that the Jacobian of the substrate vector $\boldsymbol{x}$ is given by the diagonal matrix $\underline{E}=\mathrm{diag}([E_i])$, such that its relaxation rate is proportional to its trace. Thus under the condition where $\gamma \gg \mathrm{tr}(\underline{E})$, we can assume that the timescales of substrate and complexes are well-separated. The result can be interpreted as follows: $\mathcal{L}_1$ represents the fast nonequilibrium fluctuations of the complex and $\mathcal{L}_3$ reflects the slow Poissonian decay of the substrate fluctuations.

Finally we turn our attention to the interaction, we put $\mathcal{L}_2 \to \gamma^{1/2}\mathcal{L}_2$ and rearrange to emphasize its dependence on the complex variables

$$\mathcal{L}_2^{(i)} = \sum\nolimits_j \chi_{ij}z_j + \sum\nolimits_j \varphi_{ij}\partial_z^j,$$

$$\chi_{ij} = \partial_x^j M_{ij} + \frac{1}{\gamma}(\delta_{ij}K_2^i(\partial_x^{i+1} - \partial_x^i) + \nu\partial_x^j x_i),$$

$$\varphi_{ij} = (\delta_{ij}[E_i]x_i - \partial_x^j D_{ij} - \nu\partial_x^j x_i G_{ij}[C_j]). \tag{37}$$

For notational convenience, we have considered only terms proportional to $\boldsymbol{z}$ and ∇_z, namely, those which give nonvanishing contributions in the following. We emphasize that Eqs. (33) and (37), together with Eq. (35) yield the contribution to the mesoscopic equation under the condition of well-separated substrate- and complex timescales.

It is clear that the reduced substrate density function $h(\boldsymbol{x})=\int \mathrm{d}\boldsymbol{z} g(\boldsymbol{x},\boldsymbol{z})$ cannot simply be deduced from $\mathcal{L}_3$, but

must retain a contribution from the inherent coupling to the rapid fluctuations of the complex species. However, it is known[13] that in the limit of well-separated timescales, a particularly simple result can be obtained by projecting the evolution equation on the steady-state of the fast variables. The technique is the adiabatic elimination of the fast variables and can be accomplished by the use of the projector $Pg = \pi(z)\int dz g(\boldsymbol{x},z)$, with π being the steady-state distribution of the fast variable vector $\boldsymbol{z}$, determined by the steady-state condition $\mathcal{L}_1\pi=0$.

The applicability of the method follows directly from the asymptotic form Eq. (27), which has been derived here explicitly for the large-scale enzyme-reaction networks under consideration, together with the conditions $P\mathcal{L}_3=\mathcal{L}_3P$, $P\mathcal{L}_1=\mathcal{L}_1P=0$, $P\mathcal{L}_2P=0$,[13] which can be easily verified from Eqs. (33), (35), and (37). The result in the limit of large γ has been formally derived by Gardiner[13] and reads

$$-\partial_t(Pg) = \lim_{\gamma\to\infty} P\mathcal{L}(\gamma)g = (\mathcal{L}_3 - P\mathcal{L}_2\mathcal{L}_1^{-1}\mathcal{L}_2)(Pg). \tag{38}$$

Integration over $\boldsymbol{z}$ yields the reduced evolution equation

$$-\partial_t h(\boldsymbol{x},t) = \mathcal{L}'h(\boldsymbol{x},t), \tag{39}$$

$$\mathcal{L}' = \lim_{\gamma\to\infty} \mathcal{L}(\gamma) = \mathcal{L}_3 - \langle \mathcal{L}_2\mathcal{L}_1^{-1}\mathcal{L}_2\rangle_\pi. \tag{40}$$

The mesoscopic description is now reduced to that of the substrate only. The angled brackets $\langle\cdot\rangle_\pi$ denote the trace $\int d\boldsymbol{z}\cdot\pi(z)$ taken over the rapid steady-state fluctuations of the complex species. We now substitute Eq. (37) into Eq. (40)

$$\mathcal{L}' = \mathcal{L}_3 - \sum_{ij}(\chi\langle z\mathcal{L}_1^{-1}z^T\rangle_\pi\chi^T)_{ij} - \sum_{ij}(\chi\langle z\mathcal{L}_1^{-1}\nabla_z^T\rangle_\pi\varphi^T)_{ij}. \tag{41}$$

The explicit form of the correlators involved has been derived in Appendix B using the steady-state condition. The result is

$$-\langle z\mathcal{L}_1^{-1}z^T\rangle_\pi = \underline{\mathrm{S}} \equiv \underline{\mathrm{M}}^{-1}\underline{\mathrm{D}}^T\underline{\mathrm{M}}^{-T}, \quad -\langle z\mathcal{L}_1^{-1}\nabla_z^T\rangle_\pi = \underline{\mathrm{M}}^{-1}. \tag{42}$$

Hence using Eq. (41) together with Eq. (42), we get the final form for the reduced Liouvillian in a convenient matrix form, which yields the mesoscopic implementation of the QSSA

$$\mathcal{L}' = \mathcal{L}_3 + \sum_{ij}(\chi\underline{\mathrm{S}}\chi^T)_{ij} + \sum_{ij}(\chi\underline{\mathrm{M}}^{-1}\varphi^T)_{ij}. \tag{43}$$

IV. FINITE-VOLUME CORRECTIONS

In this section we will use the reduced mesoscopic description to compute the finite-volume corrections to the macroscopic steady-state concentrations of all substrate species in the network. By definition, the concentration of species A_α according to the stochastic model is given by $\rho_\alpha = \langle n_\alpha/V\rangle$. Using Eq. (16) together with Eq. (18) it is straightforward to show that

$$\rho_\alpha = [A_\alpha] + \nu\langle x_\alpha\rangle. \tag{44}$$

Hence the size of the finite-volume correction to the macroscopic concentration of species A_α is simply given by $\langle x_\alpha\rangle$. The time-evolution equation for the latter can be computed directly from our reduced mesoscopic equation, Eq. (39) together with Eq. (43), yielding

$$-\partial_\tau\langle x_\alpha\rangle = \int d\boldsymbol{x}\, x_\alpha\mathcal{L}'h(\boldsymbol{x}). \tag{45}$$

We evaluate the corrections associated with the substrate, $\alpha<N$, to leading order in x and ν. Defining the transfer matrix $\underline{\mathrm{T}}_{\alpha,i}=K_2^i(\delta_{i+1,\alpha}-\delta_{i\alpha})$, one can show the following results:

$$\int d\boldsymbol{x}\, \boldsymbol{x}\mathcal{L}_3h = \underline{\mathrm{E}}\langle\boldsymbol{x}\rangle, \tag{46}$$

$$-\int d\boldsymbol{x}\, \boldsymbol{x}\sum_{ij}(\chi\underline{\mathrm{M}}^{-1}\varphi^T)_{ij}h = (\underline{1}+\underline{\mathrm{T}}\underline{\mathrm{M}}^{-1})\underline{\mathrm{E}}\langle\boldsymbol{x}\rangle + \frac{\nu}{\gamma}\underline{\mathrm{M}}\boldsymbol{S}, \tag{47}$$

$$\int d\boldsymbol{x}\, \boldsymbol{x}\sum_{ij}(\chi\underline{\mathrm{S}}\chi^T)_{ij}h = \frac{\nu}{\gamma}(\underline{\mathrm{M}}+\underline{\mathrm{T}})\boldsymbol{S}, \tag{48}$$

where $\boldsymbol{S}$ is the vector obtained by summing over all rows of the matrix $\underline{\mathrm{S}}$, i.e., $(\boldsymbol{S})_i=\Sigma_j S_{ij}^T$. Summing up these equations, we obtain

$$\partial_\tau\langle\boldsymbol{x}\rangle = \underline{\mathrm{T}}\left(\underline{\mathrm{M}}^{-1}\underline{\mathrm{E}}\langle\boldsymbol{x}\rangle - \frac{\nu}{\gamma}\boldsymbol{S}\right) \tag{49}$$

which has the steady-state solution $\langle\boldsymbol{x}\rangle=(\nu/\gamma)\underline{\mathrm{E}}^{-1}\underline{\mathrm{M}}\boldsymbol{S}$. Note that the result is virtually independent of γ, if we rewrite

$$\langle x_\alpha\rangle = \nu\sum_i(\underline{\mathrm{E}}^{-1}\underline{\mathrm{D}}\underline{\mathrm{J}}^{-T})_{\alpha i}. \tag{50}$$

Hence it is clear that the finite-volume corrections for substrate species are nonzero; this indeed signals the breakdown of the law of mass action on mesoscopic length scales or equivalently for low copy number of molecules.

It can also be shown that the elements of the inverse Jacobian of the complex are given by

$$J_{ij}^{-1} = \frac{1}{K_M^j}\left(\delta_{ij} - \frac{a_iG_{ij}}{1+\hat{\boldsymbol{e}}_i^T\underline{\mathrm{a}}\hat{\boldsymbol{e}}_i}\right), \tag{51}$$

where $\underline{\mathrm{a}}=\mathrm{diag}([A_i]/K_M^i)$ and a_i is the ith diagonal element of the latter matrix with value $[A_i]/K_M^i$ (the reduced macroscopic substrate concentration of species i). Given the definition, Eq. (5), it is straightforward to verify from the above equation that the elements of the inverse Jacobian are always positive-valued. Since both matrices $\underline{\mathrm{D}}$ and $\underline{\mathrm{E}}^{-1}$ are also positive it then follows that the finite-volume corrections to the substrate concentration, Eq. (50), are always positive. In other words, the predictions of the REs will always underestimate the steady-state substrate concentrations for a substrate-enzyme network confined in a small volume.

The equations for the finite-volume corrections can be conveniently expressed in a form which is useful for obtain-

ing insight about the physical origin of the corrections and also for their numerical computation. We write the matrix $\underline{D}=\underline{A}\underline{G}\underline{C}$ where $\underline{A}=\mathrm{diag}([A_i])$ and $\underline{C}=\mathrm{diag}([C_i])$. The Jacobian of the free enzyme is simply the transpose of the Jacobian of the complex such that $(\boldsymbol{J}_E^{-1})_i=\Sigma_j J_{ji}^{-1}$; this can be directly computed using Eq. (51). These two equations together with the steady-state condition, Eq. (26), written here as $\underline{C}=\underline{E}\underline{a}$, the identity $(\underline{E}^{-1}\underline{G}\underline{E})_{\alpha j}=(\hat{\mathbf{e}}_\alpha)_j$ and Eqs. (44) and (50), allow us to write a simple, final equation for the steady-state substrate concentration as predicted by the stochastic model

$$\frac{\rho_\alpha}{[A_\alpha]}=1+R_\alpha, \tag{52}$$

where

$$R_\alpha=\frac{1}{V}\hat{\boldsymbol{e}}_\alpha^T\underline{a}\boldsymbol{J}_E^{-1}. \tag{53}$$

The parameter R_α is proportional to the product of the relaxation time of the enzyme species and the corresponding reduced macroscopic substrate concentration, summed over all substrates which can bind to the same enzyme. It is also straightforward to derive expressions for the relative error made by the RE model, for the absolute differences between the reduced mesoscopic and macroscopic concentrations and for the absolute differences between the sum of the reduced mesoscopic and macroscopic concentrations, respectively

$$R_{\mathrm{error}}^\alpha=\frac{\rho_\alpha-[A_\alpha]}{\rho_\alpha}=\frac{R_\alpha}{1+R_\alpha}, \tag{54}$$

$$\frac{\rho_\alpha-[A_\alpha]}{K_M^\alpha}=R_\alpha a_\alpha, \tag{55}$$

$$\sum_\alpha\frac{\rho_\alpha-[A_\alpha]}{K_M^\alpha}=\sum_\alpha R_\alpha a_\alpha. \tag{56}$$

In the following subsections, we will apply the general results developed so far, to three different cases which are commonly encountered in biology. We will verify the main theoretical predictions, i.e., Eqs. (54)–(56), by comparison with detailed stochastic simulations.

A. Single-substrate reaction with single-enzyme species

The reaction where a single-enzyme species catalyzes only a single-substrate is the classical Michaelis–Menten reaction textbook example

$$\overset{k_{\mathrm{in}}}{\rightarrow}A_0,\quad E_0+A_0\underset{k_{-1}^0}{\overset{k_1}{\rightleftharpoons}}C_0\overset{k_2^0}{\rightarrow}E_0+A_1. \tag{57}$$

The enzyme only binds to a single substrate species and hence $G_{00}=1$. It then follows that Eq. (53) evaluates to

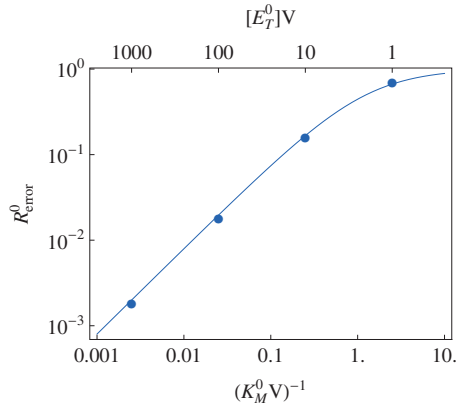

FIG. 1. *Single-substrate, single-enzyme reaction.* Finite-volume scaling of the relative error in the substrate concentration as predicted by the RE model, i.e., $R_{\mathrm{error}}^0=(\rho_0-[A_0])/\rho_0$. The error stems from a finite-volume correction to the mean concentration due to intrinsic noise. The solid line shows the theoretical estimate as given by Eq. (54) together with Eq. (58) in the text. The data points are obtained from simulations carried out using Gillespie's stochastic simulation algorithm under steady-state conditions. The parameters are $K_M^0=1/5$ ($k_1=5$, $k_2^0=0.5$, $k_{-1}^0=0.5$) and $[E_T^0]=0.5$. Timescale separation is guaranteed since $\gamma=([A_0]+K_M^0)=10[E_0]$ and near-saturation conditions ensue since $k_{\mathrm{in}}/(k_2^0[E_T^0])=0.8$ (a value of 1 indicates complete saturation). The upper x-axis indicates the corresponding discrete number of enzyme molecules $[E_T^0]V$ used in the simulation. Note that the agreement between theory and simulation is excellent even when there is just one enzyme molecule in the compartment.

$$R_0=\frac{1}{K_M^0V}\frac{a_0}{(a_0+1)}, \tag{58}$$

where $a_0=[A_0]/K_M^0$. Substituting the above in Eq. (52) gives an expression for the ratio of the mesoscopic and macroscopic substrate concentrations; this agrees exactly with that obtained by taking the limit of well-separated timescales (i.e., $\gamma=([A_0]+K_M^0)\gg[E_0]$) of the finite-volume correction previously calculated by Grima[8] using the system-size expansion including terms of order $V^{-1/2}$. Note that the finite-volume corrections are significant for large steady-state substrate concentrations, i.e., when the enzyme is working in saturated or near-saturated conditions. The predicted dependence of the relative error [i.e., Eq. (54) together with Eq. (58)] on the size of the finite volume was tested using stochastic simulations (Fig. 1). Numerics and theory agree very well over at least three orders of magnitude of the compartment volume, or equivalently over the whole typical physiological range of enzyme copy numbers (from 1 to 1000).

B. Multisubstrate network with single-enzyme species

We now consider a network where different substrates compete for catalysis by a single-enzyme species

$$\overset{k_{\mathrm{in}}}{\rightarrow}A_0,\quad E_0+A_0\underset{k_{-1}^0}{\overset{k_1}{\rightleftharpoons}}C_0\overset{k_2^0}{\rightarrow}E_0+A_1,$$

$$E_0+A_1\underset{k_{-1}^1}{\overset{k_1}{\rightleftharpoons}}C_1\overset{k_2^1}{\rightarrow}E_0+A_2,$$

…

$$E_0 + A_{N-1} \underset{k_{-1}^{N-1}}{\overset{k_1}{\rightleftharpoons}} C_{N-1} \overset{k_2^{N-1}}{\rightarrow} E_0 + A_N. \tag{59}$$

Such type of reactions is commonly associated with the methylations of proteins, DNA, and RNA by the methyltransferase class of enzymes. A strictly unidirectional methylation occurs for instance in the nitrogen trimethylation of Lys.[19] A single amino acid, embedded in a protein, can typically be methylated only a few times (e.g., three times in the case of Lys). But the mere large number of amino acids composing proteins means that the number of methylations can generally be very large. The modification of the protein from the unmethylated (or even partially methylated) form A_0 to the fully methylated form A_N is processed by the enzyme within N catalytic steps. Partially methylated intermediate states are then described by $A_1, \ldots, A_{N-1}$. This type of network has also been proposed as a candidate for explaining the kinetics of protein digestion.[20]

The enzyme can now bind to all the substrate species and hence $G_{ij}=1$. It also follows that since we have one enzyme species then $[E_T^i]=[E_T^0]=[E_T]$. It can then be shown that Eq. (53) evaluates to

$$R_i = \frac{1}{V}\sum_j \frac{1}{K_M^j}\frac{a_j}{a+1}, \tag{60}$$

where $a=\Sigma_i a_i$ is the sum of reduced macroscopic concentrations. To obtain further insight into the above equation and to evaluate Eqs. (54)–(56) it is necessary to obtain expressions for the reduced macroscopic substrate concentrations. Using the REs, Eq. (25), and applying the QSSA to the complex concentration variables we find that the reduced macroscopic equations are

$$\frac{d[A_0]}{dt} = k_{\text{in}} - v^0, \quad \frac{d[A_i]}{dt} = v^{i-1} - v^i, \tag{61}$$

where $v^i=k_2^i[E_T]a_i/(1+a)$, i.e., the catalytic velocity of step i. Imposing the steady-state condition on the macroscopic substrate concentration equations, Eq. (61), we find that the catalytic velocities of all steps must be the same and equal to the input rate, $v^i=k_{\text{in}}$. It follows that the reduced macroscopic substrate concentrations are $a_i=\eta_i(1-\eta)^{-1}$ and their sum is $a=\eta(1-\eta)^{-1}$ where η_i is the dimensionless parameter defined as

$$\eta_i = \frac{k_{\text{in}}}{[E_T]k_2^i} \tag{62}$$

and $\eta=\Sigma_i\eta_i$. Note that η_i is the ratio of the input rate and of the maximum rate at which the enzyme can catalyze substrate A_i into substrate A_{i+1}. Similarly η is the ratio of the input rate and of the overall maximum rate at which the enzyme can catalyze the initial substrate A_0 into the product A_N. Hence both η_i and η have a value between 0 and 1. It follows also that η is a measure of enzyme saturation.

Given the above results one can now evaluate Eqs. (55), (56), and (60) which leads to

$$R_i = \frac{1}{V}\sum_j \frac{\eta_j}{K_M^j}, \tag{63}$$

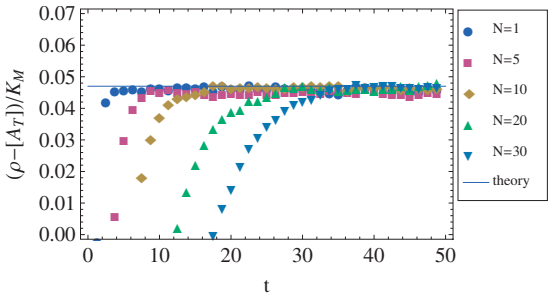

FIG. 2. *Scaling of the finite-volume corrections with network size for a multisubstrate network with single-enzyme species*. Plot of the absolute difference between the sum of the reduced mesoscopic and macroscopic substrate concentrations vs the network size N, for a homogeneous set of rate constants, $k_{-1}^i=k_{-1}^0$, $k_2^i=k_2^0$. Theory predicts that if the input rate is rescaled by the network size, $k_{\text{in}}=k_{\text{in}}^0/N$, then the absolute error in steady-state conditions [as given by Eq. (65)] should be independent of network size. Data points are obtained from stochastic simulations using Gillespie's algorithm (with a rescaling of k_{in} as mentioned above) and solid lines are the theoretical estimates using Eq. (65). Note that in the y-axis label we have used ρ to mean $\Sigma_i\rho_i$ and $[A_T]$ to mean $\Sigma_i[A_i]$. Parameter values are found in the text.

$$\frac{\rho_i - [A_i]}{K_M^i} = R_i\frac{\eta_i}{1-\eta}, \tag{64}$$

$$\sum_i \frac{\rho_i - [A_i]}{K_M^i} = R_i\frac{\eta}{1-\eta}. \tag{65}$$

From Eqs. (52), (62), and (63) one can deduce that: (i) the concentration of each different substrate species is amplified (from the deterministic concentration) by the same factor; (ii) the deviation from the predictions of the deterministic model increases with decreasing volume, decreasing Michaelis–Menten constants, increasing enzyme saturation, and increasing network size N. The latter stems from the fact that increasing the number of substrates will increase competition for the single enzyme species meaning that the copy number of free enzyme molecules at any one time becomes smaller and hence leads to increased noise-induced effects.

We have carried out extensive stochastic simulations using the Gillespie algorithm[4] to test the accuracy of our predictions. The four tests are as follows:

Test 1: Dependence of the finite-volume corrections with network size. We first impose homogeneous rate constants i.e., a fixed set of rate constants for all enzyme reactions in the network. It then follows that if we make the scaling $k_{\text{in}}=k_{\text{in}}^0/N$, the quantity $\eta=\Sigma_i\eta_i$ will be independent of the network size. Consequently under this scaling, R_i is predicted to be independent of the size and so is the sum of the absolute reduced concentrations, Eq. (65). Stochastic simulations show perfect agreement with this prediction (Fig. 2) which implicitly proves that the theory correctly predicts that finite-volume corrections increase with network size. The relevant simulation parameters are: $V=500$, $k_1=1000$, $k_{-1}^i=0.5$, $k_2^i=20$, $(K_MV)^{-1}=0.1$, $[E_T]V=5$, $k_{\text{in}}^0=0.1$, $\eta=0.5$, and $\gamma=10\ \text{tr}(\underline{E})$.

Test 2: Dependence of the relative error with enzyme saturation. We fix the Michaelis–Menten constant, $K_M^i=K_M$, and the size of the network. It then follows from Eqs. (54) and (63) that the relative error is simply given by $R_{\text{error}}^i=\eta/(\eta+K_MV)$, i.e., the relative error increases with increas-

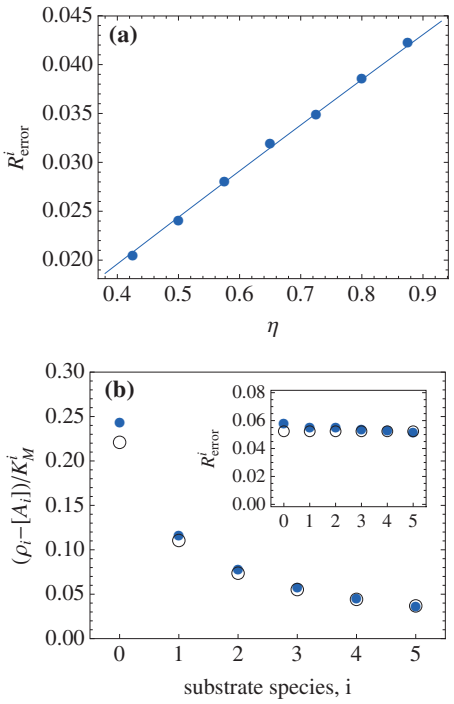

FIG. 3. *Dependence of the deviations from the RE predictions with enzyme saturation and position in the network for a multisubstrate network with single-enzyme species.* In (a) we plot the relative error as a function of η which is a measure of enzyme saturation. In (b) we plot the absolute differences in the reduced mesoscopic and macroscopic substrate concentrations as a function of the substrate species index and hence as a function of position in the six substrate network. The inset shows the relative error for all species. The data points are from stochastic simulation and the solid lines [in (a)] and the open circles [in (b)] are the theoretical predictions. The rate constants are heterogeneous in both cases but the Michaelis–Menten constants for each reaction in the network are the same for (a) while they vary according to position in the network in (b). See the text for parameter values and for a detailed discussion.

ing enzyme saturation. The prediction is verified by stochastic simulation [Fig. 3(a)], where the rate constants were chosen so that $k_2^i = k_2^0$ for i even and $k_2^i = f^{-1}k_2^0$ for i odd and the specific parameter values used were: $V=500$, $k_1=1000$, $N=6$, $f=0.5$, $k_{\text{in}}=1/6$, $k_2^0=10$, $\gamma \geq 10\,\text{tr}(\underline{E})$, $(K_M V)^{-1}=0.05$, $[E_T]V=5$.

Test 3: Dependence of the relative error with species type. As previously mentioned, an inspection of Eq. (63) shows that the quantity R_i and hence the relative error R^i_{error} are predicted to be the same for all species in the network. This agrees with the results of stochastic simulation [inset of Fig. 3(b)] of a six substrate network where $K_M^i=(1+i)K_M^0$, $k_2^i=(1+i)k_2^0$, $k_{-1}^i=(1+i)k_{-1}^0$, and the specific parameter values were: $V=1000$, $k_{\text{in}}=1/60$, $k_2^0=9$, $k_{-1}^0=1$, $k_1=1000$, $[E_T]V=5$.

Test 4: Dependence of the finite-volume correction on the position in the network. Here we test the ability of the theory to predict the absolute difference between the reduced mesoscopic and macroscopic substrate concentrations as predicted by the stochastic and deterministic models, i.e., Eq. (64), for each individual species in the network. For rate constants chosen as in test 3, the differences are predicted to be proportional to $(1+i)^{-1}$, which is confirmed by simulations [Fig. 3(b)]. All parameter values are exactly as in the previous test.

C. Multisubstrate network with multienzyme species

Finally we consider a sequential reaction network which is typically associated with metabolism[21]

$$\overset{k_{\text{in}}}{\rightarrow} A_0, \quad E_0 + A_0 \underset{k_{-1}^0}{\overset{k_1}{\rightleftharpoons}} C_0 \overset{k_2^0}{\rightarrow} E_0 + A_1,$$

$$E_1 + A_1 \underset{k_{-1}^1}{\overset{k_1}{\rightleftharpoons}} C_1 \overset{k_2^1}{\rightarrow} E_1 + A_2,$$

$$\cdots$$

$$E_{N-1} + A_{N-1} \underset{k_{-1}^{N-1}}{\overset{k_1}{\rightleftharpoons}} C_{N-1} \overset{k_2^{N-1}}{\rightarrow} E_{N-1} + A_N. \tag{66}$$

This network is characterized by a different enzyme species for each separate catalytic step and thus $G_{ij}=\delta_{ij}$. The finite-volume corrections can be easily derived from Eq. (53), since the Jacobian of the complex species is instantly diagonal and hence the result is

$$R_i = \frac{1}{K_M^i V}\frac{a_i}{(a_i+1)}. \tag{67}$$

It is useful to write the above expression in terms of the constants defining the network by using the REs (with the QSSA applied to the complex concentration variables)

$$\frac{d[A_0]}{dt} = k_{\text{in}} - v_M^0, \quad \frac{d[A_i]}{dt} = v_M^{i-1} - v_M^i, \tag{68}$$

where $v_M^i = k_2^i[E_T^i]a_i/(1+a_i)$ is the catalytic velocity in this case. The connectivity of the network requires that all catalytic velocities must be the same and equal to the input rate, $v_M^i = k_{\text{in}}$, from which it follows that the reduced macroscopic concentrations are

$$a_i = \frac{\eta_i}{1-\eta_i}, \tag{69}$$

where $\eta_i = k_{\text{in}}/k_2^i[E_T^i]$, a measure of saturation as before. Hence it follows that Eq. (67) reduces to the simple form

$$R_i = \frac{1}{K_M^i V}\eta_i. \tag{70}$$

It can be deduced from the above equation that the deviation from the predictions of the deterministic model increases with decreasing volume, decreasing Michaelis–Menten constants and increasing enzyme saturation as for the previous network. However unlike the previous case, for the metabolic network we find no dependence on the network size N and the factor R_i is species-specific implying that the concentration of each different substrate species is amplified (from the deterministic concentration) by a different factor. Hence the finite-volume corrections in this network are locally determined whereas in the previously studied network they are determined by global quantities. This indeed highlights the influence of network topology on the finite-volume corrections.

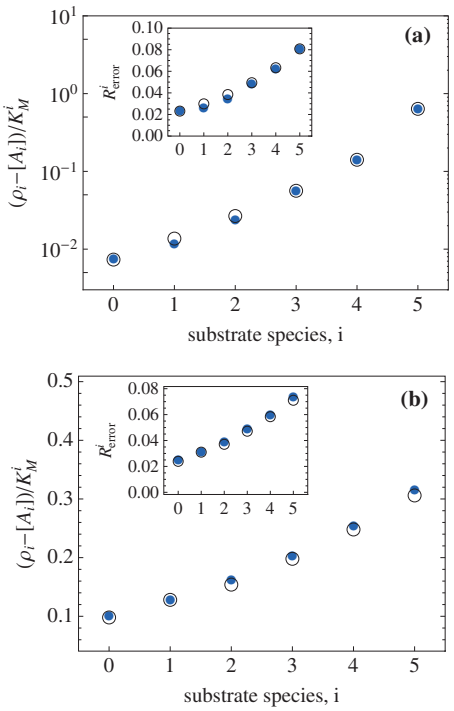

FIG. 4. *Multisubstrate network with multienzyme species.* (a) Heterogeneous catalytic constants and homogeneous Michaelis–Menten constants; (b) heterogeneous Michaelis–Menten constants and homogeneous catalytic constants. Theoretical predictions are shown as open circles while simulation results are the data points. Theory predicts that the relative error for this network is generally substrate species-specific unlike for the multisubstrate, single-enzyme network where it is the same for all species. The insets confirm this via simulation. For case (b) theory also predicts that the reduced macroscopic concentrations of all substrate species are equal but the reduced mesoscopic concentrations are different for each substrate. This is once again confirmed by simulations which indeed show that the mesoscopic concentrations increase as we go further "downstream" in the network. Detailed parameter values can be found in the text.

We have carried out a set of numerical tests in order to verify the accuracy of the theoretical predictions for the relative error [as given by Eq. (54) together with Eq. (70)] and for the absolute differences between the reduced mesoscopic and macroscopic concentrations [as given by Eq. (55) together with Eqs. (69) and (70)]. The two tests are carried out on a six substrate network and are as follows.

Test 1: Heterogeneous catalytic constants and homogeneous Michaelis–Menten constants. The catalytic constant is chosen to vary with species i by imposing $k_2^i=0.5\times 1.3^{2-i}$ while the Michaelis–Menten constants are all made equal $K_M^i=0.1$. The other relevant parameter values are $k_{\text{in}}=0.002$, $k_1=10$, $[E_T^i]V=2$, and $V=200$. Figure 4(a) shows a comparison between the theory and the stochastic simulations. Note that the relative error for each species is different, in contrast to that for the previous network [inset of Fig. 3(b)].

Test 2: Heterogeneous Michaelis–Menten constants and homogeneous catalytic constants. The catalytic constants are all made equal, $k_2^i=1$, while the Michaelis–Menten constants are made heterogeneous by choosing the complex dissociation constants to be substrate-specific, $k_{-1}^i=9\times 1.3^{2-i}$. The other parameters are $k_{\text{in}}=0.002$, $k_1=200$, $[E_T^i]V=1$, and $V=400$. Note that in this case the reduced macroscopic concentrations are all equal, $[A_i]/K_M^i=4$, since Eq. (69) only depends on the catalytic constants. In contrast, the reduced mesoscopic concentrations are predicted and confirmed by simulation [Fig. 4(b)] to vary from one species to another. This is due to the explicit dependence on the Michaelis–Menten constant in Eq. (70).

V. DISCUSSION

In this article we have developed a new approach to the calculation of finite-volume corrections to the substrate concentrations as predicted by the RE theory of enzyme-catalyzed networks. This theory is complementary to the general theory of EMREs recently developed by Grima.[9] The EMRE theory gives explicit analytical results only for relatively simple biochemical circuits; we have verified that for one such simple case, i.e., the single-substrate, single-enzyme reaction, the present theory gives the same result as the EMRE. This is an important benchmark for both theories since they were derived using completely different methods (the Poisson representation and the system-size expansion). A noteworthy achievement of the present theory over the EMRE theory is that it can produce analytical results even for very complex networks involving arbitrarily large numbers of species; this comes about by imposing a separation of timescales of all substrate and complex species in the network which results in a very substantial dimensional reduction of the CME, a result akin to that obtained from applying the QSSA on the REs. To our knowledge, this is the first time that the "mesoscopic QSSA" has been used to obtain a simple analytical picture of the stochastic dynamics of large networks in finite volumes, i.e., considering effects beyond the conventional fluctuation-dissipation theorem or equivalently beyond those which can be predicted using the linear-noise approximation. Previous studies employing the mesoscopic QSSA have been principally occupied with the need to increase the speed of stochastic simulation.[22–25]

In particular, using our approach, we have conclusively shown that the predictions of the RE models for the substrate concentrations will always underestimate the real substrate concentrations of a network confined in a finite-volume. A generic prediction is that the size of the differences increases with decreasing subcellular volume, decreasing Michaelis–Menten constants and increasing enzyme saturation. The topology of the network plays a determining role in the magnitude of these corrections and how they vary according to the "position" of the substrate species in the network. For example for the multisubstrate, single-enzyme network we find that the mesoscopic concentrations of all substrate species are larger than the RE predictions by the same factor (a global effect) whereas in the multisubstrate, multienzyme network the factor is species-specific (a local effect).

We note that in this study we have not considered the effects of diffusion and of finite particle size on the stochastic kinetics; for the case of a simple single-substrate, single-enzyme reaction, these factors have been shown to induce a renormalization of rate constants which is apparent even at deterministic length scales.[26] This suggests that the finite-volume corrections that we estimated in this paper constitute only a lower-bound on the actual differences between the

real mesoscopic description (which takes into account intrinsic noise and noise stemming from both factors) and the conventional RE description (which ignores both factors and also intrinsic noise).

Concluding we have presented a complete method of calculation by which one can study the properties of large-scale stochastic enzyme kinetic networks. We believe that the ability of the method to give simple expressions describing the kinetics of such large-scale networks is unprecedented and hence may lead to new insight into the effects of noise in biologically relevant networks.

ACKNOWLEDGMENTS

The authors gratefully acknowledge support by the German Science Foundation (P.T. and A.V.S., DFG project No. STR 1021/1-2) and by the Scottish Universities Life Sciences Alliance (R.G.).

APPENDIX A: DERIVATION OF THE POISSON REPRESENTATION FROM THE MOMENT GENERATING FUNCTION EQUATION

Consider a birth-death Markov process with moment generating function $G(z)$, which evolves as $\partial_t G(z)=\mathcal{R}^G G$, where $\mathcal{R}^G \equiv \mathcal{R}^G(z,\nabla_z)$. Note that the generic form of $\mathcal{R}^G$ has all the derivatives acting to the right. A complex Poisson representation can be obtained by expanding G

$$G(z,t) = \int d\boldsymbol{\alpha} f(\boldsymbol{\alpha},t) e^{\Sigma(z_i-1)\alpha_i V}. \tag{A1}$$

Note that here the summation in the exponent extends over all components of z and $\boldsymbol{\alpha}$. The representation is carried out in intensive variables α_i. Under the condition that f is sufficiently compact, f satisfies a Liouville equation $\partial_t f(\boldsymbol{\alpha})=\mathcal{R} f$, where $\mathcal{R} \equiv \mathcal{R}(\boldsymbol{\alpha},\nabla_\alpha)$. The relation between $\mathcal{R}$ and $\mathcal{R}^G$ can be clarified by differentiating Eq. (A1) with respect to t

$$\partial_t G(z) = \mathcal{R}^G G(z) = \int d\boldsymbol{\alpha} f(\boldsymbol{\alpha}) \mathcal{R}^G(z,\nabla_z) e^{\Sigma(z_i-1)\alpha_i V} \tag{A2}$$

$$= \int d\boldsymbol{\alpha} f(\boldsymbol{\alpha}) \mathcal{R}^G(1+V^{-1}\nabla_\alpha, V\boldsymbol{\alpha}) e^{\Sigma(z_i-1)\alpha_i V}. \tag{A3}$$

If f is sufficiently compact, such that by partial integration all boundary terms vanish, we obtain

$$\partial_t G(z) = \int d\boldsymbol{\alpha} (\mathcal{R}^G(1-V^{-1}\nabla_\alpha, V\boldsymbol{\alpha}) f(\boldsymbol{\alpha})) e^{\Sigma(z_i-1)\alpha_i V}. \tag{A4}$$

Therefore the relation between $\mathcal{R}^G$ and $\mathcal{R}$ is given by

$$\mathcal{R}(\boldsymbol{\alpha},\nabla_\alpha) \equiv \mathcal{R}^G(1-V^{-1}\nabla_\alpha, V\boldsymbol{\alpha}). \tag{A5}$$

Thus formally the Liouville operator $\mathcal{R}$ can be obtained by replacing each z_i by $(1-V^{-1}\partial_\alpha^i)$ and each ∂_z^i by $V\alpha_i$ in $\mathcal{R}^G$.

APPENDIX B: EXPLICIT FORM OF THE CORRELATORS

The reduced description of enzymatic networks involves the matrix of transport coefficient, defined by $\underline{S} \equiv -\langle z\mathcal{L}_1^{-1} z^T\rangle_\pi$, where the steady-state distribution π is obtained from

$$\mathcal{L}_1 \pi = 0. \tag{B1}$$

The quantity is related to the spectrum of the fast mesoscopic variables z described by $\mathcal{L}_1$ at zero frequency. It enters the reduced evolution as a transport coefficient as a consequence of the fast relaxation of the variable z. To see this we follow the derivation in Ref. 13 and consider $\langle z\mathcal{L}_1^{-1} z^T\rangle_\pi = \int dz z\mathcal{L}_1^{-1}(1-P) z^T\pi$, since $Pz\pi=0$, as required. Now consider

$$\int_0^\infty ds\, \exp(-\mathcal{L}_1 s) = \frac{1}{\mathcal{L}_1} \exp(-\mathcal{L}_1 s)|_\infty^0 = \mathcal{L}_1^{-1}(1-P), \tag{B2}$$

where we have used that $P=\lim_{s\to\infty} \exp(-\mathcal{L}_1 s)$. Noticing further that $\exp(-\mathcal{L}_1 s) z^T\pi$ is a solution to $-\partial_s f=\mathcal{L}_1 f$ with initial condition $z^T\pi$, we obtain

$$\langle z\mathcal{L}_1^{-1} z^T\rangle_\pi = \int_0^\infty ds \langle z(s) z^T(0)\rangle_\pi. \tag{B3}$$

Thus if the adiabatic elimination is carried out in the Poisson representation, the transport coefficients are connected only to the spectrum of the mesoscopic Liouvillian $\mathcal{L}_1$, which is independent of the macroscopic shot noise contribution. Adiabatic fluctuations are necessarily of Ornstein–Uhlenbeck form, as shown in the main text. Therefore we consider Eq. (B1) in the form of the potential condition

$$(\underline{M} z - \underline{D}\nabla_z)\pi = 0 \tag{B4}$$

from which we can easily verify that $\langle zz^T\rangle_\pi = -(\underline{D}\underline{M}^{-1})^T$; the latter is supposed to yield a symmetric expression. Then using the regression theorem, we have

$$\langle z\mathcal{L}_1^{-1} z^T\rangle_\pi = \int_0^\infty ds\, \exp(-\underline{M} s)\langle zz^T\rangle_\pi = -\underline{M}^{-1}\underline{D}^T\underline{M}^{-T}. \tag{B5}$$

Hence it follows that $\underline{S}=\underline{M}^{-1}\underline{D}^T\underline{M}^{-T}$. In a similar manner, one can also prove that $\langle z\mathcal{L}_1^{-1}\nabla_z^T\rangle_\pi = -\underline{S}(\underline{M}^{-1}\underline{D})^{-T} = -\underline{M}^{-1}$.

[1] R. Grima and S. Schnell, Essays Biochem. **45**, 41 (2008).
[2] Y. Ishihama, T. Schmidt, J. Rappsilber, M. Mann, F. U. Hartl, M. J. Kerner, and D. Frishman, BMC Genomics **9**, 102 (2008).
[3] D. T. Gillespie, Annu. Rev. Phys. Chem. **58**, 35 (2007).
[4] D. T. Gillespie, J. Phys. Chem. **81**, 2340 (1977).
[5] N. G. VanKampen, *Stochastic Processes in Physics and Chemistry* (Elsevier, New York, 2007).
[6] J. Elf and M. Ehrenberg, Genome Res. **13**, 2475 (2003).
[7] R. Grima, Phys. Rev. Lett. **102**, 218103 (2009).
[8] R. Grima, BMC Sys. Biol. **3**, 101 (2009).
[9] R. Grima, J. Chem. Phys. **133**, 035101 (2010).
[10] M. Gerstung, J. Timmer, and C. Fleck, Phys. Rev. E **79**, 011923 (2009).
[11] P. D. Drummond, T. G. Vaughan, and A. J. Drummond, J. Phys. Chem. A **114**, 10481 (2010).
[12] B. Alberts, D. Bray, J. Lewis, M. Raff, K. Roberts, and J. D. Watson, *Molecular Biology of the Cell* (Garland, New York, 1994).
[13] C. W. Gardiner, *Handbook of Stochastic Methods* (Springer, New York,

2004).
[14] C. W. Gardiner and S. Chaturvedi, J. Stat. Phys. **17**, 429 (1977).
[15] A. Cornish-Bowden, *Fundamentals of Enzyme Kinetics* (Portland Press, London, 2004).
[16] L. A. Segel and M. Slemrod, SIAM Rev. **31**, 446 (1989).
[17] C. W. Gardiner, Phys. Rev. A **29**, 2814 (1984).
[18] F. Haake, Z. Phys. B Condens. Matter **48**, 31 (1982).
[19] C. T. Walsh, *Post Translational Modification of Proteins: Expanding Nature's Inventory* (Roberts & Company, Colorado, 2006).
[20] J. Srividhya and S. Schnell, Comput. Biol. Chem. **30**, 209 (2006).
[21] D. P. Tschudy and H. L. Bonkowsky, Mol. Cell. Biochem. **2**, 55 (1973).
[22] C. V. Rao and A. P. Arkin, J. Chem. Phys. **118**, 4999 (2003).
[23] N. Sinitsyn, N. Hengartner, and I. Nemenman, Proc. Natl. Acad. Sci. U.S.A. **106**, 10546 (2009).
[24] S. Peleš, B. Munsky, and M. Khammash, J. Chem. Phys. **125**, 204104 (2006).
[25] E. A. Mastny, E. L. Haseltine, and J. B. Rawlings, J. Chem. Phys. **127**, 094106 (2007).
[26] M. Tokuyama and J. Ross, J. Chem. Phys. **91**, 4043 (1989).