The Annals of Applied Probability
2013, Vol. 23, No. 2, 529–583
DOI: 10.1214/12-AAP841
© Institute of Mathematical Statistics, 2013

SEPARATION OF TIME-SCALES AND MODEL REDUCTION FOR STOCHASTIC REACTION NETWORKS[1]

BY HYE-WON KANG AND THOMAS G. KURTZ

University of Minnesota and University of Wisconsin, Madison

A stochastic model for a chemical reaction network is embedded in a one-parameter family of models with species numbers and rate constants scaled by powers of the parameter. A systematic approach is developed for determining appropriate choices of the exponents that can be applied to large complex networks. When the scaling implies subnetworks have different time-scales, the subnetworks can be approximated separately, providing insight into the behavior of the full network through the analysis of these lower-dimensional approximations.

1. Introduction. Chemical reaction networks in biological cells involve chemical species with vastly differing numbers of molecules and reactions with rate constants that also vary over several orders of magnitude. This wide variation in number and rate yield phenomena that evolve on very different time-scales. As in many other areas of application, these differing time-scales can be exploited to obtain simplifications of complex models. Papers by Rao and Arkin (2003) and Haseltine and Rawlings (2002) stimulated considerable interest in this approach and notable contributions by Cao, Gillespie and Petzold (2005), Goutsias (2005), E, Liu and Vanden-Eijnden (2007), Mastny, Haseltine and Rawlings (2007), Crudu, Debussche and Radulescu (2009) and others. All of the cited work considers models of chemical reaction networks given by continuous time Markov chains where the state of the chain is an integer vector whose components give the numbers of molecules of each of the chemical species involved in the reaction. Most of the analysis carried out in this previous work is based on the chemical master equation (the Kolmogorov forward equation) determining the one-dimensional distributions of the process and is focused on simplifying simulation methods for the process. In contrast, the analysis in Ball et al. (2006) is based primarily on stochastic equations determining the process and focuses on the derivation of simplified models obtained as limits of rescaled versions of the original model.

The present paper gives a systematic development of many of the ideas introduced in Ball et al. (2006). First, recognizing that the variation in time-scales is

Received September 2010; revised October 2011.

[1]Supported in part by NSF Grants DMS 05-53687 and 08-05793.

MSC2010 subject classifications. 60J27, 60J80, 60F17, 92C45, 80A30.

Key words and phrases. Reaction networks, chemical reactions, cellular processes, multiple time scales, Markov chains, averaging, scaling limits, quasi-steady state assumption.

due both to variation in species number and to variation in rate constants, we normalize species numbers and rate constants by powers of a fixed constant N_0 which we assume to be "large."

Second, we replace N_0 by a parameter N to obtain a one-parameter family of models and obtain our approximate models as rigorous limits as $N \to \infty$. It is natural to compare this approach to singular perturbation analysis of deterministic models [cf. Segel and Slemrod (1989)] and many of the same ideas and problems arise. This kind of analysis is implicit in some of the earlier work and is the basis for the work in Ball et al. (2006).

Third, as in Ball et al. (2006), the different time-scales are identified with powers N_0^{γ}, and making a change of time variable (replacing t by tN^{γ}), we get different limiting/approximate models involving different subsets of the chemical species. As observed in Cao, Gillespie and Petzold (2005) and E, Liu and Vanden-Eijnden (2007), the variables in the approximate models may correspond to linear combinations of species numbers. We identify the time-scale of a species or a reaction with the exponent γ for which the asymptotic behavior is nondegenerate, that is, the quantity has a nonconstant, well-behaved limit. The time-scale of a reaction is determined by the scaling of its rate constant and by the scaling of the species numbers of the species that determine the intensity/propensity function for the reaction. The time-scale of a species will depend both on the scaling of the intensity/propensity functions (the reaction time-scales) and on the scaling of the species number. It can happen that the scaling of a species number will need to be different for different time scales, and a species may appear in the limiting model for more than one of the time scales.

Fourth, the limiting models may be stochastic, deterministic or "hybrid" involving stochastically driven differential equations, that is, *piecewise deterministic* Markov processes [see Davis (1993)]. Haseltine and Rawlings (2002) obtain hybrid models and hybrid models have been used elsewhere in reaction network modeling [e.g., Hensel, Rawlings and Yin (2009), Zeiser, Franz and Liebscher (2010)] and are a primary focus of Crudu, Debussche and Radulescu (2009).

Finally, as in Ball et al. (2006), we carry out our analysis using stochastic equations of the form

$$X(t) = X(0) + \sum_k Y_k\left(\int_0^t \lambda_k(X(s))\,ds\right)\zeta_k$$

that determine the continuous time Markov chain model. Here the Y_k are independent unit Poisson processes and the ζ_k are vectors in $\mathbb{Z}^d$. These equations are rescaled and the analysis carried out exploiting the law of large numbers and martingale properties of the Y_k. [For more information, see Kurtz (1977/78) and Ethier and Kurtz (1986), Chapter 11.] The other critical component of the analysis is averaging methods that date back at least to Khas'minskiĭ (1966a, 1966b). [We follow Kurtz (1992). See that paper for additional references.]

If N_0 is large but not large enough, the limiting model obtained by the procedure outlined above may have components that exhibit no fluctuation but corresponding to components in the original model that exhibit substantial fluctuation. This observation suggests the possibility of some kind of diffusion/Langevin approximation. Under what we will call the classical scaling (see Section 2), diffusion/Langevin approximations can be determined simply by replacing the rescaled Poisson processes by their appropriate Brownian approximations. In systems with multiple time-scales that involve averaging fast components, fluctuations around averaged quantities may also contribute to the diffusion terms, and identifying an appropriate diffusion approximation becomes more delicate. These "higher order" corrections will be discussed in a later paper [Kang, Kurtz and Popovic (2012)].

Section 2 introduces the general class of models to be considered and defines the scaling parameters used in our approach. For comparison purposes, we will also describe the "classical scaling" that leads to the deterministic law of mass action. Section 3 describes systematic approaches to the selection of the scaling parameters. Unfortunately, even with these methods there may be as much art as science in their selection, although perhaps we should claim that this is a "feature" (flexibility) rather than a "bug" (ambiguity). Section 4 discusses identification of principal time-scales and derivation of the limiting models. Section 5 reviews general averaging methods, and Section 6 gives additional examples.

We believe that these methods provide tools for the systematic reduction of highly complex models. Further evidence for that claim is provided in Kang (2011) in which the methods are applied to obtain a three time-scale reduction of a model of the heat shock response in E. coli given by Srivastava, Peterson and Bentley (2001). We should point out, however, that there are natural examples of model reductions that do not fit into our primary framework. We have focused on situations in which all species abundances remain positive, at least on average, in the limiting models. In Section 6.5, we consider examples which fail the balance conditions of Section 3, but for which reduced models can still be obtained in which some of the species are completely eliminated from the system.

1.1. *Terminology.* This paper relies on work in both the stochastic processes and the chemical physics and biochemical literature. Since the two communities use different terminology, we offer a brief translation table.

Chemistry	**Probability**
Propensity	Intensity
Master equation	Forward equation
Langevin approximation	Diffusion approximation
Van Kampen approximation	Central limit theorem
Quasi steady state/partial equilibrium analysis	Averaging

The terminology in the last line is less settled on both sides, and the methods we will discuss in Section 5 may not yield "averages" at all, although when they do not they still correspond well to the quasi-steady state assumption in the chemical literature.

2. Equations for the system state.

The standard notation for a chemical reaction

$$A + B \rightharpoonup C$$

is interpreted as "a molecule of A combines with a molecule of B to give a molecule of C."

$$A + B \rightleftharpoons C$$

means that the reaction can go in either direction, that is, in addition to the previous reaction, a molecule of C can dissociate into a molecule of A and a molecule of B. We consider a *network* of reactions involving s_0 chemical species, $S_1, \dots, S_{s_0}$, and r_0 chemical reactions

$$\sum_{i=1}^{s_0} \nu_{ik} S_i \rightharpoonup \sum_{i=1}^{s_0} \nu'_{ik} S_i, \qquad k = 1, \dots, r_0,$$

where the ν_{ik} and ν'_{ik} are nonnegative integers. If the kth reaction occurs, then for $i = 1, \dots, s_0$, ν_{ik} molecules of S_i are consumed and ν'_{ik} molecules are produced. We write reversible reactions as two separate reactions.

Let $X(t) \in \mathbb{N}^{s_0}$ be the vector whose components give the numbers of molecules of each species in the system at time t. Let ν_k be the vector with components ν_{ik} and ν'_k the vector with components ν'_{ik}. If the kth reaction occurs at time t, then the state satisfies

$$X(t) = X(t-) + \nu'_k - \nu_k.$$

If $R_k(t)$ is the number of times that the kth reaction occurs by time t, then

$$X(t) = X(0) + \sum_k R_k(t)(\nu'_k - \nu_k) = X(0) + (\nu' - \nu)R(t),$$

where ν' is the $s_0 \times r_0$-matrix with columns given by the ν'_k, ν is the matrix with columns given by the ν_k, and $R(t) \in \mathbb{N}^{r_0}$ is the vector with components $R_k(t)$.

Modeling X as a continuous time Markov chain, we can write

$$R_k(t) = Y_k\left(\int_0^t \lambda_k(X(s))\,ds\right), \tag{2.1}$$

where the Y_k are independent unit Poisson processes and $\lambda_k(x)$ is the rate at which the kth reaction occurs if the chain is in state x, that is, $\lambda_k(X(t))$ gives the *intensity* (*propensity* in the chemical literature) for the kth reaction. Then X is the solution of

$$X(t) = X(0) + \sum_k Y_k\left(\int_0^t \lambda_k(X(s))\,ds\right)(\nu_k' - \nu_k). \tag{2.2}$$

Define $\zeta_k = \nu_k' - \nu_k$. The generator of the process has the form

$$\mathbb{B}f(x) = \sum_k \lambda_k(x)\big(f(x+\zeta_k) - f(x)\big).$$

Assuming that the solution of (2.2) exists for all time, that is, X jumps only finitely often in a finite time interval,

$$f(X(t)) - f(X(0)) - \int_0^t \mathbb{B}f(X(s))\,ds \tag{2.3}$$

is at least a local martingale for all functions on the state space of the process X.

If (2.3) is a martingale, then its expectation is zero and

$$\sum_x f(x)p(x,t) = \sum_x f(x)p(x,0) + \int_0^t \mathbb{B}f(x)p(x,s)\,ds, \tag{2.4}$$

where $p(x,t) = P\{X(t) = x\}$. Taking $f(x) = \mathbf{1}_{\{y\}}(x)$, (2.4) gives the Kolmogorov forward equations (or *master equation* in the chemical literature)

$$\dot{p}(y,t) = \sum_k \lambda_k(y-\zeta_k)p(y-\zeta_k,t) - \sum_k \lambda_k(y)p(y,t). \tag{2.5}$$

The stochastic equation (2.2), the martingales (2.3) and the forward equation (2.5) provide three different ways of specifying the same model. This paper focuses primarily on the stochastic equation which seems to be the simplest approach to identifying and analyzing the rescaled families of models that we will introduce.

In what follows, we will focus on reactions that are at most binary (i.e., consume at most two molecules), so $\lambda_k(x)$ must have one of the following forms:

$\boldsymbol{\lambda_k}$	**Reaction**	$\boldsymbol{\nu_k}$
κ_k'	$\varnothing \to$ stuff	0
$\kappa_k' x_i$	$S_i \to$ stuff	e_i
$\kappa_k' V^{-1} x_i(x_i - 1)$	$2S_i \to$ stuff	$2e_i$
$\kappa_k' V^{-1} x_i x_j$	$S_i + S_j \to$ stuff	$e_i + e_j$

Here V denotes some measure of the volume of the system, and the form of the rates reflects the fact that the rate of a binary reaction in a well-stirred system should vary inversely with the volume of the system. Note that if $\zeta_{ik} < 0$, then

$\lambda_k(x)$ must have x_i as a factor. Higher order reactions can be included at the cost of more complicated expressions for the λ_k.

Our intent is to embed the model of primary interest X into a family of models X^N indexed by a large parameter N. The model X corresponds to a particular value of the parameter $N = N_0$, that is, $X = X^{N_0}$.

For each species i, let $\alpha_i \geq 0$ and define the *normalized abundance* (or simply, the abundance) for the Nth model by

$$Z_i^N(t) = N^{-\alpha_i} X_i^N(t).$$

Note that the abundance may be the species number ($\alpha_i = 0$), the species concentration or something else. The exponent α_i should be selected so that $Z_i^N = O(1)$. To be precise, we want $\{Z_i^N(t)\}$ to be stochastically bounded, that is, for each $\varepsilon > 0$, there exists $K_{\varepsilon,t} < \infty$ such that

$$\inf_N P\Big\{\sup_{s \leq t} Z_i^N(s) \leq K_{\varepsilon,t}\Big\} \geq 1 - \varepsilon.$$

In other words, we want α_i to be "large enough." On the other hand, we do not want α_i to be so large that Z_i^N converges to zero as $N \to \infty$. For example, the existence of $\delta_{\varepsilon,t}$ such that

$$\inf_N P\Big\{\inf_{s \leq t} Z_i^N(s) \geq \delta_{\varepsilon,t}\Big\} \geq 1 - \varepsilon$$

would suffice; however, there are natural situations in which $\alpha_i = 0$ and Z_i^N is occasionally or even frequently zero, so this requirement would in general be too restrictive. For the moment, we just keep in mind that α_i cannot be "too big."

The rate constants may also vary over several orders of magnitude, so we define κ_k by setting $\kappa_k' = \kappa_k N_0^{\beta_k}$ for unary reactions and $\kappa_k' V^{-1} = \kappa_k N_0^{\beta_k}$ for binary reactions. The β_k should be selected so that the κ_k are of order one, although we again avoid being too precise regarding the meaning of "order one." For a unary reaction, the intensity for the model of primary interest becomes

$$\kappa_k' x_i = N_0^{\beta_k + \alpha_i} \kappa_k z_i = N_0^{\beta_k + \nu_k \cdot \alpha} \kappa_k z_i$$

and for binary reactions,

$$\kappa_k' V^{-1} x_i x_j = N_0^{\beta_k + \alpha_i + \alpha_j} \kappa_k z_i z_j = N_0^{\beta_k + \nu_k \cdot \alpha} \kappa_k z_i z_j$$

and

$$\kappa_k' V^{-1} x_i (x_i - 1) = N_0^{\beta_k + 2\alpha_i} \kappa_k z_i (z_i - N_0^{-\alpha_i}) = N_0^{\beta_k + \nu_k \cdot \alpha} \kappa_k z_i (z_i - N_0^{-\alpha_i}). \tag{2.6}$$

The Nth model in the scaled family is given by the system

$$Z_i^N(t) = Z_i^N(0) + \sum_k N^{-\alpha_i} Y_k\left(\int_0^t N^{\beta_k + \nu_k \cdot \alpha} \lambda_k(Z^N(s))\, ds\right)(\nu_{ik}' - \nu_{ik}).$$

For binary reactions of the form $2S_i \to$ stuff with $\alpha_i > 0$, $\lambda_k(z) = \kappa_k z_i(z_i - N^{-\alpha_i})$ depends on N, but to simplify notation we still write λ_k rather than λ_k^N.

Let $\Lambda_N = \operatorname{diag}(N^{-\alpha_1}, \dots, N^{-\alpha_{s_0}})$, $\rho_k = \beta_k + \nu_k \cdot \alpha$, and $\zeta_k = \nu_k' - \nu_k$. The generator for Z^N is

$$\mathbb{B}_N f(z) = \sum_k N^{\rho_k} \lambda_k(z)\big(f(z + \Lambda_N \zeta_k) - f(z)\big).$$

Even after the β_k and α_i are selected, we still have the choice of time-scale on which to study the model, that is, we can consider

$$\begin{aligned} Z_i^{N,\gamma}(t) &= Z_i^N(tN^\gamma) \\ &= Z_i^N(0) + \sum_k N^{-\alpha_i} Y_k\left(\int_0^t N^{\gamma+\beta_k+\nu_k\cdot\alpha} \lambda_k(Z^{N,\gamma}(s))\,ds\right) \\ &\qquad \times (\nu_{ik}' - \nu_{ik}) \end{aligned} \tag{2.7}$$

for any $\gamma \in \mathbb{R}$. Different choices of γ may give interesting approximations for different subsets of species. To identify that approximation, note that if $\lim_{N\to\infty} Z_i^{N,\gamma} = Z_i^\gamma$ and N_0 is "large," then we should have

$$X_i(t) \equiv X_i^{N_0}(t) \approx N_0^{\alpha_i} Z_i^\gamma(tN_0^{-\gamma}).$$

In what we will call the *classical scaling* [see, e.g., Kurtz (1972, 1977/78)] N_0 has the interpretation of volume times Avogadro's number and $\alpha_i = 1$, for all i, so $Z_i^{N_0}$ is the concentration of S_i. Taking $\beta_k = 0$ for a unary reaction and $\beta_k = -1$ for a binary reaction, the intensities are all of the form $N\lambda_k(z)$, and, hence, taking $\gamma = 0$, $Z^N = Z^{N,0}$ converges to the solution of

$$Z_i(t) = Z_i(0) + \sum_k \int_0^t \kappa_k Z(s)^{\nu_k}\,ds(\nu_{ik}' - \nu_{ik}), \tag{2.8}$$

where $z^{\nu_k} = \prod_i z_i^{\nu_{ik}}$. Note that (2.8) is just the usual *law of mass action* model for the network.

3. Determining the scaling exponents. For systems with a diversity of scales because of wide variations in species numbers or rate constants or both, the challenge is to select the α_i and the β_k in ways that capture this variation and produce interesting approximate models. Once the exponents and N_0 are selected,

$$X_i^N(0) = \left\lfloor \left(\frac{N}{N_0}\right)^{\alpha_i} X_i(0) \right\rfloor,$$

and the family of models to be studied is determined.

Suppose

$$\kappa_1' \ge \kappa_2' \ge \cdots \ge \kappa_{r_0}'.$$

Then it is reasonable to select the β_i so that $\beta_1 \geq \cdots \geq \beta_{r_0}$, although it may be natural to impose this order separately for unary and binary reactions. (See the "classical" scaling.)

Typically, we want to select the α_i so that $Z_i^N(t) = N^{-\alpha_i} X_i^N(t) = O(1)$, or, more precisely, assuming $\lim_{N\to\infty} Z_i^N(0) = Z_i(0) > 0$, for all i, we want to avoid α, β and γ for which $\lim_{N\to\infty} Z_i^N(tN^\gamma) = 0$, for all $t > 0$ or $\lim_{N\to\infty} Z_i^N(tN^\gamma) = \infty$, for all $t > 0$. This goal places constraints on α, β and possibly γ.

3.1. *Species balance.* Consider the reaction system

$$S_1 + S_2 \rightharpoonup S_3 + S_4,$$
$$S_3 + S_5 \rightharpoonup S_6.$$

Then the equation for $Z_3^{N,\gamma}$ is

$$Z_3^{N,\gamma}(t) = Z_3^N(0) + N^{-\alpha_3} Y_1\left(N^{\gamma+\beta_1+\alpha_1+\alpha_2} \int_0^t \kappa_1 Z_1^{N,\gamma}(s) Z_2^{N,\gamma}(s)\, ds\right)$$
$$- N^{-\alpha_3} Y_2\left(N^{\gamma+\beta_2+\alpha_3+\alpha_5} \int_0^t \kappa_2 Z_3^{N,\gamma}(s) Z_5^{N,\gamma}(s)\, ds\right).$$

Assuming that $Z_i^{N,\gamma} = O(1)$ for $i \neq 3$ and $Z_3^N(0) = O(1)$, $Z_3^{N,\gamma} = O(1)$ if

$$(\beta_1 + \alpha_1 + \alpha_2 + \gamma) \vee (\beta_2 + \alpha_3 + \alpha_5 + \gamma) \leq \alpha_3$$

(the power of N outside the Poisson processes dominates the power inside) or if

$$\beta_1 + \alpha_1 + \alpha_2 = \beta_2 + \alpha_3 + \alpha_5. \tag{3.1}$$

Assuming (3.1), if $Z_3^{N,\gamma}(s) > \frac{\kappa_1 Z_1^{N,\gamma}(s) Z_2^{N,\gamma}(s)}{\kappa_2 Z_5^{N,\gamma}(s)}$, the rate of consumption of S_3 exceeds the rate of production, and if the inequality is reversed, the rate of production exceeds the rate of consumption ensuring that $Z_3^{N,\gamma}$ neither explodes nor is driven to zero.

In general, let $\Gamma_i^+ = \{k : \nu'_{ik} > \nu_{ik}\}$, that is, Γ_i^+ gives the set of reactions that result in an increase in the ith species, and let $\Gamma_i^- = \{k : \nu'_{ik} < \nu_{ik}\}$. Then for each i, we want either

$$\max_{k\in\Gamma_i^-} (\beta_k + \nu_k \cdot \alpha) = \max_{k\in\Gamma_i^+} (\beta_k + \nu_k \cdot \alpha) \tag{3.2}$$

or

$$\max_{k\in\Gamma_i^+\cup\Gamma_i^-} (\beta_k + \nu_k \cdot \alpha) + \gamma \leq \alpha_i. \tag{3.3}$$

We will refer to (3.2) as the *balance equation* for species i and to (3.3) as a *time-scale constraint* since it is equivalent to

$$\gamma \le \alpha_i - \max_{k\in\Gamma_i^+\cup\Gamma_i^-} (\beta_k + \nu_k \cdot \alpha).$$

The requirement that either a species be balanced or the time-scale constraint be satisfied will be called the *species balance condition*.

Equation (3.2) is the requirement that the maximum rate at which a species is produced is of the same order of magnitude as the rate at which it is consumed. Since consumption rates are proportional to the normalized species state Z_i, Z_i should remain $O(1)$, provided the same is true for the other Z_j even if the normalized reaction numbers blow up. If (3.2) fails to hold, then (3.3) ensures that $Z_i(t) = O(1)$, again provided the other Z_j remain $O(1)$.

Note that if $\zeta_{ik} \neq 0$, then

$$\gamma = \alpha_i - (\beta_k + \nu_k \cdot \alpha) \tag{3.4}$$

is in some sense the *natural time-scale* for the normalized reaction number

$$N^{-\alpha_i} R_k^{N,\gamma}(t) = N^{-\alpha_i} Y_k\left(N^{\gamma+\beta_k+\nu_k\cdot\alpha} \int_0^t \lambda_k(Z^{N,\gamma}(s))\,ds\right).$$

Then, regardless of whether (3.2) or (3.3) holds,

$$\gamma_i = \alpha_i - \max_{k\in\Gamma_i^+\cup\Gamma_i^-} (\beta_k + \nu_k \cdot \alpha) \tag{3.5}$$

is the natural time-scale for species S_i. With reference to (2.7), if $\gamma < \gamma_i$, we expect $Z_i^{N,\gamma}(t)$ to converge to $\lim_{N\to\infty} Z_i^N(0)$. If $\gamma = \gamma_i$ and $\alpha_i > 0$, then we expect

$$\lim_{N\to\infty} Z_i^{N,\gamma_i}(t) = \lim_{N\to\infty}\left(Z_i^N(0) + \sum_{k\in\Gamma_{i,0}} \int_0^t \lambda_k(Z^{N,\gamma_i}(s))\,ds\,(\nu'_{ik} - \nu_{ik})\right),$$

where

$$\Gamma_{i,0} = \left\{l : \beta_l + \nu_l \cdot \alpha = \max_{k\in\Gamma_i^+\cup\Gamma_i^-} (\beta_k + \nu_k \cdot \alpha)\right\}$$

and each integral on the right-hand side is nonconstant but well behaved. If $\alpha_i = 0$, we expect

$$\lim_{N\to\infty} Z_i^{N,\gamma_i}(t) = \lim_{N\to\infty}\left(Z_i^N(0) + \sum_{k\in\Gamma_{i,0}} Y_k\left(\int_0^t \lambda_k(Z^{N,\gamma_i}(s))\,ds\right)(\nu'_{ik} - \nu_{ik})\right).$$

It is important to notice that we associate “time-scales” with species (and as we will see below, with collections of species) and that one reaction may determine different time-scales associated with different species.

3.2. *Collective species balance*. The species balance condition, however, does not by itself ensure that the normalized species numbers are asymptotically all $O(1)$. There may also be subsets of species such that the collective rate of production is of a different order of magnitude than the collective rate of consumption. Consider the following simple network:

$$\varnothing \overset{\kappa_1'}{\rightharpoonup} S_1 \underset{\kappa_3'}{\overset{\kappa_2'}{\rightleftharpoons}} S_2 \overset{\kappa_4'}{\rightharpoonup} \varnothing.$$

If $0 < \beta_4 < \beta_1 < \beta_2 = \beta_3$ and $\alpha_1 = \alpha_2 = 0$, then

$$\begin{aligned} Z_1^N(t) &= Z_1^N(0) + Y_1(\kappa_1 N^{\beta_1} t) + Y_3\left(\kappa_3 N^{\beta_3} \int_0^t Z_2^N(s)\,ds\right) \\ &\quad - Y_2\left(\kappa_2 N^{\beta_2} \int_0^t Z_1^N(s)\,ds\right), \\ Z_2^N(t) &= Z_2^N(0) + Y_2\left(\kappa_2 N^{\beta_2} \int_0^t Z_1^N(s)\,ds\right) - Y_3\left(\kappa_3 N^{\beta_3} \int_0^t Z_2^N(s)\,ds\right) \\ &\quad - Y_4\left(\kappa_4 N^{\beta_4} \int_0^t Z_2^N(s)\,ds\right). \end{aligned} \tag{3.6}$$

Since $\beta_2 = \beta_3 \vee \beta_1$ and $\beta_2 = \beta_3 \vee \beta_4$, the species balance condition is satisfied for all species, but noting that

$$Z_1^N(t) + Z_2^N(t) = Z_1^N(0) + Z_2^N(0) + Y_1(\kappa_1 N^{\beta_1} t) - Y_4\left(\kappa_4 N^{\beta_4} \int_0^t Z_2^N(s)\,ds\right),$$

the species numbers still go to infinity as $N \to \infty$. This example suggests the need to consider linear combinations of species. These linear combinations may, in fact, play the role of "virtual" species or auxiliary variables needed in the specification of the reduced models [cf. Cao, Gillespie and Petzold (2005) and E, Liu and Vanden-Eijnden (2005, 2007)].

To simplify notation, define

$$\rho_k = \beta_k + \nu_k \cdot \alpha,$$

so the scaled model satisfies

$$\begin{aligned} Z^{N,\gamma}(t) &= Z^{N,\gamma}(0) + \Lambda_N \sum_k Y_k\left(N^{\beta_k + \nu_k \cdot \alpha + \gamma} \int_0^t \lambda_k(Z^{N,\gamma}(s))\,ds\right)\zeta_k \\ &= Z^{N,\gamma}(0) + \Lambda_N \sum_k Y_k\left(N^{\rho_k + \gamma} \int_0^t \lambda_k(Z^{N,\gamma}(s))\,ds\right)\zeta_k, \end{aligned}$$

where Λ_N is the diagonal matrix with entries $N^{-\alpha_i}$.

DEFINITION 3.1. For $\theta \in [0,\infty)^{s_0}$, define $\Gamma_\theta^+ = \{k : \theta \cdot \zeta_k > 0\}$ and $\Gamma_\theta^- = \{k : \theta \cdot \zeta_k < 0\}$.

Then, noting that

$$\theta^T \Lambda_N^{-1} Z^{N,\gamma}(t) = \sum_{i=1}^{s_0} \theta_i N^{\alpha_i} Z_i^{N,\gamma}(t) = \sum_{i=1}^{s_0} \theta_i X_i^N(N^\gamma t),$$

$$\begin{aligned}\theta^T \Lambda_N^{-1} Z^{N,\gamma}(t) &= \theta^T \Lambda_N^{-1} Z^{N,\gamma}(0) + \sum_k (\theta \cdot \zeta_k) Y_k\biggl(N^{\rho_k+\gamma} \int_0^t \lambda_k(Z^{N,\gamma}(s))\,ds\biggr) \\ &= \theta^T \Lambda_N^{-1} Z^{N,\gamma}(0) + \sum_{k\in\Gamma_\theta^+} (\theta \cdot \zeta_k) R_k^{N,\gamma}(t) \\ &\quad - \sum_{k\in\Gamma_\theta^-} |(\theta \cdot \zeta_k)| R_k^{N,\gamma}(t).\end{aligned}$$

To avoid some kind of degeneracy in the limit, either the positive and negative sums must cancel, or they must grow no faster than N^{α_θ}, where $\alpha_\theta = \max\{\alpha_i : \theta_i > 0\}$. Consequently, we extend the species balance condition to linear combinations of species.

CONDITION 3.2. For each $\theta \in [0, \infty)^{s_0}$,

$$\max_{k\in\Gamma_\theta^-} (\beta_k + \nu_k \cdot \alpha) = \max_{k\in\Gamma_\theta^+} (\beta_k + \nu_k \cdot \alpha) \tag{3.7}$$

or

$$\gamma \le \gamma_\theta \equiv \max_{i:\theta_i>0} \alpha_i - \max_{k\in\Gamma_\theta^+\cup\Gamma_\theta^-} (\beta_k + \nu_k \cdot \alpha) = \alpha_\theta - \max_{k\in\Gamma_\theta^+\cup\Gamma_\theta^-} \rho_k. \tag{3.8}$$

Of course, if $\theta_i > 0$ for only a single species, then this requirement is just the species balance condition, so Condition 3.2 includes that condition. Again, we will refer to (3.7) as the balance equation for the linear combination $\theta \cdot X = \sum_i \theta_i X_i$. In the special case of $\theta = e_i$, the vector with ith component 1 and other components 0, we say that X_i is balanced or that the species S_i is balanced. If (3.7) fails for θ, we say that $\theta \cdot X$ is *unbalanced*. The inequalities given by (3.8) are again called *time-scale constraints*, as they imply

$$\gamma \le \min_{\theta\cdot X \text{ unbalanced}} \gamma_\theta. \tag{3.9}$$

For example, consider the network

$$\varnothing \xrightarrow{\kappa_1'} S_1 \underset{\kappa_3'}{\overset{\kappa_2'}{\rightleftharpoons}} S_2$$

and assume that $\kappa_k' = \kappa_k N_0^{\beta_k}$, where $\beta_1 = \beta_2 > \beta_3$. For S_2 to be balanced, we must have $\beta_2 + \alpha_1 = \beta_3 + \alpha_2$ and for S_1 to be balanced, we must have

$$\beta_1 \vee (\beta_3 + \alpha_2) = \beta_2 + \alpha_1.$$

Let $\alpha_1 = 0$ and $\alpha_2 = \beta_2 - \beta_3$ so S_1 and S_2 are balanced. For $\theta = (1, 1)$, $\Gamma_\theta^+ = \{1\}$, and $\Gamma_\theta^- = \varnothing$. Consequently, (3.7) fails, so we require

$$\gamma \le \alpha_1 \vee \alpha_2 - \beta_1 = -\beta_3. \tag{3.10}$$

There are two time-scales of interest in this model: $\gamma = -\beta_1$, the natural time-scale of S_1, and $\gamma = -\beta_3$, the natural time-scale of S_2. The system of equations is

$$\begin{aligned}
Z_1^{N,\gamma}(t) = Z_1^N(0) &+ Y_1(\kappa_1 N^{\gamma+\beta_1} t) \\
&- Y_2\left(\kappa_2 N^{\gamma+\beta_2} \int_0^t Z_1^{N,\gamma}(s)\,ds\right) \\
&+ Y_3\left(\kappa_3 N^{\gamma+\beta_3+\alpha_2} \int_0^t Z_2^{N,\gamma}(s)\,ds\right), \\
Z_2^{N,\gamma}(t) = Z_2^N(0) &+ N^{-\alpha_2} Y_2\left(\kappa_2 N^{\gamma+\beta_2} \int_0^t Z_1^{N,\gamma}(s)\,ds\right) \\
&- N^{-\alpha_2} Y_3\left(\kappa_3 N^{\gamma+\beta_3+\alpha_2} \int_0^t Z_2^{N,\gamma}(s)\,ds\right).
\end{aligned}$$

For $\gamma = -\beta_1$, since $\beta_1 = \beta_2 = \beta_3 + \alpha_2$, the limit of $Z^{N,\gamma}$ satisfies

$$\begin{aligned}
Z_1(t) &= Z_1(0) + Y_1(\kappa_1 t) - Y_2\left(\kappa_2 \int_0^t Z_1(s)\,ds\right) + Y_3\left(\kappa_3 \int_0^t Z_2(s)\,ds\right) \\
&= Z_1(0) + Y_1(\kappa_1 t) - Y_2\left(\kappa_2 \int_0^t Z_1(s)\,ds\right) + Y_3(\kappa_3 Z_2(0) t), \\
Z_2(t) &= Z_2(0).
\end{aligned}$$

For $\gamma = -\beta_3$, if we divide the equation for $Z_1^{N,\gamma}$ by $N^{\alpha_2} = N^{\beta_1-\beta_3}$, we see that

$$\begin{aligned}
0 &= \lim_{N\to\infty} N^{-\alpha_2} Z_1^{N,\gamma}(t) \\
&= \lim_{N\to\infty} N^{-\alpha_2} Z_1^N(0) + N^{-\alpha_2} Y_1(\kappa_1 N^{\gamma+\beta_1} t) \\
&\qquad - N^{-\alpha_2} Y_2\left(\kappa_2 N^{\gamma+\beta_2} \int_0^t Z_1^{N,\gamma}(s)\,ds\right) \qquad (3.11) \\
&\qquad + N^{-\alpha_2} Y_3\left(\kappa_3 N^{\gamma+\beta_3+\alpha_2} \int_0^t Z_2^{N,\gamma}(s)\,ds\right) \\
&= \lim_{N\to\infty}\left(\kappa_1 t + \kappa_3 \int_0^t Z_2^{N,\gamma}(s)\,ds - \kappa_2 \int_0^t Z_1^{N,\gamma}(s)\,ds\right)
\end{aligned}$$

and $Z_2^{N,\gamma}$ converges to

$$Z_2(t) = Z_2(0) + \kappa_1 t.$$

With reference to (3.10), if $\gamma > -\beta_3$, then $Z_2^{N,\gamma}(t) \to \infty$, for each $t > 0$, demonstrating the significance of the time-scale constraints.

For $\gamma = -\beta_3$, $Z_1^{N,\gamma}$ fluctuates rapidly and does not converge in a functional sense. Its behavior is captured, at least to some extent, by its occupation measure

$$V_1^{N,\gamma}(C \times [0,t]) = \int_0^t \mathbf{1}_C(Z_1^{N,\gamma}(s))\,ds.$$

Applying the generator to functions of z_1 and using the fact that $\beta_1 - \beta_3 = \beta_2 - \beta_3 = \alpha_2$, $\mathbb{B}^{N,\gamma} f(z_1, z_2) = N^{\alpha_2}\mathbb{C}_{z_2} f(z_1)$, where

$$\begin{aligned}\mathbb{C}_{z_2} f(z_1) = {} & (\kappa_1 + \kappa_3 z_2)\big(f(z_1+1) - f(z_1)\big) \\ & + \kappa_2 z_1\big(f(z_1-1) - f(z_1)\big).\end{aligned}$$

Then

$$f(Z_1^{N,\gamma}(t)) - f(Z_1^{N,\gamma}(0)) - N^{\alpha_2}\int_{\mathbb{N}\times[0,t]} \mathbb{C}_{Z_2^{N,\gamma}(s)} f(z_1) V_1^{N,\gamma}(dz_1 \times ds)$$

is a martingale, and dividing by N^{α_2} and passing to the limit, it is not difficult to see that $V_1^{N,\gamma}$ converges to a measure satisfying

$$\int_{\mathbb{N}\times[0,t]} \mathbb{C}_{Z_2(s)} f(z_1) V_1(dz_1 \times ds) = 0.$$

(See Section 5.) Writing $V_1(dz_1 \times ds) = v_s(dz_1)\,ds$, it follows that v_s is the Poisson distribution with mean $\frac{\kappa_1+\kappa_3 Z_2(s)}{\kappa_2}$. We will refer to v_s as the *conditional-equilibrium* or *local-averaging* distribution.

3.3. *Auxiliary variables*. While (3.5) gives the natural time-scale for individual species, it is clear from examples considered by E, Liu and Vanden-Eijnden (2005) that the species time-scales may not be the only time-scales of interest. As above, define

$$\alpha_\theta = \max_{i:\theta_i>0} \alpha_i \tag{3.12}$$

and

$$Z_\theta^{N,\gamma}(t) = N^{-\alpha_\theta}\theta \cdot \Lambda_N^{-1} Z^{N,\gamma}(t) = N^{-\alpha_\theta}\sum_{i=1}^{s_0} \theta_i X_i^N(N^\gamma t). \tag{3.13}$$

Then the natural time scale for $Z_\theta^{N,\gamma}$ is

$$\gamma_\theta = \alpha_\theta - \max_{k\in\Gamma_\theta^+\cup\Gamma_\theta^-} \rho_k. \tag{3.14}$$

For example, E, Liu and Vanden-Eijnden (2005) consider the network

$$S_1 \underset{\kappa_2'}{\overset{\kappa_1'}{\rightleftharpoons}} S_2 \underset{\kappa_4'}{\overset{\kappa_3'}{\rightleftharpoons}} S_3 \underset{\kappa_6'}{\overset{\kappa_5'}{\rightleftharpoons}} S_4$$

with the rate constants for reactions 3 and 4 much smaller than the others. The scaled model is given by

$$\begin{aligned}
Z_1^N(t) &= Z_1^N(0) + N^{-\alpha_1} Y_2\left(\kappa_2 N^{\beta_2+\alpha_2} \int_0^t Z_2^N(s)\,ds\right) \\
&\quad - N^{-\alpha_1} Y_1\left(\kappa_1 N^{\beta_1+\alpha_1} \int_0^t Z_1^N(s)\,ds\right), \\
Z_2^N(t) &= Z_2^N(0) + N^{-\alpha_2} Y_1\left(\kappa_1 N^{\beta_1+\alpha_1} \int_0^t Z_1^N(s)\,ds\right) \\
&\quad - N^{-\alpha_2} Y_2\left(\kappa_2 N^{\beta_2+\alpha_2} \int_0^t Z_2^N(s)\,ds\right) \\
&\quad + N^{-\alpha_2} Y_4\left(\kappa_4 N^{\beta_4+\alpha_3} \int_0^t Z_3^N(s)\,ds\right) \\
&\quad - N^{-\alpha_2} Y_3\left(\kappa_3 N^{\beta_3+\alpha_2} \int_0^t Z_2^N(s)\,ds\right), \\
Z_3^N(t) &= Z_3^N(0) + N^{-\alpha_3} Y_6\left(\kappa_6 N^{\beta_6+\alpha_4} \int_0^t Z_4^N(s)\,ds\right) \\
&\quad - N^{-\alpha_3} Y_5\left(\kappa_5 N^{\beta_5+\alpha_3} \int_0^t Z_3^N(s)\,ds\right) \\
&\quad + N^{-\alpha_3} Y_3\left(\kappa_3 N^{\beta_3+\alpha_2} \int_0^t Z_2^N(s)\,ds\right) \\
&\quad - N^{-\alpha_3} Y_4\left(\kappa_4 N^{\beta_4+\alpha_3} \int_0^t Z_3^N(s)\,ds\right), \\
Z_4^N(t) &= Z_4^N(0) + N^{-\alpha_4} Y_5\left(\kappa_5 N^{\beta_5+\alpha_3} \int_0^t Z_3^N(s)\,ds\right) \\
&\quad - N^{-\alpha_4} Y_6\left(\kappa_6 N^{\beta_6+\alpha_4} \int_0^t Z_4^N(s)\,ds\right).
\end{aligned}$$

The rate constants used in E, Liu and Vanden-Eijnden (2005) would correspond to $\beta_1 = \beta_2 = \beta_5 = \beta_6 > \beta_3 = \beta_4$, but in order to introduce some complexity in the solution of the balance conditions, assume that $\beta_1 = \beta_2 > \beta_5 = \beta_6 > \beta_3 > \beta_4$. Then if we look for a scaling under which all $\theta \cdot X$ are balanced, $\alpha_1 = \alpha_2$, $\alpha_3 = \alpha_4$, and $\alpha_2 + \beta_3 = \alpha_3 + \beta_4$, so $\alpha_3 = \alpha_2 + \beta_3 - \beta_4$. For definiteness, take $\alpha_1 = \alpha_2 = 0$.

The natural time-scale for S_1 and S_2 is $-\beta_1$, and the natural time-scale for S_3 and S_4 is $-\beta_5$, but on either of these time-scales Z_1+Z_2 and Z_3+Z_4 are constant. In particular,

$$\begin{aligned} U_1^{N,\gamma}(t) &\equiv Z_1^{N,\gamma}(t)+Z_2^{N,\gamma}(t) \\ &= Z_1^N(0)+Z_2^N(0)+Y_4\biggl(\kappa_4 N^{\gamma+\beta_4+\alpha_3}\int_0^t Z_3^{N,\gamma}(s)\,ds\biggr) \\ &\quad - Y_3\biggl(\kappa_3 N^{\gamma+\beta_3}\int_0^t Z_2^{N,\gamma}(s)\,ds\biggr), \\ U_2^{N,\gamma}(t) &\equiv Z_3^{N,\gamma}(t)+Z_4^{N,\gamma}(t) \\ &= Z_3^N(0)+Z_4^N(0)-N^{-\alpha_3}Y_4\biggl(\kappa_4 N^{\gamma+\beta_4+\alpha_3}\int_0^t Z_3^{N,\gamma}(s)\,ds\biggr) \\ &\quad + N^{-\alpha_3}Y_3\biggl(\kappa_3 N^{\gamma+\beta_3}\int_0^t Z_2^{N,\gamma}(s)\,ds\biggr). \end{aligned}$$

For $\gamma_1=\gamma_2=-\beta_1=-\beta_2$, $(Z_1^{N,\gamma_1}, Z_2^{N,\gamma_1})$ converges to

$$\begin{aligned} Z_1^{\gamma_1}(t) &= Z_1(0)+Y_2\biggl(\kappa_2\int_0^t Z_2^{\gamma_1}(s)\,ds\biggr)-Y_1\biggl(\kappa_1\int_0^t Z_1^{\gamma_1}(s)\,ds\biggr), \\ Z_2^{\gamma_1}(t) &= Z_2(0)+Y_1\biggl(\kappa_1\int_0^t Z_1^{\gamma_1}(s)\,ds\biggr)-Y_2\biggl(\kappa_2\int_0^t Z_2^{\gamma_1}(s)\,ds\biggr) \end{aligned}$$

and for $\gamma_3=\gamma_4=-\beta_5=-\beta_6$,

$$\begin{aligned} Z_3^{\gamma_3}(t) &= Z_3(0)+\kappa_6\int_0^t Z_4^{\gamma_3}(s)\,ds-\kappa_5\int_0^t Z_3^{\gamma_3}(s)\,ds, \\ Z_4^{\gamma_3}(t) &= Z_4(0)+\kappa_5\int_0^t Z_3^{\gamma_3}(s)\,ds-\kappa_6\int_0^t Z_4^{\gamma_3}(s)\,ds. \end{aligned}$$

Let $\gamma_{12}=\gamma_\theta$ for $\theta=(1,1,0,0)$. Then $\gamma_{12}=-\beta_3=-(\alpha_3+\beta_4)$ and dividing the equation for $Z_4^{N,\gamma_{12}}$ by $N^{\beta_5-\beta_3}$, we see that

$$\kappa_5\int_0^t Z_3^{N,\gamma_{12}}(s)\,ds-\kappa_6\int_0^t Z_4^{N,\gamma_{12}}(s)\,ds\to 0 \tag{3.15}$$

and, hence,

$$\int_0^t Z_3^{N,\gamma_{12}}(s)\,ds-\frac{\kappa_6}{\kappa_5+\kappa_6}\int_0^t U_2^{N,\gamma_{12}}(s)\,ds\to 0. \tag{3.16}$$

Similarly, dividing the equation for $Z_1^{N,\gamma_{12}}$ by $N^{\beta_2-\beta_3}$,

$$\int_0^t Z_2^{N,\gamma_{12}}(s)\,ds-\frac{\kappa_1}{\kappa_1+\kappa_2}\int_0^t U_1^{N,\gamma_{12}}(s)\,ds\to 0.$$

Since $U_2^{N,\gamma_{12}}$ converges to $U_2(0)$ uniformly on bounded time intervals, $U_1^{N,\gamma_{12}}$ converges to the solution of

$$U_1(t) = U_1(0) + Y_4\left(\frac{\kappa_4\kappa_6}{\kappa_5+\kappa_6}U_2(0)t\right) - Y_3\left(\frac{\kappa_3\kappa_1}{\kappa_1+\kappa_2}\int_0^t U_1(s)\,ds\right).$$

Finally, for $\theta = (0, 0, 1, 1)$ and $\gamma_{34} = \gamma_\theta$, $\gamma_{34} = -\beta_4$ and, as in (3.16),

$$\int_0^t Z_3^{N,\gamma_{34}}(s)\,ds - \frac{\kappa_6}{\kappa_5+\kappa_6}\int_0^t U_2^{N,\gamma_{34}}(s)\,ds \to 0.$$

Dividing the equation for $U_1^{N,\gamma_{34}}$ by $N^{\beta_3-\beta_4}$,

$$\int_0^t Z_2^{N,\gamma_{34}}(s)\,ds - \frac{\kappa_4}{\kappa_3}\int_0^t Z_3^{N,\gamma_{34}}(s)\,ds \to 0.$$

Consequently, even on this faster time-scale, $U_2^{N,\gamma_{34}}$ converges to $U_2(0)$ uniformly on bounded time intervals.

3.4. *Checking the balance conditions*. Condition 3.2 only depends on the support of θ, $\mathrm{supp}(\theta) = \{i : \theta_i \neq 0\}$, and on the signs of $\theta \cdot \zeta_k$, so the condition needs to be checked for only finitely many θ. For $k \in \{1, \dots, r_0\}$, define

$$\Lambda_k^+ = \{\theta \in [0,\infty)^{s_0} : \theta \cdot \zeta_k > 0\}, \qquad \Lambda_k^- = \{\theta \in [0,\infty)^{s_0} : \theta \cdot \zeta_k < 0\},$$
$$\Lambda_k^0 = \{\theta \in [0,\infty)^{s_0} : \theta \cdot \zeta_k = 0\}$$

and for disjoint Γ_-, Γ_+, Γ_0 satisfying $\Gamma_- \cup \Gamma_+ \cup \Gamma_0 = \{1, \dots, r_0\}$, define

$$\Lambda_{\Gamma_-,\Gamma_+,\Gamma_0} = \left(\bigcap_{k\in\Gamma_-}\Lambda_k^-\right) \cap \left(\bigcap_{k\in\Gamma_+}\Lambda_k^+\right) \cap \left(\bigcap_{k\in\Gamma_0}\Lambda_k^0\right).$$

The following lemma is immediate.

LEMMA 3.3. *Fix γ. Condition 3.2 holds for all $\theta \in [0,\infty)^{s_0}$, provided*

$$\max_{k\in\Gamma_-}(\beta_k + \nu_k \cdot \alpha) = \max_{k\in\Gamma_+}(\beta_k + \nu_k \cdot \alpha) \tag{3.17}$$

or

$$\gamma \le \min_{\theta\in\Lambda_{\Gamma_-,\Gamma_+,\Gamma_0}} \max_{i:\theta_i>0} \alpha_i - \max_{k\in\Gamma_+\cup\Gamma_-}(\beta_k + \nu_k \cdot \alpha) \tag{3.18}$$

for all partitions $\{\Gamma_-, \Gamma_+, \Gamma_0\}$ for which $\Lambda_{\Gamma_-,\Gamma_+,\Gamma_0} \neq \varnothing$.

Checking the conditions of Lemma 3.3 could still be a formidable task. The next lemmas significantly reduce the effort required. Observe that for $\theta^1, \theta^2 \in [0,\infty)^{s_0}$

and $c_1, c_2 > 0$, $k \in \Gamma^+_{c_1\theta^1+c_2\theta^2}$ implies $k \in \Gamma^+_{\theta^1} \cup \Gamma^+_{\theta^2}$ and similarly for $\Gamma^-_{c_1\theta^1+c_2\theta^2}$, so

$$\max_{k\in\Gamma^+_{c_1\theta^1+c_2\theta^2}} \rho_k \le \max_{k\in\Gamma^+_{\theta^1}} \rho_k \vee \max_{k\in\Gamma^+_{\theta^2}} \rho_k \tag{3.19}$$

and

$$\max_{k\in\Gamma^-_{c_1\theta^1+c_2\theta^2}} \rho_k \le \max_{k\in\Gamma^-_{\theta^1}} \rho_k \vee \max_{k\in\Gamma^-_{\theta^2}} \rho_k. \tag{3.20}$$

Let G be a directed graph in which the nodes are identified with the species and a directed edge is drawn from S_i to S_j if there is a reaction that consumes S_i and produces S_j. A subgraph $G_0 \subset G$ is *strongly connected* if and only if for each pair $S_i, S_j \in G_0$, there is a directed path in G_0 beginning at S_i and ending at S_j. Single nodes are understood to form strongly connected subgraphs. Recall that G has a unique decomposition $G = \bigcup_j G_j$ into maximal strongly connected subgraphs.

The following lemma may significantly reduce the work needed to verify Condition 3.2.

LEMMA 3.4. *Let $\theta \in [0,\infty)^{s_0}$, and fix γ. Write*

$$\theta = \sum_{j=1}^{m} \theta^j, \tag{3.21}$$

where $\mathrm{supp}(\theta^j) \subset G_j$ *for some maximal strongly connected subgraph G_j and $G_j \neq G_i$ for $i \neq j$. If Condition 3.2 holds for each θ^j, then it holds for θ. More specifically, if the balance equation (3.7) holds for each θ^j, then the balance equation holds for θ, and if (3.8) holds for each θ^j, then (3.8) holds for θ.*

Consequently, if Condition 3.2 holds for each $\theta \in [0,\infty)^{s_0}$ with support in some strongly connected subgraph, then Condition 3.2 holds for all $\theta \in [0,\infty)^{s_0}$; if (3.7) holds for each $\theta \in [0,\infty)^{s_0}$ with support in some strongly connected subgraph, then (3.7) holds for all $\theta \in [0,\infty)^{s_0}$; and if (3.8) holds for each $\theta \in [0,\infty)^{s_0}$ with support in some strongly connected subgraph, then (3.8) holds for all $\theta \in [0,\infty)^{s_0}$.

PROOF. Assume that Condition 3.2 holds for each θ^j, $j = 1,\dots,m$. First, assume that $\Gamma^+_\theta \neq \varnothing$. Select $l_1 \in \Gamma^+_\theta$ satisfying

$$\rho_{l_1} = \max_{k\in\Gamma^+_\theta} \rho_k. \tag{3.22}$$

Since $\Gamma^+_\theta \subset \bigcup_j \Gamma^+_{\theta^j}$, there exists j_1 such that $l_1 \in \Gamma^+_{\theta^{j_1}}$, and using (3.22), we have

$$\max_{k\in\Gamma^+_\theta} \rho_k = \rho_{l_1} \le \max_{k\in\Gamma^+_{\theta^{j_1}}} \rho_k. \tag{3.23}$$

We have three possible cases. First, if $\max_{k\in\Gamma^+_{\theta^{j_1}}} \rho_k \neq \max_{k\in\Gamma^-_{\theta^{j_1}}} \rho_k$, then by (3.8), there exists $i_1 \in \operatorname{supp}(\theta^{j_1})$ such that

$$\gamma + \max_{k\in\Gamma^+_{\theta^{j_1}}\cup\Gamma^-_{\theta^{j_1}}} \rho_k \leq \alpha_{i_1} \tag{3.24}$$

and by (3.23),

$$\gamma + \max_{k\in\Gamma^+_\theta} \rho_k \leq \alpha_{i_1} \leq \max_{i\in\operatorname{supp}(\theta)} \alpha_i. \tag{3.25}$$

Second, if $\max_{k\in\Gamma^+_{\theta^{j_1}}} \rho_k = \max_{k\in\Gamma^-_{\theta^{j_1}}} \rho_k \leq \max_{k\in\Gamma^-_\theta} \rho_k$, then by (3.23), we obtain

$$\max_{k\in\Gamma^+_\theta} \rho_k \leq \max_{k\in\Gamma^+_{\theta^{j_1}}} \rho_k = \max_{k\in\Gamma^-_{\theta^{j_1}}} \rho_k \leq \max_{k\in\Gamma^-_\theta} \rho_k. \tag{3.26}$$

Finally, if

$$\max_{k\in\Gamma^+_{\theta^{j_1}}} \rho_k = \max_{k\in\Gamma^-_{\theta^{j_1}}} \rho_k > \max_{k\in\Gamma^-_\theta} \rho_k, \tag{3.27}$$

we select l_2 in $\Gamma^-_{\theta^{j_1}}$ with $\rho_{l_2} = \max_{k\in\Gamma^-_{\theta^{j_1}}} \rho_k$. The fact that $\rho_{l_2} > \max_{k\in\Gamma^-_\theta} \rho_k$ ensures the existence of j_2 such that $l_2 \in \Gamma^+_{\theta^{j_2}}$. Then we have

$$\max_{k\in\Gamma^+_{\theta^{j_1}}} \rho_k = \max_{k\in\Gamma^-_{\theta^{j_1}}} \rho_k = \rho_{l_2} \leq \max_{k\in\Gamma^+_{\theta^{j_2}}} \rho_k. \tag{3.28}$$

We recursively select l_n and j_n with $l_n \in \Gamma^+_{\theta^{j_n}}$ such that

$$\max_{k\in\Gamma^+_{\theta^{j_{n-1}}}} \rho_k = \max_{k\in\Gamma^-_{\theta^{j_{n-1}}}} \rho_k = \rho_{l_n} \leq \max_{k\in\Gamma^+_{\theta^{j_n}}} \rho_k$$

until we find l_n for which this is no longer possible. Since the G_j are maximal strongly connected subgraphs, there is no possibility that the same θ^j is selected more than once. Thus, the process will terminate for some n and when it does $\max_{k\in\Gamma^+_{\theta^{j_n}}} \rho_k \neq \max_{k\in\Gamma^-_{\theta^{j_n}}} \rho_k$ and

$$\gamma + \max_{k\in\Gamma^+_\theta} \rho_k \leq \gamma + \max_{k\in\Gamma^+_{\theta^{j_n}}} \rho_k \leq \max_{i\in\operatorname{supp}(\theta^{j_n})} \alpha_i \leq \max_{i\in\operatorname{supp}(\theta)} \alpha_i. \tag{3.29}$$

Consequently, we always have either

$$\gamma + \max_{k\in\Gamma^+_\theta} \rho_k \leq \max_{i\in\operatorname{supp}(\theta)} \alpha_i \tag{3.30}$$

or

$$\max_{k\in\Gamma^+_\theta} \rho_k \leq \max_{k\in\Gamma^-_\theta} \rho_k. \tag{3.31}$$

If $\Gamma_\theta^- \neq \varnothing$, interchanging $-$ and $+$, we see that either

$$\gamma + \max_{k\in\Gamma_\theta^-} \rho_k \le \max_{i\in\mathrm{supp}(\theta)} \alpha_i \tag{3.32}$$

or

$$\max_{k\in\Gamma_\theta^-} \rho_k \le \max_{k\in\Gamma_\theta^+} \rho_k. \tag{3.33}$$

Assume that both Γ_θ^+ and Γ_θ^- are nonempty. If both (3.31) and (3.33) hold, then (3.7) is satisfied. If (3.30) and (3.32) hold, then taking the maximum of the left and right-hand sides, (3.8) holds. If (3.30) and (3.33) hold, then (3.8) holds and similarly for (3.31) and (3.32).

If (3.7) holds for all θ^j, then the first and third cases above cannot hold, so (3.26) must hold, giving (3.31) and by the same argument (3.33). Consequently, (3.7) must hold for θ. If (3.8) holds for all θ^j, then the first case above holds, giving (3.30) and by the same argument (3.32), so (3.8) must hold for θ.

If $\Gamma_\theta^+ = \varnothing$ and $\Gamma_\theta^- \neq \varnothing$, then (3.32) must hold and (3.8) holds for θ and similarly with the $+$ and $-$ interchanged.

If both Γ_θ^+ and Γ_θ^- are empty, then (3.7) holds ($-\infty = -\infty$). In particular, $\theta \cdot \zeta_k = 0$ for all ζ_k. $\square$

The remaining lemmas in this section may be useful in verifying Condition 3.2 for the cases that remain, that is, for θ with support in some strongly connected subgraph.

LEMMA 3.5. *Fix $\gamma \in \mathbb{R}$, and suppose (3.8) holds for $\theta^1, \dots, \theta^m \in [0,\infty)^{s_0}$. Then for $c_j > 0$, $j = 1, \dots, m$, (3.8) holds for $\theta = \sum_{j=1}^m c_j \theta^j$.*

PROOF. Since $\theta \cdot \zeta_k > 0$ implies $c_j\theta^j \cdot \zeta_k > 0$ for some j and $\theta \cdot \zeta_k < 0$ implies $c_j\theta^j \cdot \zeta_k < 0$ for some j,

$$\max_{k\in\Gamma_\theta^+\cup\Gamma_\theta^-} \rho_k \le \max_{1\le j\le m} \max_{k\in\Gamma_{\theta^j}^+\cup\Gamma_{\theta^j}^-} \rho_k$$

and there exists j such that

$$\gamma \le \max_{i:\theta_i^j>0} \alpha_i - \max_{k\in\Gamma_{\theta^j}^+\cup\Gamma_{\theta^j}^-} \rho_k \le \max_{i:\theta_i>0} \alpha_i - \max_{k\in\Gamma_\theta^+\cup\Gamma_\theta^-} \rho_k. \qquad \square$$

LEMMA 3.6. *For $\theta^1, \theta^2 \in [0,\infty)^{s_0}$, suppose that*

$$\max_{k\in\Gamma_{\theta^1}^-} \rho_k = \max_{k\in\Gamma_{\theta^1}^+} \rho_k > \max_{k\in\Gamma_{\theta^2}^+\cup\Gamma_{\theta^2}^-} \rho_k. \tag{3.34}$$

Then for $c_1, c_2 > 0$, (3.7) holds for $c_1\theta^1 + c_2\theta^2$.

PROOF. If $l \in \Gamma^+_{\theta^1}$ and $\rho_l = \max_{k\in\Gamma^+_{\theta^1}} \rho_k$, then by (3.34), $l \notin \Gamma^-_{\theta^2}$. Consequently, $l \in \Gamma^+_{c_1\theta^1+c_2\theta^2}$ and by (3.19), we must have

$$\max_{k\in\Gamma^+_{c_1\theta^1+c_2\theta^2}} \rho_k = \max_{k\in\Gamma^+_{\theta^1}} \rho_k.$$

By the same argument,

$$\max_{k\in\Gamma^-_{c_1\theta^1+c_2\theta^2}} \rho_k = \max_{k\in\Gamma^-_{\theta^1}} \rho_k,$$

and it follows that (3.7) holds for $c_1\theta^1 + c_2\theta^2$. □

LEMMA 3.7. *Fix γ, and suppose that* (3.7) *holds for θ^1 and* (3.8) *for θ^2. Then for $c_1, c_2 > 0$, Condition* 3.2 *holds for $c_1\theta^1 + c_2\theta^2$.*

PROOF. If

$$\max_{k\in\Gamma^-_{\theta^1}} \rho_k = \max_{k\in\Gamma^+_{\theta^1}} \rho_k > \max_{k\in\Gamma^-_{\theta^2}} \rho_k \vee \max_{k\in\Gamma^+_{\theta^2}} \rho_k, \tag{3.35}$$

then Lemma 3.6 implies $c_1\theta^1 + c_2\theta^2$ satisfies (3.7), so assume that

$$\max_{k\in\Gamma^-_{\theta^1}} \rho_k = \max_{k\in\Gamma^+_{\theta^1}} \rho_k \le \max_{k\in\Gamma^-_{\theta^2}} \rho_k \vee \max_{k\in\Gamma^+_{\theta^2}} \rho_k. \tag{3.36}$$

Then

$$\max_{k\in\Gamma^-_{c_1\theta^1+c_2\theta^2}} \rho_k \le \max_{k\in\Gamma^-_{\theta^1}} \rho_k \vee \max_{k\in\Gamma^-_{\theta^2}} \rho_k \le \max_{k\in\Gamma^-_{\theta^2}} \rho_k \vee \max_{k\in\Gamma^+_{\theta^2}} \rho_k$$

and

$$\max_{k\in\Gamma^+_{c_1\theta^1+c_2\theta^2}} \rho_k \le \max_{k\in\Gamma^+_{\theta^1}} \rho_k \vee \max_{k\in\Gamma^+_{\theta^2}} \rho_k \le \max_{k\in\Gamma^-_{\theta^2}} \rho_k \vee \max_{k\in\Gamma^+_{\theta^2}} \rho_k,$$

so

$$\max_{k\in\Gamma^-_{c_1\theta^1+c_2\theta^2}} \rho_k \vee \max_{k\in\Gamma^+_{c_1\theta^1+c_2\theta^2}} \rho_k \le \max_{k\in\Gamma^-_{\theta^2}} \rho_k \vee \max_{k\in\Gamma^+_{\theta^2}} \rho_k,$$

and since $\operatorname{supp}(c_1\theta^1 + c_2\theta^2) \supset \operatorname{supp}(\theta^2)$, (3.8) for θ^2 implies (3.8) for $c_1\theta^1 + c_2\theta^2$. □

If Condition 3.2 holds for θ^1 and θ^2 and $c_1, c_2 > 0$, then the previous lemmas imply Condition 3.2 holds for $c_1\theta^1 + c_2\theta^2$ except in one possible situation, that is,

$$\max_{k\in\Gamma^-_{\theta^1}} \rho_k = \max_{k\in\Gamma^+_{\theta^1}} \rho_k = \max_{k\in\Gamma^-_{\theta^2}} \rho_k = \max_{k\in\Gamma^+_{\theta^2}} \rho_k. \tag{3.37}$$

Since the species balance condition does not imply Condition 3.2 for $\theta = (1,1)$ for the system (Z_1^N, Z_2^N) given by (3.6), some additional condition must be required to be able to conclude Condition 3.2 holds for $c_1\theta^1 + c_2\theta^2$ when (3.37) holds. The following lemmas give such conditions.

LEMMA 3.8. *Fix $\gamma \in \mathbb{R}$, and suppose that Condition 3.2 holds for $\theta^1, \theta^2 \in [0,\infty)^{s_0}$. If $\Gamma^+_{\theta^1} \cap \Gamma^-_{\theta^2} = \varnothing$ or $\Gamma^-_{\theta^1} \cap \Gamma^+_{\theta^2} = \varnothing$ and $c_1, c_2 > 0$, then Condition 3.2 holds for $c_1\theta^1 + c_2\theta^2$.*

If (3.7) holds for θ_1 and θ_2, $\Gamma^+_{\theta^1} \cap \Gamma^-_{\theta^2} = \varnothing$ or $\Gamma^-_{\theta^1} \cap \Gamma^+_{\theta^2} = \varnothing$, and $c_1, c_2 > 0$, then (3.7) holds for $c_1\theta^1 + c_2\theta^2$.

REMARK 3.9. If no reaction that consumes a species in the support of θ^1 produces a species in the support of θ^2, then $\Gamma^-_{\theta^1} \cap \Gamma^+_{\theta^2} = \varnothing$. That condition is, of course, equivalent to the requirement that a reaction that produces a species in the support of θ^2 does not consume a species in the support of θ^1.

PROOF OF LEMMA 3.8. As noted, the previous lemmas cover all possible situations except in the case that (3.37) holds. Suppose $\Gamma^-_{\theta^1} \cap \Gamma^+_{\theta^2} = \varnothing$. If $\theta^1 \cdot \zeta_k < 0$, then $\theta^2 \cdot \zeta_k \le 0$ and $(c_1\theta^1 + c_2\theta^2) \cdot \zeta_k < 0$, and if $(c_1\theta^1 + c_2\theta^2) \cdot \zeta_k < 0$, then either $\theta^1 \cdot \zeta_k < 0$ or $\theta^2 \cdot \zeta_k < 0$, so

$$\max_{k \in \Gamma^-_{\theta^1}} \rho_k \le \max_{k \in \Gamma^-_{c_1\theta^1 + c_2\theta^2}} \rho_k \le \max_{k \in \Gamma^-_{\theta^1}} \rho_k \vee \max_{k \in \Gamma^-_{\theta^2}} \rho_k. \tag{3.38}$$

Similarly, noting that $\theta^2 \cdot \zeta_k > 0$ implies $\theta^1 \cdot \zeta_k \ge 0$,

$$\max_{k \in \Gamma^+_{\theta^2}} \rho_k \le \max_{k \in \Gamma^+_{c_1\theta^1 + c_2\theta^2}} \rho_k \le \max_{k \in \Gamma^+_{\theta^1}} \rho_k \vee \max_{k \in \Gamma^+_{\theta^2}} \rho_k. \tag{3.39}$$

But (3.37) implies equality holds throughout (3.38) and (3.39) and (3.7) holds for $c_1\theta^1 + c_2\theta^2$. $\square$

LEMMA 3.10. *Suppose (3.7) holds for θ^1 and θ^2 and for $\theta^1 - \frac{\theta^1 \cdot \zeta_k}{\theta^2 \cdot \zeta_k}\theta^2$ for all $k \in (\Gamma^+_{\theta^1} \cap \Gamma^-_{\theta^2}) \cup (\Gamma^-_{\theta^1} \cap \Gamma^+_{\theta^2})$. (Note that $-\frac{\theta^1 \cdot \zeta_k}{\theta^2 \cdot \zeta_k} > 0$.) Then (3.7) holds for $c_1\theta^1 + c_2\theta^2$ for all $c_1, c_2 > 0$.*

PROOF. By Lemma 3.6, we can restrict our attention to the case (3.37), and it is enough to consider $\theta^1 + c\theta^2$ for $c > 0$. For c sufficiently small, $\Gamma^+_{\theta^1} \subset \Gamma^+_{\theta^1 + c\theta^2} \subset \Gamma^+_{\theta^1} \cup \Gamma^+_{\theta^2}$ and $\Gamma^-_{\theta^1} \subset \Gamma^-_{\theta^1 + c\theta^2} \subset \Gamma^-_{\theta^1} \cup \Gamma^-_{\theta^2}$, so assuming (3.37), with reference to (3.19) and (3.20),

$$\max_{k \in \Gamma^+_{\theta^1 + c\theta^2}} \rho_k = \max_{k \in \Gamma^+_{\theta^1}} \rho_k = \max_{k \in \Gamma^-_{\theta^1 + c\theta^2}} \rho_k = \max_{k \in \Gamma^-_{\theta^1}} \rho_k.$$

Let

$$c_0 = \inf\Big\{c : \max_{k\in\Gamma^+_{\theta^1+c\theta^2}} \rho_k \neq \max_{k\in\Gamma^+_{\theta^1}} \rho_k \text{ or } \max_{k\in\Gamma^-_{\theta^1+c\theta^2}} \rho_k \neq \max_{k\in\Gamma^-_{\theta^1}} \rho_k\Big\},$$

and note that for $0 < c < c_0$, (3.7) holds for $\theta = \theta^1 + c\theta^2$. If $c_0 < \infty$, then for $\varepsilon > 0$ there must exist $c_0 \le c \le c_0 + \varepsilon$ and k such that either $k \in \Gamma^+_{\theta^1}$ and $(\theta^1 + c\theta^2)\cdot\zeta_k \le 0$ or $k \in \Gamma^-_{\theta^1}$ and $(\theta^1 + c\theta^2)\cdot\zeta_k \ge 0$. In either case, $0 < -\frac{\theta^1\cdot\zeta_k}{\theta^2\cdot\zeta_k} \le c$. Since for each such k and $c' < c_0$, $-\frac{\theta^1\cdot\zeta_k}{\theta^2\cdot\zeta_k} \ge c'$, it follows that $c_0 = -\frac{\theta^1\cdot\zeta_k}{\theta^2\cdot\zeta_k} > 0$. Consequently, by the assumptions of the lemma,

$$\max_{k\in\Gamma^+_{\theta^1+c_0\theta^2}} \rho_k = \max_{k\in\Gamma^-_{\theta^1+c_0\theta^2}} \rho_k < \max_{k\in\Gamma^+_{\theta^1}} \rho_k = \max_{k\in\Gamma^-_{\theta^1}} \rho_k. \tag{3.40}$$

But (3.40) can hold only if there exists $l^+ \in \Gamma^+_{\theta_1}$ such that $\rho_{l^+} = \max_{k\in\Gamma^+_{\theta^1}} \rho_k$ and $c_0 = -\frac{\theta^1\cdot\zeta_{l^+}}{\theta^2\cdot\zeta_{l^+}}$ and $l^- \in \Gamma^-_{\theta_1}$ such that $\rho_{l^-} = \max_{k\in\Gamma^-_{\theta^1}} \rho_k$ and $c_0 = -\frac{\theta^1\cdot\zeta_{l^-}}{\theta^2\cdot\zeta_{l^-}}$. Then, for $c > c_0$, $(\theta^1 + c\theta^2)\zeta_{l^+} < (\theta^1 + c_0\theta^2)\zeta_{l^+} = 0$, so $l^+ \in \Gamma^-_{\theta^1+c\theta^2}$. Similarly, $l^- \in \Gamma^+_{\theta^1+c\theta^2}$, and the lemma follows. $\square$

4. Derivation of limiting models. As can be seen from the examples, derivation of the limiting models can frequently be carried out by straightforward analysis of the stochastic equations. The results of this section take a more general approach and may be harder to apply than direct analysis of the stochastic equations, but they should give added confidence that the limits hold in great generality for complex models.

We assume throughout this section that $\lim_{N\to\infty} Z_i^{N,\gamma}(0)$ exists and is positive for all i. If

$$\gamma = r_1 \equiv \min_i \gamma_i = \min_i\Big(\alpha_i - \max_{k\in\Gamma_i^+\cup\Gamma_i^-}(\beta_k + \nu_k\cdot\alpha)\Big), \tag{4.1}$$

then $\lim_{N\to\infty} Z^{N,\gamma}$ exists, at least on some interval $[0, \tau_\infty)$ with $\tau_\infty > 0$, and is easy to calculate since on any time interval over which $\sup_{t\le T}|Z^{N,\gamma}(t)| < \infty$, each term

$$N^{-\alpha_i} Y_k\left(\int_0^t N^{\gamma+\rho_k}\lambda_k(Z^{N,\gamma}(s))\,ds\right)$$

either converges to zero (if $\alpha_i > \gamma + \rho_k$), is dependent on N only through $Z^{N,\gamma}$ (if $\alpha_i = \gamma + \rho_k = 0$), or is asymptotic to

$$\int_0^t \lambda_k(Z^{N,\gamma}(s))\,ds$$

(if $\alpha_i = \gamma + \rho_k > 0$), since

$$\lim_{N\to\infty} \sup_{u\le u_0} |N^{-\alpha_i} Y_k(N^{\alpha_i} u) - u| = 0, \qquad u_0 > 0.$$

The caveat regarding the interval $[0, \tau_\infty)$ reflects the fact that we have not ruled out "reaction" networks of the form $2S_1 \to 3S_1$, $S_1 \to \varnothing$ which would be modeled by

$$X_1(t) = X_1(0) + Y_1\biggl(\kappa_1 \int_0^t X_1(s)\bigl(X_1(s) - 1\bigr)\,ds\biggr) - Y_2\biggl(\kappa_2 \int_0^t X_1(s)\,ds\biggr)$$

and has positive probability of exploding in finite time, if $X_1(0) > 1$.

For $\alpha \ge 0$ and $\gamma \in \mathbb{R}$, define

$$\Gamma_\alpha^\gamma = \{k : \gamma + \rho_k = \alpha, D^\alpha \zeta_k \neq 0\}, \tag{4.2}$$

where

$$D^\alpha = \operatorname{diag}\bigl(\dots \mathbf{1}_{\{\alpha_i = \alpha\}} \dots\bigr). \tag{4.3}$$

THEOREM 4.1. *For r_1 defined by* (4.1), $Z^{N,r_1} \Rightarrow Z^{r_1}$ *on* $[0, \tau_\infty)$, *where if* $\alpha_i > 0$,

$$Z_i^{r_1}(t) = Z_i(0) + \sum_{k\in\Gamma_{\alpha_i}^{r_1}} \int_0^t \lambda_k(Z^{r_1}(s))\,ds(\nu'_{ik} - \nu_{ik}),$$

if $\alpha_i = 0$,

$$Z_i^{r_1}(t) = Z_i(0) + \sum_{k\in\Gamma_{\alpha_i}^{r_1}} Y_k\biggl(\int_0^t \lambda_k(Z^{r_1}(s))\,ds\biggr)(\nu'_{ik} - \nu_{ik})$$

and

$$\tau_\infty = \lim_{c\to\infty} \tau_c \equiv \inf\Bigl\{t : \sup_{s\le t} |Z^{r_1}(s)| \ge c\Bigr\}.$$

REMARK 4.2. By $Z^{N,r_1} \Rightarrow Z^{r_1}$ on $[0, \tau_\infty)$, we mean that there exist $\tau^{N,n}$ and τ^n such that $(Z^{N,r_1}(\cdot \wedge \tau^{N,n}), \tau^{N,n}) \Rightarrow (Z^{r_1}(\cdot \wedge \tau^n), \tau^n)$ and $\lim_{n\to\infty} \tau^n = \tau_\infty$.

We can write

$$\begin{aligned} Z^{r_1}(t) = Z(0) &+ \sum_{k : r_1+\rho_k>0} \int_0^t \lambda_k(Z^{r_1}(s)) D^{r_1+\rho_k} \zeta_k \\ &+ \sum_{k : r_1+\rho_k=0} Y_k\biggl(\int_0^t \lambda_k(Z^{r_1}(s))\,ds\biggr) D^0 \zeta_k. \end{aligned}$$

PROOF OF THEOREM 4.1. Let $\tau_{N,c} = \inf\{t : \sup_{s \le t} |Z^{N,r_1}(s)| \ge c\}$. The relative compactness of $\{Z^{N,r_1}(\cdot \wedge \tau_{N,c})\}$ follows from the uniform boundedness of $\lambda_k(Z^{N,r_1}(\cdot \wedge \tau_{N,c}))$. Then $(Z^{N,r_1}(\cdot \wedge \tau_{N,c}), \tau_{N,c}) \Rightarrow (Z^{r_1}(\cdot \wedge \tau_c), \tau_c)$ at least for all but countably many c. □

Note that $\gamma_\theta \ge \min_{i:\theta_i>0} \gamma_i$, so $r_1 = \min_{\theta \in [0,\infty)^{s_0}} \gamma_\theta$, and Condition 3.2 always holds for $\gamma = r_1$. Recall that $\alpha_\theta = \max_{i:\theta_i>0} \alpha_i$ and

$$Z_\theta^{N,\gamma}(t) = N^{-\alpha_\theta}\theta \cdot \Lambda_N^{-1} Z^{N,\gamma}(t) = N^{-\alpha_\theta} \sum_{i=1}^{s_0} \theta_i X_i^N(N^\gamma t).$$

If $\gamma_\theta = r_1$, then

$$\lim_{N\to\infty} Z_\theta^{N,\gamma_\theta} \Rightarrow \theta \cdot D^{\alpha_\theta} Z^{r_1}$$

on $[0, \tau_\infty)$.

If Condition 3.2 holds for some $\gamma > r_1$, then the balance equality (3.7) must hold for all $\theta \in [0,\infty)^{s_0}$ with $\gamma_\theta = r_1$. Let

$$\widehat{\gamma} = \sup\{\gamma\colon \text{Condition 3.2 holds}\}.$$

Either $\widehat{\gamma} = \infty$, that is, (3.7) holds for all θ, or $\widehat{\gamma} = \gamma_\theta$ for some θ. Assume that there is at least one $\theta \in [0,\infty)^{s_0}$ such that $\gamma_\theta > r_1$, that is, there is more than one natural time-scale. If $\widehat{\gamma} > r_1$, then

$$r_1 < r_2 \equiv \inf\{\gamma_\theta : \gamma_\theta > r_1\} \le \widehat{\gamma},$$

and r_2 should be the second time-scale for the system. Note that $D^\alpha \Lambda_N = \Lambda_N D^\alpha$ and that we can write

$$\begin{aligned}
Z^{N,r_2}(t) &= Z^N(0) + \sum_k Y_k\left(N^{r_2+\rho_k}\int_0^t \lambda_k(Z^{N,r_2}(s))\,ds\right)\Lambda_N\zeta_k \\
&= Z^N(0) + \sum_k N^{-(r_1+\rho_k)} Y_k\left(N^{r_2+\rho_k}\int_0^t \lambda_k(Z^{N,r_2}(s))\,ds\right) D^{r_1+\rho_k}\zeta_k \\
&\quad + \sum_k N^{-(r_2+\rho_k)} Y_k\left(N^{r_2+\rho_k}\int_0^t \lambda_k(Z^{N,r_2}(s))\,ds\right) D^{r_2+\rho_k}\zeta_k \\
&\quad + \sum_k Y_k\left(N^{r_2+\rho_k}\int_0^t \lambda_k(Z^{N,r_2}(s))\,ds\right)\Lambda_N(I - D^{r_1+\rho_k} - D^{r_2+\rho_k})\zeta_k,
\end{aligned} \tag{4.4}$$

where the third sum on the right should converge to zero.

Let $\mathbb{L}_1$ be the space spanned by

$$\mathbb{S}_1 = \{e_i : \exists k, e_i \cdot D^{r_1+\rho_k}\zeta_k \ne 0\},$$

and $\mathbb{L}_2$ be the space spanned by

$$\mathbb{S}_2 = \{\theta \in [0,\infty)^{s_0} : \theta \cdot D^{r_1+\rho_k}\zeta_k = 0, \forall k\}.$$

Let Π_1 be the projection onto $\mathbb{L}_1$ and Π_2 be the projection onto $\mathbb{L}_2$. Of course, $\mathbb{S}_2$ contains $\{e_i : e_i \notin \mathbb{S}_1\}$, but as in the example of Section 3.3, it may be larger. Consequently, the projections Π_1 and Π_2 are not necessarily orthogonal, but for any $x \in \mathbb{R}^{s_0}$, $x - \Pi_2 x \in \mathbb{L}_1$.

LEMMA 4.3. *For each* $x \in \mathbb{R}^{s_0}$, $x - \Pi_2 x \in \mathbb{L}_1$.

PROOF. Note that $\mathbb{L}_1 = \{x \in \mathbb{R}^{s_0} : e_i \cdot x = 0, \forall e_i \in \mathbb{S}_2\}$ and that for $e_i \in \mathbb{S}_2$, $e_i \cdot \Pi_2 x = e_i \cdot x$. Consequently, for $e_i \in \mathbb{S}_2$, $e_i \cdot (x - \Pi_2 x) = 0$ and $x - \Pi_2 x \in \mathbb{L}_1$. $\square$

With reference to (4.4),

$$\Pi_2 Z^{N,r_2}(t) \approx \Pi_2 Z^N(0) + \sum_k N^{-(r_2+\rho_k)} Y_k\left(N^{r_2+\rho_k}\int_0^t \lambda_k(Z^{N,r_2}(s))\,ds\right)\Pi_2 D^{r_2+\rho_k}\zeta_k,$$

since the projection of the first sum on the right in (4.4) is zero and the third sum on the right goes to zero.

Unfortunately, while r_2 can naturally be viewed as the second time scale, we cannot guarantee a priori that the system will converge to a nondegenerate model on that time scale. For example, consider the network

$$\varnothing \to S_1, \qquad \varnothing \to S_2, \qquad \varnothing \to S_3,$$
$$S_1 + S_2 \to \varnothing, \qquad S_1 + S_3 \to \varnothing$$

and assume that the parameters scale so that

$$X_1(t) = X_1(0) + Y_1(\kappa_1 t) - Y_2\left(\kappa_2 \int_0^t X_1(s)X_2(s)\,ds\right) - Y_5\left(\kappa_5 N^{-1}\int_0^t X_1(s)X_3(s)\,ds\right),$$

$$X_2(t) = X_2(0) + Y_3(\kappa_3 t) - Y_2\left(\kappa_2 \int_0^t X_1(s)X_2(s)\,ds\right),$$

$$X_3(t) = X_3(0) + Y_4(\kappa_4 N^{-1} t) - Y_5\left(\kappa_5 N^{-1}\int_0^t X_1(s)X_3(s)\,ds\right).$$

Then (3.7) is satisfied for all θ, $r_1 = 0$, and $r_2 = 1$. But if $\kappa_1 > \kappa_3$, $X_1(Nt) \to \infty$ and $X_2(Nt) \to 0$ for all $t > 0$.

The problem is that even though the balance equations are satisfied for the fast subnetwork (X_1, X_2), the subnetwork is not stable. Consequently, to guarantee convergence on the second time scale, we need some additional condition to ensure stability for the fast subnetwork so that the influence of the fast components can be averaged in the system on the second time scale.

Of course, with reference to (3.11) and (3.15), it is frequently possible to verify convergence without any special techniques, but we will outline a more systematic approach.

Define the random measure on $\mathbb{L}_1 \times [0,\infty)$ by

$$V_1^{N,r_2}(C \times [0,t]) = \int_0^t \mathbf{1}_C\big((I-\Pi_2)Z^{N,r_2}(s)\big)\,ds.$$

Assume that

$$V_1^{N,r_2} \Rightarrow V_1 \tag{4.5}$$

in the sense that

$$\int_{\mathbb{L}_1\times[0,t]} f(x)V_1^{N,r_2}(dx\times ds) \Rightarrow \int_{\mathbb{L}_1\times[0,t]} f(x)V_1(dx\times ds)$$

for all $f \in C_b(\mathbb{L}_1)$ and all $t > 0$. This requirement is essentially an ergodicity assumption on the fast subsystem.

For $q > 0$, define $\tau_q^N = \inf\{t : |\Pi_2 Z^{N,r_2}(t)| \ge q\}$ and

$$h_q(y) = \sup\biggl\{\sum_{k:\,\Pi_2 D^{r_2+\rho_k}\zeta_k \neq 0} \lambda_k(x) : |\Pi_2 x| \le q,\, x - \Pi_2 x = y\biggr\}.$$

Assume that $\psi_q : [0,\infty) \to [0,\infty)$ satisfies $\lim_{r\to\infty} r^{-1}\psi_q(r) = \infty$ and

$$\biggl\{\int_{\mathbb{L}_1\times[0,t\wedge\tau_q^N]} \psi_q(h_q(y))V_1^{N,r_2}(dy\times ds)\biggr\} \tag{4.6}$$

is stochastically bounded. In addition, assume

$$\sum_k |N^{r_2+\rho_k}\Lambda_N(I - D^{r_1+\rho_k} - D^{r_2+\rho_k})\zeta_k| \\ \times \int_{\mathbb{L}_1\times[0,t\wedge\tau_q^N]} \lambda_k\big(\Pi_2 Z^{N,r_2}(s) + y\big)V_1^{N,r_2}(dy\times ds) \to 0.$$

[Recall $|N^{r_2+\rho_k}\Lambda_N(I - D^{r_1+\rho_k} - D^{r_2+\rho_k})\zeta_k| \to 0$.] Then at least along a subsequence, for all but countably many q, $\Pi_2 Z^{N,r_2}(\cdot\wedge\tau_q^N)$ converges in distribution to a process $\widehat{Z}^{r_2}(\cdot\wedge\tau_q)$ and for k such that $\Pi_2 D^{r_2+\rho_k}\zeta_k \neq 0$, by Lemma A.6,

$$\int_0^{t\wedge\tau_q^N} \lambda_k(Z^{N,r_2}(s))\,ds \Rightarrow \int_{\mathbb{L}_1\times[0,t\wedge\tau_q]} \lambda_k\big(\widehat{Z}^{r_2}(s) + y\big)V_1(dy\times ds). \tag{4.7}$$

THEOREM 4.4. *Under the above assumptions, there exists a $\mathbb{L}_2$-valued process $\widehat{Z}^{r_2}$ and a random variable $\tau_\infty > 0$ such $\Pi_2 Z^{N,r_2}$ converges in distribution to $\widehat{Z}^{r_2}$ on $[0, \tau_\infty)$ where*

$$\widehat{Z}^{r_2}(t) = \Pi_2 Z(0) + \sum_{k:\, r_2+\rho_k>0} \int_{\mathbb{L}_1\times[0,t]} \lambda_k\big(\widehat{Z}^{r_2}(s)+y\big)V_1(dy\times ds)D^{r_2+\rho_k}\zeta_k$$
$$+ \sum_{k:\, r_2+\rho_k=0} Y_k\Big(\int_{\mathbb{L}_1\times[0,t]} \lambda_k\big(\widehat{Z}^{r_2}(s)+y\big)V_1(dy\times ds)\Big)D^{r_2+\rho_k}\zeta_k$$

for $t \in [0, \tau_\infty)$.

REMARK 4.5. The statement of this theorem is somewhat misleading. We are assuming V_1^{N,r_2} converges to V_1. Then given V_1, $\widehat{Z}^{r_2}$ is uniquely determined. However, as we will see in the next section, typically V_1 depends on $\widehat{Z}^{r_2}$. There we will give conditions under which the sequence of pairs $\{(V_1^{N,r_2}, Z^{N,r_2})\}$ is relatively compact. Then any limit point $(V_1, \widehat{Z}^{r_2})$ will satisfy the equations given by the present theorem, but it will still be necessary to show that the *pair* is uniquely determined.

PROOF OF THEOREM 4.4. As for the first time-scale, stopping the process at

$$\tau_q^N = \inf\{t : |\Pi_2 Z^{N,r_2}(t)| \geq q\}$$

ensures that $\{\Pi_2 Z^{N,r_2}(\cdot \wedge \tau_q^N)\}$ is relatively compact, and (4.7) ensures that any limit process satisfies the stochastic equations. Uniqueness for the limiting system then follows by the smoothness of the λ_k. $\square$

5. Averaging. Stochastic averaging methods go back at least to Khas'minskiĭ (1966a, 1966b). In this section we summarize the approach taken in Kurtz (1992). See that article for additional detail and references.

Recall that $\Lambda_N = \mathrm{diag}(N^{-\alpha_1}, \ldots, N^{-\alpha_{s_0}})$, $\rho_k = \beta_k + \nu_k \cdot \alpha$, and $\zeta_k = \nu_k' - \nu_k$. The generator for $Z^{N,0}$ is

$$\mathbb{B}_N f(z) = \sum_k N^{\rho_k}\lambda_k(z)\big(f(z+\Lambda_N\zeta_k) - f(z)\big).$$

Another way of characterizing r_1 is as the largest γ (possibly negative) such that $\lim_{N\to\infty} N^\gamma \mathbb{B}_N f(z)$ exists for each $f \in C_c^2(\mathbb{R}^m)$ and $z \in \mathbb{R}^m$. As before, define $D^\alpha = \mathrm{diag}(\ldots \mathbf{1}_{\{\alpha_i=\alpha\}}\ldots)$ and $\Gamma_\alpha^{r_1} = \{k : r_1 + \rho_k = \alpha, D^\alpha\zeta_k \neq 0\}$. Then

$$\begin{aligned}\mathbb{C}_0 f(x) &\equiv \lim_{N\to\infty} N^{r_1}\mathbb{B}_N f(x)\\ &= \sum_{k:\, r_1+\rho_k=0} \lambda_k(x)\big(f(x+D^0\zeta_k) - f(x)\big)\\ &\quad + \sum_{k:\, r_1+\rho_k>0} \lambda_k(x)D^{r_1+\rho_k}\zeta_k \cdot \nabla f(x),\end{aligned}$$

which is the generator for the limit of the system on the first time scale. The state space for the limit process is $\mathbb{E} = \prod_{i=1}^{s_0} \mathbb{E}_i$, where $\mathbb{E}_i = \mathbb{N}$ if $\alpha_i = 0$ and $\mathbb{E}_i = [0, \infty)$ if $\alpha_i > 0$.

By the definition of $\mathbb{L}_2$, $\Pi_2 D^{r_1+\rho_k}\zeta_k = 0$. Consequently, for $z \in \Pi_2\mathbb{E}$ and

$$\begin{aligned}\mathbb{E}_z &= \{y \in \mathbb{L}_1 : y = (I - \Pi_2)x, \Pi_2 x = z, x \in \mathbb{E}\},\\ \mathbb{C}^z f(y) &\equiv \mathbb{C}_0 f(z + y)\end{aligned}$$

defines a generator with state space $\mathbb{E}_z$.

As before, define

$$V_1^{N,r_2}(C \times [0,t]) = \int_0^t \mathbf{1}_C\big((I - \Pi_2)Z^{N,r_2}(s)\big)\,ds$$

and observe that

$$\begin{aligned}M_f^N(t) &= f(Z^{N,r_2}(t)) - f(Z^{N,r_2}(0)) - \int_0^t N^{r_2}\mathbb{B}_N f(Z^{N,r_2}(s))\,ds\\ &= f(Z^{N,r_2}(t)) - f(Z^{N,r_2}(0))\\ &\quad - \int_{\mathbb{L}_1\times[0,t]} N^{r_2}\mathbb{B}_N f\big(\Pi_2 Z^{N,r_2}(s) + y\big)V_1^{N,r_2}(dy \times ds)\end{aligned}$$

is a martingale. Since f and $N^{r_1}\mathbb{B}_N f$ are bounded by constants, $N^{r_1-r_2}M_f^N$ is bounded by a constant on any bounded time interval. It follows that $\{N^{r_1-r_2}M_f^N\}$ is relatively compact, any limit point is a martingale with initial value zero, and any limit point is Lipschitz continuous with Lipschitz constant $\sup_z |\mathbb{C}_0 f(z)|$. Since any continuous martingale with finite variation paths is constant, it follows that the limit must be zero. Combining these observations with those of the previous section, we have the following theorem.

THEOREM 5.1. *Suppose that $\{V_1^{N,r_2}\}$ is relatively compact and that for each $q > 0$, (4.6) is stochastically bounded. Selecting a convergent subsequence if necessary, let $Z_2^{r_2}$ and τ_∞ be as in the conclusion of Theorem 4.4. Then for all $f \in C_c^2(\mathbb{R}^{s_0})$,*

$$\begin{aligned}&\int_{\mathbb{L}_1\times[0,\tau_\infty)} \mathbb{C}_0 f\big(Z_2^{r_2}(s) + y\big)V_1(dy \times ds)\\ &\qquad = \int_{\mathbb{L}_1\times[0,\tau_\infty)} \mathbb{C}^{Z_2^{r_2}(s)} f(y)V_1(dy \times ds) = 0.\end{aligned}$$

If for each $z \in \Pi_2\mathbb{E}$, π^z is the unique stationary distribution for $\mathbb{C}^z$, then

$$V_1(dy \times ds) = \pi^{Z_2^{r_2}(s)}(dy)\,ds,$$

and the limiting equation in Theorem 4.4 becomes

$$Z_2^{r_2}(t) = \Pi_2 Z(0) + \sum_{k:\, r_2+\rho_k>0} \int_0^t \int_{\mathbb{L}_1} \lambda_k\big(Z_2^{r_2}(s)+y\big)\pi^{Z_2^{r_2}(s)}(dy)\,ds\,D^{r_2+\rho_k}\zeta_k$$
$$+ \sum_{k:\, r_2+\rho_k=0} Y_k\bigg(\int_0^t \int_{\mathbb{L}_1} \lambda_k\big(Z_2^{r_2}(s)+y\big)\pi^{Z_2^{r_2}(s)}(dy)\,ds\bigg) D^{r_2+\rho_k}\zeta_k$$

for $t \in [0, \tau_\infty)$.

REMARK 5.2. Assuming uniqueness, the system determines a piecewise deterministic Markov process in the sense of Davis (1993). If one defines

$$\beta_k(z) = \int_{\mathbb{L}_1} \lambda_k(z+y)\pi^z(dy), \qquad z \in \Pi_2\mathbb{E},$$

the description of the system will simplify.

We still need to address conditions for the relative compactness of the sequence of occupation measures. If $(I-\Pi_2)\mathbb{E}$ is compact, relative compactness is immediate. Otherwise, it is natural to look for some kind of Lyapunov function. Note that if $\gamma_c^N = \inf\{t : |Z^{N,r_2}(t)| \geq c\}$, then

$$f\big(Z^{N,r_2}(t\wedge\gamma_c^N)\big) - f(Z^{N,r_2}(0)) - \int_0^{t\wedge\gamma_c^N} N^{r_2}\mathbb{B}_N f(Z^{N,r_2}(s))\,ds$$

is a martingale for all locally bounded f.

LEMMA 5.3. *Let h_q and ψ_q be as in (4.6). Suppose that f_q^N are nonnegative functions and that there exist positive constants c_1, c_2 such that*

$$\sup_N N^{r_2}\mathbb{B}_N f_q^N(z) < c_1 - c_2\psi_q\big(h_q\big((I-\Pi_2)z\big)\big)$$

for all z satisfying $|\Pi_2 z| \leq q$ and for each $c \in \mathbb{R}$,

$$\sup\Big\{|(I-\Pi_2)z| : |\Pi_2 z| \text{ and } \sup_N N^{r_2}\mathbb{B}_N f_q^N(z) \geq c\Big\} < \infty.$$

Then for each $t > 0$, $\{V_1^{N,r_2}\}$ is relatively compact and (4.6) is stochastically bounded.

6. Examples. We give some additional examples that demonstrate how identifying exponents satisfying the balance condition leads to reasonable approximations to the original model. For a "production level" example, see the analysis of an E. coli heat shock model in Kang (2011).

6.1. *Goutsias's model of regulated transcription*. We consider the following model of transcription regulation introduced in Goutsias (2005) and studied further in Macnamara, Burrage and Sidje (2007). The model involves six species:

$$
\begin{aligned}
&X_1 = \#\text{ of } M && \text{Protein monomer},\\
&X_2 = \#\text{ of } D && \text{Transcription factor},\\
&X_3 = \#\text{ of } RNA && mRNA,\\
&X_4 = \#\text{ of } DNA && \text{Unbound } DNA,\\
&X_5 = \#\text{ of } DNA\cdot D && DNA \text{ bound at one site},\\
&X_6 = \#\text{ of } DNA\cdot 2D && DNA \text{ bound at two sites},
\end{aligned}
$$

and ten reactions:

$$
\begin{aligned}
RNA &\to RNA + M,\\
M &\to \varnothing,\\
DNA\cdot D &\to RNA + DNA\cdot D,\\
RNA &\to \varnothing,\\
DNA + D &\to DNA\cdot D,\\
DNA\cdot D &\to DNA + D,\\
DNA\cdot D + D &\to DNA\cdot 2D,\\
DNA\cdot 2D &\to DNA\cdot D + D,\\
M + M &\to D,\\
D &\to 2M.
\end{aligned}
$$

Taking the volume $V = 1$, the corresponding system of equations becomes

$$
\begin{aligned}
X_1(t) = X_1(0) &+ Y_1\biggl(\kappa_1'\int_0^t X_3(s)\,ds\biggr) + 2Y_{10}\biggl(\kappa_{10}'\int_0^t X_2(s)\,ds\biggr)\\
&- Y_2\biggl(\kappa_2'\int_0^t X_1(s)\,ds\biggr) - 2Y_9\biggl(\kappa_9'\int_0^t X_1(s)\bigl(X_1(s)-1\bigr)\,ds\biggr),\\
X_2(t) = X_2(0) &+ Y_6\biggl(\kappa_6'\int_0^t X_5(s)\,ds\biggr) + Y_8\biggl(\kappa_8'\int_0^t X_6(s)\,ds\biggr)\\
&+ Y_9\biggl(\kappa_9'\int_0^t X_1(s)\bigl(X_1(s)-1\bigr)\,ds\biggr) - Y_5\biggl(\kappa_5'\int_0^t X_2(s)X_4(s)\,ds\biggr)\\
&- Y_7\biggl(\kappa_7'\int_0^t X_2(s)X_5(s)\,ds\biggr) - Y_{10}\biggl(\kappa_{10}'\int_0^t X_2(s)\,ds\biggr),
\end{aligned}
$$

$$X_3(t) = X_3(0) + Y_3\left(\kappa_3' \int_0^t X_5(s)\,ds\right) - Y_4\left(\kappa_4' \int_0^t X_3(s)\,ds\right),$$

$$X_4(t) = X_4(0) + Y_6\left(\kappa_6' \int_0^t X_5(s)\,ds\right) - Y_5\left(\kappa_5' \int_0^t X_2(s)X_4(s)\,ds\right),$$

$$\begin{aligned} X_5(t) = X_5(0) &+ Y_5\left(\kappa_5' \int_0^t X_2(s)X_4(s)\,ds\right) + Y_8\left(\kappa_8' \int_0^t X_6(s)\,ds\right) \\ &- Y_6\left(\kappa_6' \int_0^t X_5(s)\,ds\right) - Y_7\left(\kappa_7' \int_0^t X_2(s)X_5(s)\,ds\right), \end{aligned}$$

$$X_6(t) = X_6(0) + Y_7\left(\kappa_7' \int_0^t X_2(s)X_5(s)\,ds\right) - Y_8\left(\kappa_8' \int_0^t X_6(s)\,ds\right).$$

6.2. *A scaling with two fast reactions*. In his analysis of the model, Goutsias assumes two time-scales and identifies reactions 9 and 10 as "fast" reactions. In our approach, that is the same as assuming $\beta_9 = \beta_{10} > \beta_1 = \cdots = \beta_8$, so we take $N_0 = 100$, $\beta_9 = \beta_{10} = 0$ and $\beta_1 = \cdots = \beta_8 = -1$. Recall the relationships $\kappa_k' = \kappa_k N_0^{\beta_k}$ (we are assuming the volume $V = 1$) and $\rho_k = \beta_k + \nu_k \cdot \alpha$. Employing the rate constants from Goutsias (2005), and taking $\alpha_i = 0$ for all i, we have Table 1.

Then, for $\gamma = 0$, $(Z_1^{N,0}, Z_2^{N,0})$ converges to the solution of

$$Z_1^0(t) = X_1(0) + 2Y_{10}\left(\kappa_{10} \int_0^t Z_2^0(s)\,ds\right) - 2Y_9\left(\kappa_9 \int_0^t Z_1^0(s)\big(Z_1^0(s) - 1\big)\,ds\right),$$

$$Z_2^0(t) = X_2(0) + Y_9\left(\kappa_9 \int_0^t Z_1^0(s)\big(Z_1^0(s) - 1\big)\,ds\right) - Y_{10}\left(\kappa_{10} \int_0^t Z_2^0(s)\,ds\right),$$

and for $k > 2$, $Z_k^{N,0}$ converges to $X_k(0)$.

TABLE 1
Scaling exponents for reaction rates

Rates		Scaled rates		ρ	
κ_1'	4.30×10^{-2}	κ_1	4.30	ρ_1	-1
κ_2'	7.00×10^{-4}	κ_2	0.07	ρ_2	-1
κ_3'	7.15×10^{-2}	κ_3	7.15	ρ_3	-1
κ_4'	3.90×10^{-3}	κ_4	0.390	ρ_4	-1
κ_5'	1.99×10^{-2}	κ_5	1.99	ρ_5	-1
κ_6'	4.79×10^{-1}	κ_6	47.9	ρ_6	-1
κ_7'	1.99×10^{-4}	κ_7	0.0199	ρ_7	-1
κ_8'	8.77×10^{-12}	κ_8	8.77×10^{-10}	ρ_8	-1
κ_9'	8.30×10^{-2}	κ_9	0.0830	ρ_9	0
κ_{10}'	5.00×10^{-1}	κ_{10}	0.500	ρ_{10}	0

For $\gamma = 1$, the kind of argument employed in (3.15) implies

$$\kappa_9 \int_0^t Z_1^{N,1}(s)\big(Z_1^{N,1}(s) - 1\big)\,ds - \int_0^t \kappa_{10} Z_2^{N,1}(s)\,ds \to 0, \tag{6.1}$$

but does not lead to a closed system for the limit of $(Z_3^{N,1}, \dots, Z_6^{N,1})$. To obtain a closed limiting system, we introduce the following auxiliary variable:

$$\begin{aligned}
Z_{12}^{N,1}(t) &= Z_1^{N,1}(t) + 2Z_2^{N,1}(t) \\
&= Z_{12}^N(0) + Y_1\biggl(\kappa_1 \int_0^t Z_3^{N,1}(s)\,ds\biggr) \\
&\quad + 2Y_6\biggl(\kappa_6 \int_0^t Z_5^{N,1}(s)\,ds\biggr) + 2Y_8\biggl(\kappa_8 \int_0^t Z_6^{N,1}(s)\,ds\biggr) \\
&\quad - 2Y_5\biggl(\kappa_5 \int_0^t Z_2^{N,1}(s) Z_4^{N,1}(s)\,ds\biggr) \\
&\quad - 2Y_7\biggl(\kappa_7 \int_0^t Z_2^{N,1}(s) Z_5^{N,1}(s)\,ds\biggr) \\
&\quad - Y_2\biggl(\kappa_2 \int_0^t Z_1^{N,1}(s)\,ds\biggr)
\end{aligned}$$

and observe that the conditional equilibrium distribution satisfies

$$\begin{aligned}
&\kappa_9(z_1+2)(z_1+1)\mu_s(z_1+2, z_2-1) + \kappa_{10}(z_2+1)\mu_s(z_1-2, z_2+1) \\
&\quad = \big(\kappa_9 z_1(z_1-1) + \kappa_{10} z_2\big)\mu_s(z_1, z_2)
\end{aligned}$$

and is uniquely determined by the requirement that

$$z_1 + 2z_2 = Z_{12}^1(s),$$

where Z_{12}^1 is the limit of $Z_{12}^{N,1}$. For $m = z_1 + 2z_2$, the conditional equilibrium distribution is

$$\mu_m(z_1, z_2) = M_m \frac{(\kappa_{10}/\kappa_9)^{z_1+z_2}}{z_1! z_2!}, \tag{6.2}$$

where M_m is a normalizing constant making μ_m a probability distribution on the collection of (z_1, z_2) such that z_1 and z_2 are nonnegative integers satisfying $z_1 + 2z_2 = m$. Define

$$\alpha(m) = \int z_2 \mu_m(dz_1, dz_2) = M_m \sum_{1 \le z_2 \le m/2} \frac{(\kappa_{10}/\kappa_9)^{(m-z_2)}}{(m-2z_2)!(z_2-1)!} \tag{6.3}$$

and observe that $m - 2\alpha(m) = \int z_1 \mu_m(dz_1, dz_2)$. Then $(Z_{12}^{N,1}, Z_3^{N,1}, \ldots, Z_6^{N,1})$ converges to the solution of

$$
\begin{aligned}
Z_{12}^1(t) = Z_{12}^1(0) &+ Y_1\biggl(\kappa_1 \int_0^t Z_3^1(s)\,ds\biggr) + 2Y_6\biggl(\kappa_6 \int_0^t Z_5^1(s)\,ds\biggr) \\
&+ 2Y_8\biggl(\kappa_8 \int_0^t Z_6^1(s)\,ds\biggr) - 2Y_5\biggl(\kappa_5 \int_0^t \alpha(Z_{12}^1(s)) Z_4^1(s)\,ds\biggr) \\
&- 2Y_7\biggl(\kappa_7 \int_0^t \alpha(Z_{12}^1(s)) Z_5^1(s)\,ds\biggr) \\
&- Y_2\biggl(\kappa_2 \int_0^t \bigl(Z_{12}^1(s) - 2\alpha(Z_{12}^1(s))\bigr)\,ds\biggr), \\
Z_3^1(t) = Z_3^1(0) &+ Y_3\biggl(\kappa_3 \int_0^t Z_5^1(s)\,ds\biggr) - Y_4\biggl(\kappa_4 \int_0^t Z_3^1(s)\,ds\biggr), \\
Z_4^1(t) = Z_4^1(0) &+ Y_6\biggl(\kappa_6 \int_0^t Z_5^1(s)\,ds\biggr) - Y_5\biggl(\kappa_5 \int_0^t \alpha(Z_{12}^1(s)) Z_4^1(s)\,ds\biggr), \\
Z_5^1(t) = Z_5^1(0) &+ Y_5\biggl(\kappa_5 \int_0^t \alpha(Z_{12}^1(s)) Z_4^1(s)\,ds\biggr) + Y_8\biggl(\kappa_8 \int_0^t Z_6^1(s)\,ds\biggr), \\
&- Y_6\biggl(\kappa_6 \int_0^t Z_5^1(s)\,ds\biggr) - Y_7\biggl(\kappa_7 \int_0^t \alpha(Z_{12}^1(s)) Z_5^1(s)\,ds\biggr), \\
Z_6^1(t) = Z_6^1(0) &+ Y_7\biggl(\kappa_7 \int_0^t \alpha(Z_{12}^1(s)) Z_5^1(s)\,ds\biggr) - Y_8\biggl(\kappa_8 \int_0^t Z_6^1(s)\,ds\biggr),
\end{aligned}
$$

which is essentially the approximation obtained by Goutsias. Note that the "fast" reactions, reactions 9 and 10, have been eliminated from the model.

This system is not entirely satisfactory as $\alpha(m)$ is not computable analytically. For simulations, values of $\alpha(m)$ could be precomputed using (6.3). E, Liu and Vanden-Eijnden (2007) suggest a Monte Carlo approach for computing $\alpha(m)$ as needed. Goutsias suggests a way of approximating the transition rates which is equivalent to the following: The limit in (6.1) implies

$$
\kappa_{10}\alpha(m) = \kappa_9 \int z_1(z_1 - 1)\mu_m(dz_1, dz_2) \tag{6.4}
$$

as can be verified directly from the definition of μ_m. A moment closure argument suggests replacing (6.4) by

$$
\begin{aligned}
\kappa_{10}\alpha(m) &= \kappa_9 \int z_1 \mu_m(dz_1, dz_2) \int (z_1 - 1)\mu_m(dz_1, dz_2) \\
&= \kappa_9\bigl(m - 2\alpha(m)\bigr)\bigl(m - 2\alpha(m) - 1\bigr),
\end{aligned}
$$

which gives a quadratic equation for the approximation for $\alpha(m)$.

TABLE 2
Balance equations

Variable	**Balance equation**
X_1	$\rho_1 \vee \rho_{10} = \rho_2 \vee \rho_9$
X_2	$\rho_6 \vee \rho_8 \vee \rho_9 = \rho_5 \vee \rho_7 \vee \rho_{10}$
X_3	$\rho_3 = \rho_4$
X_4	$\rho_5 = \rho_6$
X_5	$\rho_5 \vee \rho_8 = \rho_6 \vee \rho_7$
X_6	$\rho_7 = \rho_8$
$X_1 + 2X_2 + 2X_5 + 4X_6$	$\rho_1 = \rho_2$
$X_2 + X_5 + 2X_6$	$\rho_9 = \rho_{10}$
$X_5 + X_6$	$\rho_5 = \rho_6$
$X_4 + X_5 + X_6$	$0 = 0$
$X_4 + X_5$	$\rho_8 = \rho_7$

6.3. *Alternative scaling*. Observe that $\kappa_9' < \kappa_6'$, so reaction 6 is actually "faster" than reaction 9. Consequently, it is reasonable to look for a different solution of the balance conditions with $\beta_{10} = \beta_6 > \beta_9$. Drop the assumption that $\alpha_i = 0$, and consider a subset of the balance equations. Recall that $\rho_k = \beta_k + \nu_k \cdot \alpha$.

We take $N_0 = 100$, $\alpha_1 = \alpha_2 = 1$, and $\alpha_i = 0$ for $3 \leq i \leq 6$. We see that the following exponents satisfy the balance conditions and the additional requirement that $\kappa_k' \geq \kappa_l'$ implies $\beta_k \geq \beta_l$, except for β_8, the exponent associated with the extremely small rate constant κ_8'. Recall that κ_k is determined by the requirement $\kappa_k' = \kappa_k N_0^{\beta_k}$.

TABLE 3
Scaling exponents for reaction rates

	Rates		**Exponents**		**Scaled rates**		$\boldsymbol{\rho}$
κ_1'	4.30×10^{-2}	β_1	-1	κ_1	4.30	ρ_1	-1
κ_2'	7.00×10^{-4}	β_2	-2	κ_2	7.00	ρ_2	-1
κ_3'	7.15×10^{-2}	β_3	-1	κ_3	7.15	ρ_3	-1
κ_4'	3.90×10^{-3}	β_4	-1	κ_4	0.390	ρ_4	-1
κ_5'	1.99×10^{-2}	β_5	-1	κ_5	1.99	ρ_5	0
κ_6'	4.79×10^{-1}	β_6	0	κ_6	0.479	ρ_6	0
κ_7'	1.99×10^{-4}	β_7	-3	κ_7	199	ρ_7	-2
κ_8'	8.77×10^{-12}	β_8	-2	κ_8	8.77×10^{-8}	ρ_8	-2
κ_9'	8.30×10^{-2}	β_9	-1	κ_9	8.30	ρ_9	1
κ_{10}'	5.00×10^{-1}	β_{10}	0	κ_{10}	0.500	ρ_{10}	1

Defining $Z_i^{N,\gamma}(t) = N^{-\alpha_i} X_i^N(N^\gamma t)$ and $\kappa_k = N_0^{-\beta_k}\kappa_k'$,

$$\begin{aligned}
Z_1^{N,\gamma}(t) &= Z_1^N(0) + N^{-1}Y_1\left(\int_0^t \kappa_1 N^{\gamma-1} Z_3^{N,\gamma}(s)\,ds\right) \\
&\quad + 2N^{-1}Y_{10}\left(\int_0^t \kappa_{10} N^{\gamma+1} Z_2^{N,\gamma}(s)\,ds\right) \\
&\quad - N^{-1}Y_2\left(\int_0^t \kappa_2 N^{\gamma-1} Z_1^{N,\gamma}(s)\,ds\right) \\
&\quad - 2N^{-1}Y_9\left(\int_0^t \kappa_9 N^{\gamma+1} Z_1^{N,\gamma}(s)\big(Z_1^{N,\gamma}(s) - N^{-1}\big)\,ds\right), \\
Z_2^{N,\gamma}(t) &= Z_2^N(0) + N^{-1}Y_6\left(\int_0^t \kappa_6 N^{\gamma} Z_5^{N,\gamma}(s)\,ds\right) \\
&\quad + N^{-1}Y_8\left(\int_0^t \kappa_8 N^{\gamma-2} Z_6^{N,\gamma}(s)\,ds\right) \\
&\quad + N^{-1}Y_9\left(\int_0^t \kappa_9 N^{\gamma+1} Z_1^{N,\gamma}(s)\big(Z_1^{N,\gamma}(s) - N^{-1}\big)\,ds\right) \\
&\quad - N^{-1}Y_5\left(\int_0^t \kappa_5 N^{\gamma} Z_2^{N,\gamma}(s) Z_4^{N,\gamma}(s)\,ds\right) \\
&\quad - N^{-1}Y_7\left(\int_0^t \kappa_7 N^{\gamma-2} Z_2^{N,\gamma}(s) Z_5^{N,\gamma}(s)\,ds\right) \\
&\quad - N^{-1}Y_{10}\left(\int_0^t \kappa_{10} N^{\gamma+1} Z_2^{N,\gamma}(s)\,ds\right), \\
Z_3^{N,\gamma}(t) &= Z_3^N(0) + Y_3\left(\int_0^t \kappa_3 N^{\gamma-1} Z_5^{N,\gamma}(s)\,ds\right) \\
&\quad - Y_4\left(\int_0^t \kappa_4 N^{\gamma-1} Z_3^{N,\gamma}(s)\,ds\right), \\
Z_4^{N,\gamma}(t) &= Z_4^N(0) + Y_6\left(\int_0^t \kappa_6 N^{\gamma} Z_5^{N,\gamma}(s)\,ds\right) \\
&\quad - Y_5\left(\int_0^t \kappa_5 N^{\gamma} Z_2^{N,\gamma}(s) Z_4^{N,\gamma}(s)\,ds\right), \\
Z_5^{N,\gamma}(t) &= Z_5^N(0) + Y_5\left(\int_0^t \kappa_5 N^{\gamma} Z_2^{N,\gamma}(s) Z_4^{N,\gamma}(s)\,ds\right) \\
&\quad + Y_8\left(\int_0^t \kappa_8 N^{\gamma-2} Z_6^{N,\gamma}(s)\,ds\right) - Y_6\left(\int_0^t \kappa_6 N^{\gamma} Z_5^{N,\gamma}(s)\,ds\right) \\
&\quad - Y_7\left(\int_0^t \kappa_7 N^{\gamma-2} Z_2^{N,\gamma}(s) Z_5^{N,\gamma}(s)\,ds\right),
\end{aligned}$$

$$Z_6^{N,\gamma}(t) = Z_6^N(0) + Y_7\left(\int_0^t \kappa_7 N^{\gamma-2} Z_2^{N,\gamma}(s) Z_5^{N,\gamma}(s)\,ds\right) - Y_8\left(\int_0^t \kappa_8 N^{\gamma-2} Z_6^{N,\gamma}(s)\,ds\right).$$

Useful auxiliary variables include

$$\begin{aligned}
&N Z_1^{N,\gamma}(t) + 2N Z_2^{N,\gamma}(t) + 2Z_5^{N,\gamma}(t) + 4Z_6^{N,\gamma}(t)\\
&\quad = N Z_1^N(0) + 2N Z_2^N(0) + 2Z_5^N(0) + 4Z_6^N(0)\\
&\qquad + Y_1\left(\int_0^t \kappa_1 N^{\gamma-1} Z_3^{N,\gamma}(s)\,ds\right) - Y_2\left(\int_0^t \kappa_2 N^{\gamma-1} Z_1^{N,\gamma}(s)\,ds\right),\\
&N Z_2^{N,\gamma}(t) + Z_5^{N,\gamma}(t) + 2Z_6^{N,\gamma}(t)\\
&\quad = N Z_2^N(0) + Z_5^N(0) + 2Z_6^N(0)\\
&\qquad + Y_9\left(\int_0^t \kappa_9 N^{\gamma+1} Z_1^{N,\gamma}(s)\big(Z_1^{N,\gamma}(s) - N^{-1}\big)\,ds\right)\\
&\qquad - Y_{10}\left(\int_0^t \kappa_{10} N^{\gamma+1} Z_2^{N,\gamma}(s)\,ds\right),\\
&Z_5^{N,\gamma}(t) + Z_6^{N,\gamma}(t)\\
&\quad = Z_5^N(0) + Z_6^N(0) + Y_5\left(\int_0^t \kappa_5 N^{\gamma} Z_2^{N,\gamma}(s) Z_4^{N,\gamma}(s)\,ds\right)\\
&\qquad - Y_6\left(\int_0^t \kappa_6 N^{\gamma} Z_5^{N,\gamma}(s)\,ds\right),\\
&Z_4^{N,\gamma}(t) + Z_5^{N,\gamma}(t) + Z_6^{N,\gamma}(t) = Z_4^N(0) + Z_5^N(0) + Z_6^N(0),\\
&Z_4^{N,\gamma}(t) + Z_5^{N,\gamma}(t)\\
&\quad = Z_4^N(0) + Z_5^N(0) + Y_8\left(\int_0^t \kappa_8 N^{\gamma-2} Z_6^{N,\gamma}(s)\,ds\right)\\
&\qquad - Y_7\left(\int_0^t \kappa_7 N^{\gamma-2} Z_2^{N,\gamma}(s) Z_5^{N,\gamma}(s)\,ds\right).
\end{aligned}$$

For $\gamma = 0$, the limiting system is the piecewise deterministic model

$$\begin{aligned}
Z_1^0(t) &= Z_1(0) + \int_0^t \big(2\kappa_{10} Z_2^0(s) - 2\kappa_9 Z_1^0(s)^2\big)\,ds,\\
Z_2^0(t) &= Z_2(0) + \int_0^t \big(\kappa_9 Z_1^0(s)^2 - \kappa_{10} Z_2^0(s)\big)\,ds,\\
Z_4^0(t) &= Z_4(0) + Y_6\left(\int_0^t \kappa_6 Z_5^0(s)\,ds\right) - Y_5\left(\int_0^t \kappa_5 Z_2^0(s) Z_4^0(s)\,ds\right), \qquad (6.5)
\end{aligned}$$

$$Z_5^0(t) = Z_5(0) + Y_5\left(\int_0^t \kappa_5 Z_2^0(s) Z_4^0(s)\, ds\right) - Y_6\left(\int_0^t \kappa_6 Z_5^0(s)\, ds\right)$$

with $Z_3^0(t) \equiv Z_3(0)$ and $Z_6^0(t) \equiv Z_6(0)$.

For $\gamma = 1$, we introduce the auxiliary variables

$$\begin{aligned} Z_{12}^{N,1}(t) &\equiv Z_1^{N,1}(t) + 2Z_2^{N,1}(t), \\ Z_{45}^{N,1}(t) &\equiv Z_4^{N,1}(t) + Z_5^{N,1}(t) \\ &= Z_4^N(0) + Z_5^N(0) + Y_8\left(\int_0^t \kappa_8 N^{-1} Z_6^{N,1}(s)\, ds\right) \\ &\quad - Y_7\left(\int_0^t \kappa_7 N^{-1} Z_2^{N,1}(s) Z_5^{N,1}(s)\, ds\right). \end{aligned}$$

Observing that $Z_{12}^{N,1}$ is asymptotically the same as $Z_1^{N,1} + 2Z_2^{N,1} + 2N^{-1}Z_5^{N,1} + 4N^{-1}Z_6^{N,1}$, $Z_{12}^{N,1}$ converges to $Z_{12}^1(t) \equiv Z_{12}(0) = \lim_{N\to\infty}(Z_1^N(0) + 2Z_2^N(0))$. In particular, Z_{12}^1 is constant in time. We also have $Z_{45}^1(t) \equiv Z_{45}(0) = \lim_{N\to\infty}(Z_4^N(0) + Z_5^N(0))$.

Let $V^{N,1}$ denote the occupation measure for $(Z_1^{N,1}, Z_2^{N,1}, Z_4^{N,1}, Z_5^{N,1})$. The stochastic boundedness of $Z_{12}^{N,1}$ and $Z_{45}^{N,1}$ ensures the relative compactness of $\{V^{N,1}\}$, and as in Section 5, $V^{N,1}$ converges to $V^1(dz, ds) = v_s(dz)\, ds$, where v_s satisfies

$$\int \mathbb{C} f\, v_s(dz) = 0$$

and

$$\begin{aligned} \mathbb{C} f(z_1, z_2, z_4, z_5) = (\kappa_{10} z_2 - \kappa_9 z_1^2)&\big(2\partial_{z_1} f(z) - \partial_{z_2} f(z)\big) \\ &+ \kappa_6 z_5\big(f(z + e_4 - e_5) - f(z)\big) \\ &+ \kappa_5 z_2 z_4\big(f(z - e_4 + e_5) - f(z)\big). \end{aligned}$$

Consequently, v_s is uniquely determined for each s by the requirement that $z_1 + 2z_2 = Z_{12}^1(s) = Z_{12}(0)$ and $z_4 + z_5 = Z_{45}^1(s) = Z_{45}(0)$, and, hence,

$$\begin{aligned} v_s(dz) &= \delta_{\varphi_1(Z_{12}(0))}(dz_1)\delta_{\varphi_2(Z_{12}(0))}(dz_2) \\ &\quad \times B\left(Z_{45}(0), \frac{\kappa_6}{\kappa_6 + \kappa_5 \varphi_2(Z_{12}(0))}, dz_4, dz_5\right), \end{aligned}$$

where

$$\varphi_1(y) = \frac{\sqrt{\kappa_{10}^2 + 8\kappa_9\kappa_{10} y} - \kappa_{10}}{4\kappa_9},$$

$$\varphi_2(y) = \frac{4\kappa_9 y + \kappa_{10} - \sqrt{\kappa_{10}^2 + 8\kappa_9\kappa_{10} y}}{8\kappa_9}$$

and $B(n, p, dz_4, dz_5)$ is given by the binomial distribution

$$P\{Z_4 = k, Z_5 = n-k\} = \binom{n}{k} p^k (1-p)^{n-k}.$$

Averaging gives

$$\begin{aligned} Z_3^1(t) &= Z_3(0) + Y_3\left(\int_0^t \frac{\kappa_3\kappa_5\varphi_2(Z_{12}(0))}{\kappa_6 + \kappa_5\varphi_2(Z_{12}(0))} Z_{45}(0)\,ds\right) \\ &\quad - Y_4\left(\int_0^t \kappa_4 Z_3^1(s)\,ds\right). \end{aligned} \tag{6.6}$$

Finally, for $\gamma = 2$, dividing the equation for $Z_3^{N,2}$ by N, we see that

$$\int_0^t Z_3^{N,2}(s)\,ds \approx \frac{\kappa_3}{\kappa_4}\int_0^t Z_5^{N,2}(s)\,ds,$$

and $(Z_{12}^{N,2}, Z_{45}^{N,2}, Z_6^{N,2})$ converges to the solution of

$$\begin{aligned} Z_{12}^2(t) &= Z_{12}(0) + \int_0^t \left(\frac{\kappa_1\kappa_3}{\kappa_4}\overline{Z}_5^2(s) - \kappa_2\varphi_1(Z_{12}^2(s))\right) ds, \\ Z_{45}^2(t) &= Z_{45}(0) + Y_8\left(\int_0^t \kappa_8 Z_6^2(s)\,ds\right) - Y_7\left(\int_0^t \kappa_7\varphi_2(Z_{12}^2(s))\overline{Z}_5^2(s)\,ds\right), \\ Z_6^2(t) &= Z_6(0) + Y_7\left(\int_0^t \kappa_7\varphi_2(Z_{12}^2(s))\overline{Z}_5^2(s)\,ds\right) - Y_8\left(\int_0^t \kappa_8 Z_6^2(s)\,ds\right), \\ \overline{Z}_5^2(t) &= \frac{\kappa_5\varphi_2(Z_{12}^2(t))}{\kappa_6 + \kappa_5\varphi_2(Z_{12}^2(t))} Z_{45}^2(t). \end{aligned}$$

6.3.1. *Simulation results.* We compare simulation results for the full model with the approximations given by the limiting systems. The mean and standard deviations of the number of molecules for each species or for the auxiliary variables of interest are given from 100 simulations of the full model and from 1000 simulations of the limiting systems. The evolution of the processes in the full model is approximated by the evolution of the processes in the limiting system using the relationship

$$X_i(t) \equiv X_i^{N_0}(t) \approx N_0^{\alpha_i} Z_i^{\gamma}(tN_0^{-\gamma}).$$

The initial values are taken as $X_1(0) = 2$, $X_2(0) = 6$, $X_5(0) = 2$ and all other values equal to zero.

For $\gamma = 0$, we observe the evolution of the processes during the time interval $[0, 100]$. The full model is reduced to the four-dimensional hybrid model (6.5) in which Z_1^0 and Z_2^0 are the solution of a pair of ordinary differential equations and Z_4^0 and Z_5^0 are discrete with transition intensities depending on Z_2^0. The evolution of X_1, X_2, X_4 and X_5 in the full model is given in Figure 1 and the evolution of

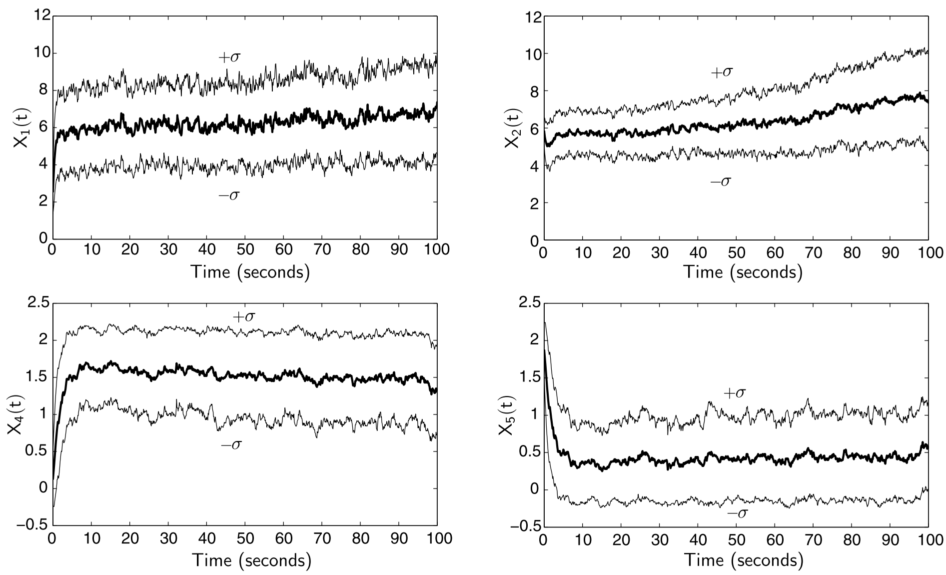

FIG. 1. *Simulation of the full model during $t = 0$ to $t = 100$.*

the approximation is given in Figure 2. The exact simulations of the full model are done using Gillespie's stochastic simulation algorithm (SSA) from Gillespie (1977). For the approximation, Z_1^0 and Z_2^0 are solved by the Matlab ODE solver,

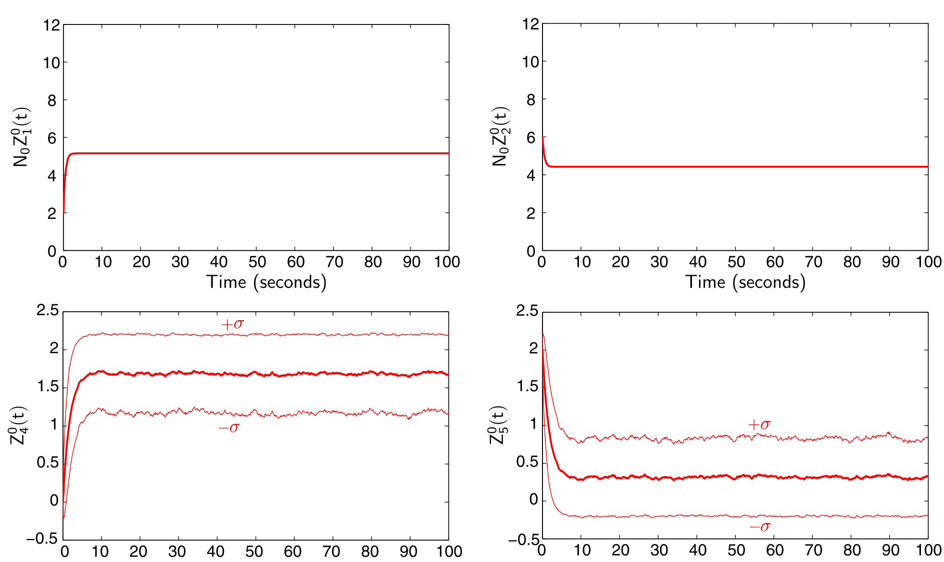

FIG. 2. *Approximation using the limiting model for $\gamma = 0$ in the alternative scaling.*

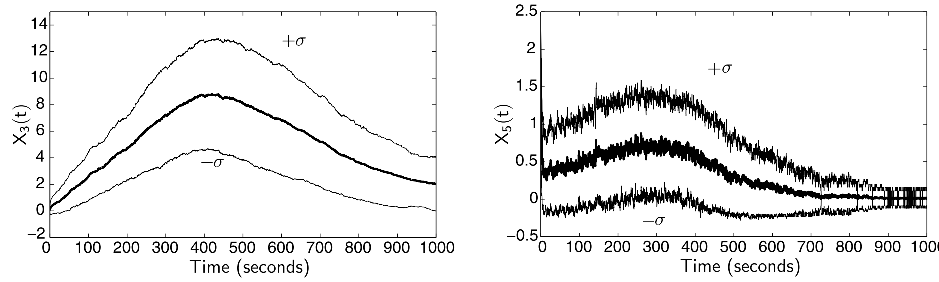

FIG. 3. *Simulation of the full model during $t = 0$ to $t = 1000$.*

and Z_4^0 and Z_5^0 are computed by Gillespie's SSA taking Z_2^0 from the solution of ODE. The evolution of X_1 and X_2 are well captured by Z_1^0 and Z_2^0 in Figure 2. These deterministic values approximate the evolution of the mean of X_1 and X_2 given in Figure 1 except for a slight increase over time in the simulation of the full model. Note that in the approximate model $Z_1^0(t) + 2Z_2^0(t)$ is constant, but that is not the case in the full model.

For $\gamma = 1$, we consider the evolution of the processes on the time interval $[0, 1000]$. The full model is reduced to the one-dimensional limiting system (6.6) with a single jump process Z_3^1. Comparing the governing equations for $Z_3^{N,1}$ and Z_3^1, the different behavior of the evolution of the two processes comes from the difference between $Z_5^{N,1}$ and $\overline{Z}_5^1(t) = \frac{\kappa_5 \varphi_2(Z_{12}(0))}{\kappa_6 + \kappa_5 \varphi_2(Z_{12}(0))} Z_{45}(0)$. Therefore, plots of the evolution of both X_3 and X_5 in the exact simulation are given in Figure 3. In Figure 4, the evolution of Z_3^1 and of $\overline{Z}_5^1$ is given. For both exact and approximate simulations, we use Gillespie's SSA. In Figure 3, $Z_5^{N,1}$ increases slightly and then decreases to zero. Since $\overline{Z}_5^1$ is approximated as constant in Figure 4, the increase during the early time and the decrease to zero of X_3 is not captured by the approximation.

For $\gamma = 2$, the simulation is carried out on the time interval $[0, 10{,}000]$. The three-dimensional limiting model is piecewise deterministic and includes the aux-

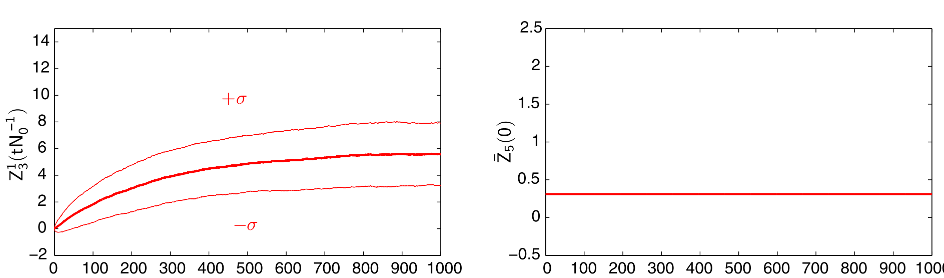

FIG. 4. *Approximation using the limiting model for $\gamma = 1$ in the alternative scaling.*

iliary variables Z_{12}^2, Z_{45}^2 and the species abundance Z_6^2. Z_{12}^2 is governed by a random differential equation driven by a component of the jump process, Z_{45}^2. Z_{45}^2 and Z_6^2 are discrete with transition intensities that depend on Z_{12}^2. Since there is mutual dependence between the continuous and discrete components, we modify Gillespie's SSA to simulate the limiting system. Here is a brief description of the simulation method for the limiting system.

(1) Assume that the process has been simulated up to t_i, the ith jump time of the jump process. Simulate a unit exponential random variable Δ by simulating a uniform [0, 1] random number r_1 and setting $\Delta = \log \frac{1}{r_1}$.

(2) Solve the differential equation for Z_{12}^2 starting at $Z_{12}^2(t_i)$ holding $Z_{45}^2(t) = Z_{45}^2(t_i)$ and $Z_6^2(t) = Z_6^2(t_i)$ until time t_{i+1} satisfying

$$\begin{aligned}
&\int_{t_i}^{t_{i+1}} \big(\kappa_7 \varphi_2(Z_{12}^2(s))\overline{Z}_5^2(s) + \kappa_8 Z_6^2(s)\big)\, ds \\
&\qquad = Z_{45}^2(t_i) \int_{t_i}^{t_{i+1}} \frac{\kappa_5\kappa_7\varphi_2(Z_{12}^2(s))^2}{\kappa_6 + \kappa_5\varphi_2(Z_{12}^2(s))}\, ds + \kappa_8 Z_6^2(t_i)(t_{i+1} - t_i) \\
&\qquad = \Delta.
\end{aligned}$$

(We compute the integral by the trapezoid rule using the grid points from the ODE solver.)

(3) Simulate a uniform [0, 1] random number r_2. If

$$\begin{aligned}
r_2 &\le \frac{\kappa_7\varphi_2(Z_{12}^2(t_{i+1}))\overline{Z}_5^2(t_{i+1}-)}{\kappa_7\varphi_2(Z_{12}^2(t_{i+1}))\overline{Z}_5^2(t_{i+1}-) + \kappa_8 Z_6^2(t_{i+1}-)} \\
&= \frac{\kappa_5\kappa_7\varphi_2(Z_{12}^2(t_{i+1}))^2 Z_{45}^2(t_i)}{\kappa_5\kappa_7\varphi_2(Z_{12}^2(t_{i+1}))^2 Z_{45}^2(t_i) + \kappa_8 Z_6^2(t_i)(\kappa_6 + \kappa_5\varphi_2(Z_{12}^2(t_{i+1})))},
\end{aligned} \tag{6.7}$$

set

$$\begin{pmatrix} Z_{45}^2(t_{i+1}) \\ Z_6^2(t_{i+1}) \end{pmatrix} = \begin{pmatrix} Z_{45}^2(t_i) \\ Z_6^2(t_i) \end{pmatrix} + \begin{pmatrix} -1 \\ 1 \end{pmatrix}$$

and if the reverse inequality holds in (6.7), set

$$\begin{pmatrix} Z_{45}^2(t_{i+1}) \\ Z_6^2(t_{i+1}) \end{pmatrix} = \begin{pmatrix} Z_{45}^2(t_i) \\ Z_6^2(t_i) \end{pmatrix} + \begin{pmatrix} 1 \\ -1 \end{pmatrix}.$$

(4) Go back to step (1).

Comparing plots for $X_1(t) + 2X_2(t)$ in Figure 5 and for $N_0 Z_{12}^2(t N_0^{-2})$ in Figure 6, the plot in the approximation increases more rapidly at early times and starts to drop earlier than the plot in the exact simulation. Also, the peak level in the approximation is much lower than the peak level in the exact simulation.

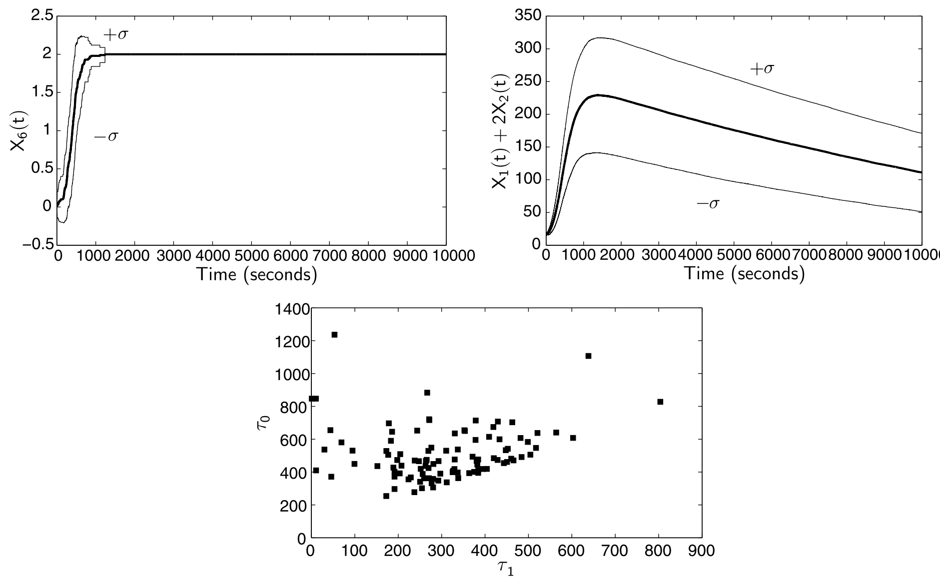

FIG. 5. *Simulation of the full model during $t = 0$ to $t = 10{,}000$.*

Since $\kappa_8 = 8.77 \times 10^{-8}$ is small compared to the time interval, reaction 8 will rarely occur on the time scales we are considering. We retained this reaction in the limiting model only to emphasize that a long time after the model appears to

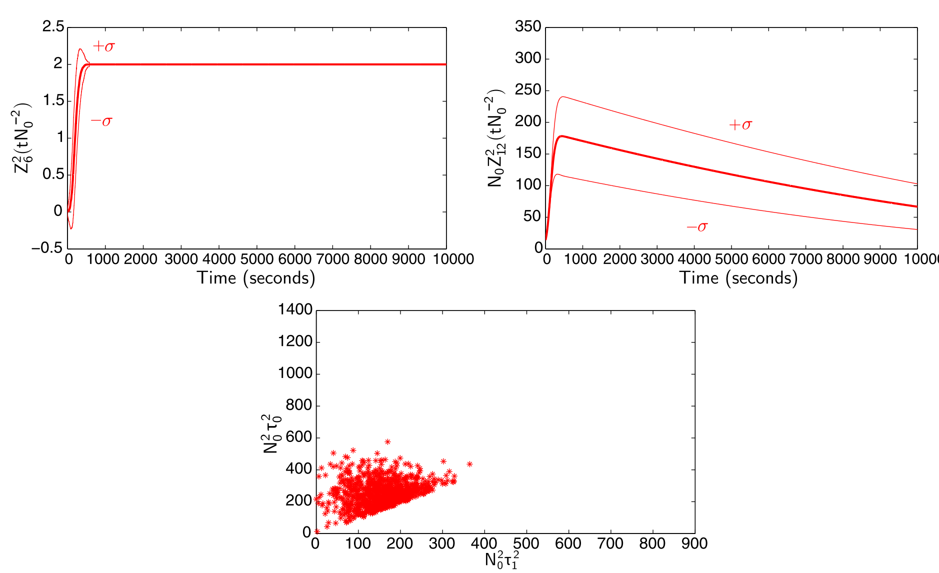

FIG. 6. *Approximation using the limiting model for $\gamma = 2$ in the alternative scaling.*

equilibrate, action may restart after the dissociation

$$DNA \cdot 2D \rightharpoonup DNA \cdot D + D.$$

If reaction 8 does not occur, the stochastic behavior of the limiting model just depends on the two jump times

$$\tau_1^2 = \inf\{t : Z_{45}^2(t) = 1\}, \qquad \tau_0^2 = \inf\{t : Z_{45}^2(t) = 0\},$$

so we compare these random variables to the corresponding variables

$$\tau_1 = \inf\{t : X_4(t) + X_5(t) = 1\}, \qquad \tau_0 = \inf\{t : X_4(t) + X_5(t) = 0\}$$

from the original model or, more precisely, because of the change of time scale, we compare $(N_0^2\tau_1^2, N_0^2\tau_0^2)$ to (τ_1, τ_0).

In Figure 5, plots for τ_1 and τ_0 for 100 exact simulations are given. Taking the average, the mean of first hitting time of $X_4(t) + X_5(t)$ to 1 is 305.44 and the mean of the first hitting time of $X_4(t) + X_5(t)$ to 0 is 512.45. In Figure 6, plots for 1000 simulations of τ_1^2 and τ_0^2 are given. The mean of the first hitting time of $Z_{45}^2(tN_0^{-2})$ to 1 is 155.95 and the mean of the first hitting time of $Z_{45}^2(tN_0^{-2})$ to 0 is 261.01. Comparing the two stopping times in the simulations of the full model and of the approximation, the mean hitting time to 1 and 0 in the approximation is much faster than in the full model. Consequently, the quicker decrease of Z_{45}^2 to 0 gives a discrepancy in the peak levels and the peak times in the full model and in the approximation.

6.4. *Derivation of Michaelis–Menten equation.* Darden (1979, 1982) derives the Michaelis–Menten equation from a stochastic reaction network model. His result can be obtained as a special case of the methods developed here.

Consider the reaction system

$$S_1 + S_2 \underset{\kappa_2'}{\overset{\kappa_1'}{\rightleftharpoons}} S_3 \overset{\kappa_3'}{\rightharpoonup} S_4 + S_2,$$

where S_1 is the substrate, S_2 the enzyme, S_3 the enzyme-substrate complex and S_4 the product. Assume that the parameters scale so that

$$\begin{aligned}
Z_1^N(t) &= Z_1^N(0) - N^{-1}Y_1\left(N\int_0^t \kappa_1 Z_1^N(s)Z_2^N(s)\,ds\right) \\
&\quad + N^{-1}Y_2\left(N\int_0^t \kappa_2 Z_3^N(s)\,ds\right), \\
Z_2^N(t) &= Z_2^N(0) - Y_1\left(N\int_0^t \kappa_1 Z_1^N(s)Z_2^N(s)\,ds\right) \\
&\quad + Y_2\left(N\int_0^t \kappa_2 Z_3^N(s)\,ds\right) + Y_3\left(N\int_0^t \kappa_3 Z_3^N(s)\,ds\right),
\end{aligned}$$

$$Z_3^N(t) = Z_2^N(0) + Y_1\left(N\int_0^t \kappa_1 Z_1^N(s) Z_2^N(s)\,ds\right) - Y_2\left(N\int_0^t \kappa_2 Z_3^N(s)\,ds\right) - Y_3\left(N\int_0^t \kappa_3 Z_3^N(s)\,ds\right),$$

$$Z_4^N(t) = N^{-1} Y_3\left(N\int_0^t \kappa_3 Z_3^N(s)\,ds\right),$$

that is, $\alpha_1 = \alpha_4 = 1$, $\alpha_2 = \alpha_3 = 0$, $\beta_1 = 0$, and $\beta_2 = \beta_3 = 1$.

Note that $M = Z_3^N(t) + Z_2^N(t)$ is constant, and define

$$V_2^N(t) = \int_0^t Z_2^N(s)\,ds.$$

THEOREM 6.1. *Assume that $Z_1^N(0) \to x_1(0)$. Then (Z_1^N, V_2^N) converges to $(x_1(t), v_2(t))$ satisfying*

$$\begin{aligned} x_1(t) &= x_1(0) - \int_0^t \kappa_1 x_1(s)\dot{v}_2(s)\,ds + \int_0^t \kappa_2\bigl(M - \dot{v}_2(s)\bigr)\,ds, \\ 0 &= -\int_0^t \kappa_1 x_1(s)\dot{v}_2(s)\,ds + \int_0^t (\kappa_2+\kappa_3)\bigl(M - \dot{v}_2(s)\bigr)\,ds \end{aligned} \tag{6.8}$$

and, hence, $\dot{v}_2(s) = \frac{(\kappa_2+\kappa_3)M}{\kappa_2+\kappa_3+\kappa_1 x_1(s)}$ and

$$\dot{x}_1(t) = -\frac{M\kappa_1\kappa_3 x_1(t)}{\kappa_2+\kappa_3+\kappa_1 x_1(t)}.$$

PROOF. Relative compactness of the sequence (Z_1^N, V_2^N) is straightforward. Dividing the second equation by N and passing to the limit, we see that any limit point (x_1, v_2) of (Z_1^N, V_2^N) must satisfy

$$0 = -\int_0^t \kappa_1 x_1(s)\,dv_2(s) + (\kappa_2+\kappa_3)Mt - \int_0^t (\kappa_2+\kappa_3)\,dv_2(s). \tag{6.9}$$

Since v_2 is Lipschitz, it is absolutely continuous, and rewriting (6.9) in terms of the derivative gives the second equation in (6.8). The first equation follows by a similar argument. □

6.5. *Limiting models when the balance conditions fail.* The balance condition, Condition 3.2, has as its goal ensuring that the normalized species numbers remain positive, at least on average, and bounded. Even if the balance condition fails, it may still be possible to obtain a limiting model in which one or more of the species abundances are driven to zero and completely disappear from the limiting model. A referee suggested the following simple example:

$$\varnothing \overset{\kappa_1'}{\rightharpoonup} S_1 \overset{\kappa_2'}{\rightharpoonup} S_2 \overset{\kappa_3'}{\rightharpoonup} S_3 \overset{\kappa_4'}{\rightharpoonup} \varnothing$$

under the assumption that $\kappa_3' \gg \kappa_1', \kappa_2', \kappa_4'$. Clearly, the natural reduced model should be

$$\varnothing \overset{\kappa_1'}{\rightharpoonup} S_1 \overset{\kappa_2'}{\rightharpoonup} S_3 \overset{\kappa_4'}{\rightharpoonup} \varnothing.$$

Taking the $\alpha_i = 0$, $\beta_1 = \beta_2 = \beta_4 = 0$, and $\beta_3 = 1$, the scaled system becomes

$$\begin{aligned}
Z_1^N(t) &= Z_1(0) + Y_1(\kappa_1 t) - Y_2\Bigl(\int_0^t \kappa_2 Z_1^N(s)\,ds\Bigr),\\
Z_2^N(t) &= Z_2(0) + Y_2\Bigl(\int_0^t \kappa_2 Z_1^N(s)\,ds\Bigr) - Y_3\Bigl(\int_0^t \kappa_3 N Z_2^N(s)\,ds\Bigr),\\
Z_3^N(t) &= Z_3(0) + Y_3\Bigl(\int_0^t \kappa_3 N Z_2^N(s)\,ds\Bigr) - Y_4\Bigl(\int_0^t \kappa_4 Z_3^N(s)\,ds\Bigr).
\end{aligned}$$

Clearly, the species balance condition fails for both species 2 and species 3. Dividing the second equation by N and passing to the limit, it follows easily that for each $T > 0$, the Lebesgue measure of the set $\{t \le T : Z_2^N(t) > 0\}$ converges to zero. Consequently, the Lebesgue measure of the set of $t \le T$ such that

$$Z_3^N(t) \ne Z_3(0) + Z_2(0) + Y_2\Bigl(\int_0^t \kappa_2 Z_1^N(s)\,ds\Bigr) - Y_4\Bigl(\int_0^t \kappa_4 Z_3^N(s)\,ds\Bigr)$$

goes to zero, and (Z_1^N, Z_3^N) converges to the solution of

$$\begin{aligned}
Z_1(t) &= Z_1(0) + Y_1(\kappa_1 t) - Y_2\Bigl(\int_0^t \kappa_2 Z_1(s)\,ds\Bigr),\\
Z_3(t) &= Z_3(0) + Z_2(0) + Y_2\Bigl(\int_0^t \kappa_2 Z_1(s)\,ds\Bigr)\\
&\quad - Y_4\Bigl(\int_0^t \kappa_4 Z_3(s)\,ds\Bigr).
\end{aligned}$$

Note that the sequence does *not* converge in the Skorohod topology on $D_{\mathbb{R}^2}[0,\infty)$ (distinct discontinuities of Z_1^N and Z_2^N coalesce in the limit), but it does converge in $D_{\mathbb{R}}[0,\infty) \times D_{\mathbb{R}}[0,\infty)$ and the finite-dimensional distributions of (Z_1^N, Z_3^N) converge to the finite-dimensional distributions of (Z_1, Z_3).

Mastny, Haseltine and Rawlings (2007) consider a more complex example in which the balance conditions fail,

$$S_1 \underset{\kappa_2'}{\overset{\kappa_1'}{\rightleftharpoons}} 2S_2, \qquad S_2 \overset{\kappa_3'}{\rightharpoonup} S_3,$$

where we assume $\kappa_2', \kappa_3' \gg \kappa_1'$. Take the scaled system to be

$$Z_1^N(t) = Z_1(0) - Y_1\Bigl(\int_0^t \kappa_1 Z_1^N(s)\,ds\Bigr) + Y_2\Bigl(N\int_0^t \kappa_2 Z_2^N(s)\bigl(Z_2^N(s)-1\bigr)\,ds\Bigr),$$

$$
\begin{aligned}
Z_2^N(t) &= Z_2(0) + 2Y_1\left(\int_0^t \kappa_1 Z_1^N(s)\,ds\right) \\
&\qquad - 2Y_2\left(N\int_0^t \kappa_2 Z_2^N(s)\big(Z_2^N(s)-1\big)\,ds\right) - Y_3\left(N\int_0^t \kappa_3 Z_2^N(s)\,ds\right), \\
Z_3^N(t) &= Z_3(0) + Y_3\left(N\int_0^t \kappa_3 Z_2^N(s)\,ds\right).
\end{aligned}
$$

Consequently, assuming $Z_2(0)=0$, for most $t>0$, $Z_2^N(t)=0$ and

$$
\begin{aligned}
2Y_1\left(\int_0^t \kappa_1 Z_1^N(s)\,ds\right) &= Y_3\left(N\int_0^t \kappa_3 Z_2^N(s)\,ds\right) \\
&\qquad + 2Y_2\left(N\int_0^t \kappa_2 Z_2^N(s)\big(Z_2^N(s)-1\big)\,ds\right).
\end{aligned}
$$

To be precise, letting Λ denote the Lebesgue measure and defining

$$
\widehat{R}_2^N(t) = \int_0^t \mathbf{1}_{\{Z_2^N(r-)=2\}}\,dR_2^N(r), \qquad \widehat{R}_3^N(t) = \int_0^t \mathbf{1}_{\{Z_2^N(r-)=2\}}\,dR_3^N(r)
$$

for each $t>0$,

$$
\begin{gathered}
\lim_{N\to\infty} \Lambda\{0\le s\le t : Z_2^N(s)\neq 0\} \le \lim_{N\to\infty}\int_0^t Z_2^N(s)\,ds = 0, \\
\limsup_{N\to\infty}\sup_{s\le t} Z_2^N(s) \le 2, \\
\lim_{N\to\infty}\int_0^t |R_2^N(s)-\widehat{R}_2^N(s)|\,ds = 0, \\
\lim_{N\to\infty}\int_0^t |R_3^N(s)-2\widehat{R}_3^N(s)|\,ds = 0,
\end{gathered}
$$

so

$$
\lim_{N\to\infty}\int_0^t |R_1^N(s)-\widehat{R}_2^N(s)-\widehat{R}_3^N(s)|\,ds = 0.
$$

Setting $Q^N(t)=\mathbf{1}_{\{Z_2^N(t)=2\}}$,

$$
\widehat{R}_2^N(t) - \int_0^t NQ^N(s)\kappa_2 2\,ds \quad\text{and}\quad \widehat{R}_3^N(t) - \int_0^t NQ^N(s)\kappa_3 2\,ds
$$

are martingales. Working first with a subsequence satisfying (A.7), by Lemma A.13, $(\widehat{R}_2^N, \widehat{R}_3^N)$ converges to counting processes $(\widehat{R}_2, \widehat{R}_3)$ with intensities

$$
\widehat{\lambda}_2(t) = \frac{\kappa_1\kappa_2}{\kappa_2+\kappa_3}Z_1(t), \qquad \widehat{\lambda}_3(t) = \frac{\kappa_1\kappa_3}{\kappa_2+\kappa_3}Z_1(t),
$$

where $Z_1(t) = Z_1(0) - \widehat{R}_3(t)$. It follows that the finite-dimensional distributions of (Z_1^N, Z_3^N) converge to those of a solution to

$$Z_1(t) = Z_1(0) - Y\left(\int_0^t \frac{\kappa_1\kappa_3}{\kappa_2+\kappa_3} Z_1(s)\,ds\right),$$

$$Z_3(t) = Z_3(0) + 2Y\left(\int_0^t \frac{\kappa_1\kappa_3}{\kappa_2+\kappa_3} Z_1(s)\,ds\right),$$

which is the reduced model obtained in Mastny, Haseltine and Rawlings (2007).

In this example, Z_1^N does not converge in the Skorohod topology, but (Z_1^N, Z_3^N) does converge in the Jakubowski topology as described in Remark A.14.

[Note the relationship between our rate constants and those of Mastny, Haseltine and Rawlings (2007): $\kappa_1 = k_1$, $\kappa_2 = \frac{1}{2}k_{-1}$ and $\kappa_3 = k_2$.]

APPENDIX

A.1. Convergence of random measures. The material in this section is taken from Kurtz (1992). Proofs of the results can be found there.

Let $(\mathbb{L}, d)$ be a complete, separable metric space, and let $\mathcal{M}(\mathbb{L})$ be the space of finite measures on $\mathbb{L}$ with the weak topology. The Prohorov metric on $\mathcal{M}(\mathbb{L})$ is defined by

$$\rho(\mu,\nu) = \inf\{\varepsilon > 0 : \mu(B) \le \nu(B^\varepsilon) + \varepsilon, \nu(B) \le \mu(B^\varepsilon) + \varepsilon, B \in \mathcal{B}(\mathbb{L})\}, \tag{A.1}$$

where $B^\varepsilon = \{x \in \mathbb{L} : \inf_{y\in B} d(x,y) < \varepsilon\}$. The following lemma is a simple consequence of Prohorov's theorem.

LEMMA A.1. *Let $\{\Gamma_n\}$ be a sequence of $\mathcal{M}(\mathbb{L})$-valued random variables. Then Γ_n is relatively compact if and only if $\{\Gamma_n(\mathbb{L})\}$ is relatively compact as a family of $\mathbb{R}$-valued random variables and for each $\varepsilon > 0$, there exists a compact $K \subset \mathbb{L}$ such that* $\sup_n P\{\Gamma_n(K^c) > \varepsilon\} < \varepsilon$.

COROLLARY A.2. *Let $\{\Gamma_n\}$ be a sequence of $\mathcal{M}(\mathbb{L})$-valued random variables. Suppose that $\sup_n E[\Gamma_n(\mathbb{L})] < \infty$ and that for each $\varepsilon > 0$, there exists a compact $K \subset \mathbb{L}$ such that*

$$\limsup_{n\to\infty} E[\Gamma_n(K^c)] \le \varepsilon.$$

Then $\{\Gamma_n\}$ is relatively compact.

Let $\mathcal{L}(\mathbb{L})$ be the space of measures on $\mathbb{L}\times[0,\infty)$ such that $\mu(\mathbb{L}\times[0,t]) < \infty$ for each $t > 0$, and let $\mathcal{L}_m(\mathbb{L}) \subset \mathcal{L}(\mathbb{L})$ be the subspace on which $\mu(\mathbb{L}\times[0,t]) = t$. For $\mu \in \mathcal{L}(\mathbb{L})$, let μ^t denote the restriction of μ to $\mathbb{L}\times[0,t]$. Let ρ_t denote the Prohorov metric on $\mathcal{M}(\mathbb{L}\times[0,t])$, and define $\widehat{\rho}$ on $\mathcal{L}(\mathbb{L})$ by

$$\widehat{\rho}(\mu,\nu) = \int_0^\infty e^{-t} 1 \wedge \rho_t(\mu^t, \nu^t)\,dt,$$

that is, $\{\mu_n\}$ converges in $\widehat{\rho}$ if and only if $\{\mu_n^t\}$ converges weakly for almost every t. In particular, if $\widehat{\rho}(\mu_n, \mu) \to 0$, then $\rho_t(\mu_n^t, \mu^t) \to 0$ if and only if $\mu_n(\mathbb{L} \times [0, t]) \to \mu(\mathbb{L} \times [0, t])$. The following lemma is an immediate consequence of Lemma A.1.

LEMMA A.3. *A sequence of* $(\mathcal{L}_m(\mathbb{L}), \widehat{\rho})$*-valued random variables* $\{\Gamma_n\}$ *is relatively compact if and only if for each* $\varepsilon > 0$ *and each* $t > 0$, *there exists a compact* $K \subset \mathbb{L}$ *such that* $\inf_n E[\Gamma_n(K \times [0, t])] \geq (1 - \varepsilon)t$.

LEMMA A.4. *Let* Γ *be an* $(\mathcal{L}(\mathbb{L}), \widehat{\rho})$*-valued random variable adapted to a complete filtration* $\{\mathcal{F}_t\}$ *in the sense that for each* $t \geq 0$ *and* $H \in \mathcal{B}(\mathbb{L})$, $\Gamma(H \times [0, t])$ *is* $\mathcal{F}_t$*-measurable. Let* $\lambda(G) = \Gamma(\mathbb{L} \times G)$. *Then there exists an* $\{\mathcal{F}_t\}$*-optional*, $\mathcal{P}(\mathbb{L})$*-valued process* γ *such that*

$$\int_{\mathbb{L} \times [0,t]} h(y, s)\Gamma(dy \times ds) = \int_0^t \int_{\mathbb{L}} h(y, s)\gamma_s(dy)\lambda(ds) \tag{A.2}$$

for all $h \in B(\mathbb{L} \times [0, \infty))$ *with probability one. If* $\lambda([0, t])$ *is continuous, then* γ *can be taken to be* $\{\mathcal{F}_t\}$*-predictable.*

LEMMA A.5. *Let* $\{(x_n, \mu_n)\} \subset D_{\mathbb{E}}[0, \infty) \times \mathcal{L}(\mathbb{L})$, *and* $(x_n, \mu_n) \to (x, \mu)$. *Let* $h \in C(\mathbb{E} \times \mathbb{L})$ *and* ψ *be a nonnegative function on* $[0, \infty)$ *satisfying* $\lim_{r\to\infty} \psi(r)/r = \infty$ *such that*

$$\sup_n \int_{\mathbb{L} \times [0,t]} \psi(|h(x_n(s), y)|)\mu_n(dy \times ds) < \infty \tag{A.3}$$

for each $t > 0$.

Define

$$u_n(t) = \int_{\mathbb{L} \times [0,t]} h(x_n(s), y)\mu_n(dy \times ds),$$
$$u(t) = \int_{\mathbb{L} \times [0,t]} h(x(s), y)\mu(dy \times ds),$$

$z_n(t) = \mu_n(\mathbb{L} \times [0, t])$ *and* $z(t) = \mu(\mathbb{L} \times [0, t])$.

(a) *If* x *is continuous on* $[0, t]$ *and* $\lim_{n\to\infty} z_n(t) = z(t)$, *then* $\lim_{n\to\infty} u_n(t) = u(t)$.

(b) *If* $(x_n, z_n, \mu_n) \to (x, z, \mu)$ *in* $D_{\mathbb{E}\times\mathbb{R}}[0, \infty) \times \mathcal{L}(\mathbb{L})$, *then* $(x_n, z_n, u_n, \mu_n) \to (x, z, u, \mu)$ *in* $D_{\mathbb{E}\times\mathbb{R}\times\mathbb{R}}[0, \infty) \times \mathcal{L}(\mathbb{L})$. *In particular*, $\lim_{n\to\infty} u_n(t) = u(t)$ *at all points of continuity of* z.

(c) *The continuity assumption on* h *can be replaced by the assumption that* h *is continuous a.e.* ν_t *for each* t, *where* $\nu_t \in \mathcal{M}(\mathbb{E} \times \mathbb{L})$ *is the measure determined by* $\nu_t(A \times B) = \mu\{(y, s) : x(s) \in A, s \leq t, y \in B\}$.

Lemma A.5 and the continuous mapping theorem give the following.

LEMMA A.6. *Suppose $(Z^N, V^N) \Rightarrow (Z, V)$ in $D_{\mathbb{E}}[0,\infty) \times \mathcal{L}_m(\mathbb{L})$. Let $h \in C(\mathbb{E} \times \mathbb{L})$ and ψ be as in Lemma A.5. If $\{\int_0^t \psi(|h(Z^N(s), y)|) V^N(dy \times ds)\}$ is stochastically bounded for all $t > 0$, then*

$$\int_{\mathbb{L}\times[0,\cdot]} h(Z^N(s), y) V^N(dy \times ds) \Rightarrow \int_{\mathbb{L}\times[0,\cdot]} h(Z(s), y) V(dy \times ds).$$

A.2. Martingale properties of counting processes. A cadlag stochastic process R is a *counting process* if $R(0) = 0$ and R is constant except for jumps of plus one. If R is adapted to a filtration $\{\mathcal{F}_t\}$, then a nonnegative $\{\mathcal{F}_t\}$-adapted process λ is an $\{\mathcal{F}_t\}$-intensity for R if

$$M(t) = R(t) - \int_0^t \lambda(s)\,ds$$

is an $\{\mathcal{F}_t\}$-local martingale. Specifically, letting τ_l denote the lth jump time of R,

$$M^{\tau_l}(t) \equiv M(t \wedge \tau_l) = R(t \wedge \tau_l) - \int_0^{t\wedge\tau_l} \lambda(s)\,ds$$

is an $\{\mathcal{F}_t\}$-martingale for each l.

For simplicity, we assume that λ is cadlag.

REMARK A.7. For R_k defined in (2.1) and $\{\mathcal{F}_t\} = \sigma(R_l(s) : s \le t, l = 1, \dots, r_0)$, the intensity for R_k is $t \to \lambda_k(X(t))$.

LEMMA A.8. *For each $t \ge 0$ and each l,*

$$l \ge E[R(t \wedge \tau_l)] = E\left[\int_0^{t\wedge\tau_l} \lambda(s)\,ds\right] \tag{A.4}$$

and

$$E[R(t)] = E\left[\int_0^t \lambda(s)\,ds\right],$$

where we allow $\infty = \infty$. If $E[R(t)] < \infty$ for all $t > 0$, then

$$R(t) - \int_0^t \lambda(s)\,ds$$

is an $\{\mathcal{F}_t\}$-martingale.

Two counting processes, R_1, R_2, are *orthogonal* if they have no simultaneous jumps.

LEMMA A.9. *Let $R_1, \dots, R_m$ be pairwise orthogonal $\{\mathcal{F}_t\}$-adapted counting processes with $\{\mathcal{F}_t\}$-intensities λ_k. Then, perhaps on a larger probability space, there exist independent unit Poisson processes $Y_1, \dots, Y_m$ such that*

$$R_k(t) = Y_k\left(\int_0^t \lambda_k(s)\,ds\right),$$

and $R = \sum_{k=1}^m R_k$ *is a counting process with intensity* $\lambda = \sum_{k=1}^m \lambda_k$.

If τ_l *is the* l*th jump time of* R, *then*

$$P\{R_k(\tau_l) - R_k(\tau_l-) = 1|\mathcal{F}_{\tau_l}\} = \frac{\lambda_k(\tau_l-)}{\lambda(\tau_l-)}. \tag{A.5}$$

REMARK A.10. Note that the right-hand side of (A.5) involves the left limits of the intensities. If the intensities are not cadlag, then $\lambda_k(\tau_l-)$ should be replaced by

$$\limsup_{h\to 0+} h^{-1} \int_{\tau_l - h}^{\tau_l} \lambda_k(s)\, ds.$$

The intensity of a counting process does not necessarily uniquely determined its distribution. For example, consider the system

$$R_1(t) = Y_1\biggl(\int_0^t \lambda(R_1(s))\, ds\biggr),$$

$$R_2(t) = Y_2\biggl(\int_0^t \lambda(R_1(s))\, ds\biggr).$$

The intensity for each component is $\lambda(R_1(t))$, but the two components will not have the same distribution.

PROOF OF LEMMA A.9. See Meyer (1971) and Kurtz (1980). □

LEMMA A.11. *Suppose that* $R_1^N, \dots, R_m^N$ *are pairwise orthogonal counting processes adapted to a filtration* $\{\mathcal{F}_t^N\}$ *with* $\{\mathcal{F}_t^N\}$*-intensities* $\lambda_1^N, \dots, \lambda_m^N$. *Let* $\Lambda_k^N(t) = \int_0^t \lambda_k^N(s)\, ds$, *and suppose that* $(\Lambda_1^N, \dots, \Lambda_m^N) \Rightarrow (\Lambda_1, \dots, \Lambda_m)$ *in the Skorohod topology on* $D_{\mathbb{R}^m}[0,\infty)$. *Then* $\{(R_1^N, \dots, R_m^N)\}$ *is relatively compact in the Skorohod topology and any limit point* $(R_1, \dots, R_m)$ *consists of pairwise orthogonal counting processes.*

At least along a further subsequence,

$$(\Lambda_1^N, \dots, \Lambda_m^N, R_1^N, \dots, R_m^N) \Rightarrow (\Lambda_1, \dots, \Lambda_m, R_1, \dots, R_m),$$

and letting $\{\mathcal{F}_t^{\Lambda,R}\}$ *be the filtration generated by* $(\Lambda_1, \dots, \Lambda_m, R_1, \dots, R_m)$, $R_k - \Lambda_k$ *are* $\{\mathcal{F}_t^{\Lambda,R}\}$*-local martingales and there exist independent unit Poisson processes* $(Y_1, \dots, Y_m)$ *such that*

$$R_k(t) = Y_k(\Lambda_k(t)), \qquad k = 1, \dots, m. \tag{A.6}$$

REMARK A.12. If the Λ_k are adapted to $\{\mathcal{F}_t^R\}$, then R will be the unique solution of (A.6) and $R^N \Rightarrow R$ in the Skorohod topology.

PROOF OF LEMMA A.11. See Kabanov, Liptser and Shiryaev (1984). $\square$

In Section 6.5, we consider an example for which the integrated intensities did not have a continuous limit. The next lemma covers that situation.

LEMMA A.13. *Suppose that $R_0^N, R_1^N, \dots, R_m^N$ are counting processes adapted to a filtration $\{\mathcal{F}_t^N\}$, and $R_1^N, \dots, R_m^N$ are pairwise orthogonal. Suppose R_0^N has $\{\mathcal{F}_t^N\}$-intensity λ_0^N, and $R_1^N, \dots, R_m^N$ have $\{\mathcal{F}_t^N\}$-intensities $\lambda_k^N = NQ^N\mu_k^N$, where $Q^N \geq 0$. Suppose*

$$(\lambda_0^N, \mu_1^N, \dots, \mu_m^N) \Rightarrow (\lambda_0, \mu_1, \dots, \mu_m) \tag{A.7}$$

and

$$\int_0^t \left| R_0^N(s) - \sum_{k=1}^m R_k^N(s) \right| ds \to 0 \tag{A.8}$$

for each $t > 0$. Then $\{(R_0^N, R_1^N, \dots, R_m^N)\}$ is relatively compact in the Jakubowski topology and for any limit point $(R_0, R_1, \dots, R_m)$,

$$R_0 = \sum_{k=1}^m R_k,$$

and $R_1, \dots, R_m$ are pairwise orthogonal counting processes with intensities

$$\lambda_k(t) = \frac{\mu_k(t)}{\sum_{l=1}^m \mu_l(t)} \lambda_0(t).$$

REMARK A.14. The sequence may not be relatively compact in the Skorohod topology since we have not ruled out the possibility that the sequence has discontinuities that coalesce. See the example in Section 6.5.

The Meyer–Zheng conditions [Meyer and Zheng (1984)] imply relative compactness in the Jakubowski topology [Jakubowski (1997)]. A sequence of cadlag functions $\{x_n\}$ converges to a cadlag function x in the Jakubowski topology if and only if there exists a sequence of time changes $\{\gamma_n\}$ such that $(x_n \circ \gamma_n, \gamma_n) \to (x \circ \gamma, \gamma)$ in the Skorohod topology. [See Kurtz (1991).] The time-changes are continuous, nondecreasing mappings from $[0, \infty)$ onto $[0, \infty)$ but are not necessarily strictly increasing. Convergence implies $\int_0^t |x_n(s) - x(s)| \wedge 1\, ds \to 0$. In contrast to the Skorohod topology, if $x_n \to x$ and $y_n \to y$ in the Jakubowski topology, then $(x_n, y_n) \to (x, y)$ in the Jakubowski topology on cadlag functions in the product space.

PROOF OF LEMMA A.13. By Lemma A.11, $\{R_0^N\}$ is relatively compact in the Skorohod topology and hence in the Jakubowski topology. Let

$$\widehat{R}_0^N = \sum_{k=1}^m R_k^N.$$

The stochastic boundedness of $\{R_0^N(t)\}$ for each $t > 0$ and (A.8) imply the stochastic boundedness of $\{\widehat{R}_0^N(t)\}$ for each $t > 0$ which by (A.4) implies the stochastic boundedness of

$$\left\{\int_0^t NQ^N(s)\sum_{k=1}^m \mu_k^N(s)\,ds\right\}.$$

Let γ_N be defined by

$$\int_0^{\gamma_N(t)}\left(1 + NQ^N(s)\sum_{k=1}^m \mu_k^N(s)\right)ds = t.$$

Since $|\gamma_N(s) - \gamma_N(t)| \le |s-t|$, $\{\gamma_N\}$ is relatively compact. Define

$$\Lambda_k^N(t) = \int_0^t \lambda_k^N(s)\,ds$$

and observe that

$$\Lambda_l^N \circ \gamma_N(t) = \int_0^t \frac{NQ^N \circ \gamma_N(s)\mu_l^N \circ \gamma_N(s)}{1 + NQ^N \circ \gamma_N(s)\sum_k \mu_k^N \circ \gamma_N(s)}\,ds$$

is also Lipschitz with Lipschitz constant 1. Since $\{\gamma_N(t), t \ge 0\}$ are stopping times,

$$R_l^N - \Lambda_l^N \circ \gamma_N$$

are martingales with respect to the filtration $\{\mathcal{F}^N_{\gamma_N(t)}\}$.

The Lipschitz properties imply the relative compactness of

$$\{(\Lambda_1^N \circ \gamma_N, \ldots, \Lambda_m^N \circ \gamma_N, \gamma_N)\}$$

in the Skorohod topology, which, in turn, by Lemma A.11, implies the relative compactness of

$$\{(\Lambda_1^N \circ \gamma_N, \ldots, \Lambda_m^N \circ \gamma_N, \gamma_N, R_1^N \circ \gamma_N, \ldots, R_m^N \circ \gamma_N)\}.$$

Relative compactness of this sequence in the Skorohod topology ensures relative compactness of $\{(R_1^N, \ldots, R_m^N)\}$ in the Jakubowski topology, which, in turn, implies relative compactness of $\{(R_0^N, R_1^N, \ldots, R_m^N)\}$ in the Jakubowski topology.

Along an appropriate subsequence, we have convergence of γ_N to a limit γ,

$$\int_0^t \frac{NQ^N \circ \gamma_N(s)\sum_k \mu_k^N \circ \gamma_N(s)}{1 + NQ^N \circ \gamma_N(s)\sum_k \mu_k^N \circ \gamma_N(s)}\,ds \Rightarrow \widehat{\Lambda},$$

convergence of $\Lambda_k^N \circ \gamma_N$ to

$$\widehat{\Lambda}_k(t) = \int_0^t \frac{\mu_k \circ \gamma(s)}{\sum_l \mu_l \circ \gamma(s)}\,d\widehat{\Lambda}(ds),$$

and convergence of $(R_0^N, R_1^N, \dots, R_m^N)$ in the Jakubowski topology to a process satisfying

$$R_0 = \sum_{k=1}^{m} R_k.$$

Since $R_0 \circ \gamma(t) - \int_0^{\gamma(t)} \lambda_0(s)\,ds$ is a martingale, we must have

$$\int_0^{\gamma(t)} \lambda_0(s)\,ds = \widehat{\Lambda}(t)$$

and

$$\widehat{\Lambda}_k(t) = \int_0^t \frac{\mu_k \circ \gamma(s)}{\sum_l \mu_l \circ \gamma(s)} \lambda_0 \circ \gamma(s)\gamma'(s)\,ds.$$

Since R_0 is a counting process, the R_k, $k = 1, \dots, m$, must be orthogonal, and R_k must have intensity $\frac{\mu_k}{\sum_l \mu_l}\lambda_0$. $\square$

Acknowledgments. The authors thank the other members of the NSF sponsored Focused Research Group on Intracellular Reaction Networks, David Anderson, George Craciun, Lea Popovic, Greg Rempala and John Yin, for many helpful conversations during the long gestation period of the ideas presented here. They provided many valuable insights and much encouragement. This work was completed while the first author held a postdoctoral appointment under Hans Othmer at the University of Minnesota and the second author was a Visiting Fellow at the Isaac Newton Institute in Cambridge, UK. The hospitality and support provided by these appointments is gratefully acknowledged.

REFERENCES

BALL, K., KURTZ, T. G., POPOVIC, L. and REMPALA, G. (2006). Asymptotic analysis of multiscale approximations to reaction networks. *Ann. Appl. Probab.* **16** 1925–1961. MR2288709

CAO, Y., GILLESPIE, D. T. and PETZOLD, L. R. (2005). The slow-scale stochastic simulation algorithm. *J. Chem. Phys.* **122** 014116. Available at http://link.aip.org/link/?JCP/122/014116/1.

CRUDU, A., DEBUSSCHE, A. and RADULESCU, O. (2009). Hybrid stochastic simplifications for multiscale gene networks. *BMC Syst. Biol.* **3** 89.

DARDEN, T. (1979). A pseudo-steady state approximation for stochastic chemical kinetics. *Rocky Mountain J. Math.* **9** 51–71. MR0517973

DARDEN, T. A. (1982). Enzyme kinetics: Stochastic vs. deterministic models. In *Instabilities, Bifurcations, and Fluctuations in Chemical Systems* (*Austin, Tex.*, 1980) 248–272. Univ. Texas Press, Austin, TX. MR0774792

DAVIS, M. H. A. (1993). *Markov Models and Optimization. Monographs on Statistics and Applied Probability* **49**. Chapman & Hall, London. MR1283589

E, W., LIU, D. and VANDEN-EIJNDEN, E. (2005). Nested stochastic simulation algorithm for chemical kinetic systems with disparate rates. *J. Chem. Phys.* **123** 194107. Available at http://link.aip.org/link/?JCP/123/194107/1.

E, W., LIU, D. and VANDEN-EIJNDEN, E. (2007). Nested stochastic simulation algorithms for chemical kinetic systems with multiple time scales. *J. Comput. Phys.* **221** 158–180. MR2290567

ETHIER, S. N. and KURTZ, T. G. (1986). *Markov Processes*: *Characterization and Convergence*. Wiley, New York. MR0838085

GILLESPIE, D. T. (1977). Exact stochastic simulation of coupled chemical reactions. *J. Phys. Chem.* **81** 2340–2361.

GOUTSIAS, J. (2005). Quasiequilibrium approximation of fast reaction kinetics in stochastic biochemical systems. *J. Chem. Phys.* **122** 184102. Available at http://link.aip.org/link/?JCP/122/184102/1.

HASELTINE, E. L. and RAWLINGS, J. B. (2002). Approximate simulation of coupled fast and slow reactions for stochastic chemical kinetics. *J. Chem. Phys.* **117** 6959–6969.

HENSEL, S. C., RAWLINGS, J. B. and YIN, J. (2009). Stochastic kinetic modeling of vesicular stomatitis virus intracellular growth. *Bull. Math. Biol.* **71** 1671–1692. MR2544630

JAKUBOWSKI, A. (1997). A non-Skorohod topology on the Skorohod space. *Electron. J. Probab.* **2** 21 pp. (electronic). MR1475862

KABANOV, Y. M., LIPTSER, R. S. and SHIRYAEV, A. N. (1984). Weak and strong convergence of distributions of counting processes. *Theory Probab. Appl.* **28** 303–336.

KANG, H.-W. (2011). A multiscale approximation in a heat shock response model of E. coli. Unpublished manuscript. Mathematical Biosciences Institute, Ohio State Univ.

KANG, H.-W., KURTZ, T. G. and POPOVIC, L. (2012). Diffusion approximations for multiscale chemical reaction models. Unpublished manuscript. School of Mathematics, Univ. Minnesota, Dept. Mathematics and Statistics, Univ. Wisconsin, Madison and Dept. Mathematics and Statistics, Concordia Univ.

KHAS'MINSKIĬ, R. Z. (1966a). On stochastic processes defined by differential equations with a small parameter. *Theory Probab. Appl.* **11** 211–228.

KHAS'MINSKIĬ, R. Z. (1966b). A limit theorem for the solutions of differential equations with random right-hand sides. *Theory Probab. Appl.* **11** 390–406.

KURTZ, T. G. (1972). The relationship between stochastic and deterministic models for chemical reactions. *J. Chem. Phys.* **57** 2976–2978.

KURTZ, T. G. (1977/78). Strong approximation theorems for density dependent Markov chains. *Stochastic Process. Appl.* **6** 223–240. MR0464414

KURTZ, T. G. (1980). Representations of Markov processes as multiparameter time changes. *Ann. Probab.* **8** 682–715. MR0577310

KURTZ, T. G. (1991). Random time changes and convergence in distribution under the Meyer–Zheng conditions. *Ann. Probab.* **19** 1010–1034. MR1112405

KURTZ, T. G. (1992). Averaging for martingale problems and stochastic approximation. In *Applied Stochastic Analysis* (*New Brunswick*, *NJ*, 1991). *Lecture Notes in Control and Information Sciences* **177** 186–209. Springer, Berlin. MR1169928

MACNAMARA, S., BURRAGE, K. and SIDJE, R. B. (2007). Multiscale modeling of chemical kinetics via the master equation. *Multiscale Model. Simul.* **6** 1146–1168. MR2393029

MASTNY, E. A., HASELTINE, E. L. and RAWLINGS, J. B. (2007). Two classes of quasi-steady-state model reductions for stochastic kinetics. *J. Chem. Phys.* **127** 094106. Available at http://link.aip.org/link/?JCP/127/094106/1.

MEYER, P. A. (1971). Démonstration simplifiée d'un théorème de Knight. In *Séminaire de Probabilités*, *V* (*Univ. Strasbourg*, *Année Universitaire* 1969–1970). *Lecture Notes in Math.* **191** 191–195. Springer, Berlin. MR0380972

MEYER, P. A. and ZHENG, W. A. (1984). Tightness criteria for laws of semimartingales. *Ann. Inst. Henri Poincaré Probab. Stat.* **20** 353–372. MR0771895

RAO, C. V. and ARKIN, A. P. (2003). Stochastic chemical kinetics and the quasi-steady-state assumption: Application to the Gillespie algorithm. *J. Chem. Phys.* **118** 4999–5010.

SEGEL, L. A. and SLEMROD, M. (1989). The quasi-steady-state assumption: A case study in perturbation. *SIAM Rev.* **31** 446–477. MR1012300

SRIVASTAVA, R., PETERSON, M. S. and BENTLEY, W. E. (2001). Stochastic kinetic analysis of Escherichia coli stress circuit using sigma(32)-targeted antisense. *Biotechnol*. *Bioeng*. **75** 120–129.

ZEISER, S., FRANZ, U. and LIEBSCHER, V. (2010). Autocatalytic genetic networks modeled by piecewise-deterministic Markov processes. *J*. *Math*. *Biol*. **60** 207–246. MR2552726

SCHOOL OF MATHEMATICS
UNIVERSITY OF MINNESOTA
206 CHURCH STREET S.E.
MINNEAPOLIS, MINNESOTA 55455
USA
E-MAIL: hkang@math.umn.edu
URL: http://www.math.umn.edu/~hkang/

DEPARTMENTS OF MATHEMATICS AND STATISTICS
UNIVERSITY OF WISCONSIN, MADISON
480 LINCOLN DRIVE
MADISON, WISCONSIN 53706-1388
USA
E-MAIL: kurtz@math.wisc.edu
URL: http://www.math.wisc.edu/~kurtz/