FRAGILITY IN A TOGASHI–KANEKO STOCHASTIC MODEL WITH MUTATIONS*

YI FU[†], HYE-WON KANG[‡], WASIUR R. KHUDABUKHSH[§], LEA POPOVIC[¶], GRZEGORZ A. REMPAŁA[‖], AND RUTH J. WILLIAMS[#]

Abstract. The Togashi–Kaneko (TK) stochastic model is a prototypical example of an autocatalytic reaction network exhibiting dramatic switching behavior. The desire to understand this unusual behavior has attracted considerable attention in recent years. In this paper, we study the TK model with additional mutations. We establish a rigorous stochastic averaging principle that describes slow dynamics in terms of certain ergodic means of fast variables. Beginning with two species, we demonstrate a sensitivity of the model to even slight departures from symmetry in the autocatalytic reactions. We accomplish this through a detailed analysis of the stationary distribution of the fast process when the state of the slow process is fixed. We call this high sensitivity property "fragility". We give some examples of behavior that can occur when there are more than two species. These preliminary explorations for multiple species point to a wealth of open questions for future research.

Relevance to Life Sciences. Autocatalysis or self-amplification plays a key role in many biochemical and biological processes, ranging from pattern formation and self-organization, through gene regulation and signaling cascades, to ecological interactions and evolutionary dynamics. Understanding the sensitivity of such stochastic systems to small parameter changes is important for the formulation of models from data and for drawing conclusions for real life systems. In this paper we explore the sensitivity of the prototypical Togashi–Kaneko model with additional mutations. We find a high sensitivity to even slight departures from symmetry in the autocatalytic reactions, which we call fragility. We believe that fragility is an important underappreciated and understudied phenomenon, that will affect the formulation and interpretation of autocatalytic models across a wide variety of applications in the life sciences.

Mathematical Content. We develop tools for analyzing and understanding the dynamic stochastic behavior of autocatalytic reaction systems, especially in asymmetric situations, by considering an extension of the standard TK model with additional mutation reactions. We prove a rigorous stochastic averaging principle that links the slow population dynamics with fast autocatalytic reactions. Through analysis of the ergodic mean of the fast variables (when the slow variables are frozen at a given value) for the two-species model, we find a high sensitivity of the model (which we call fragility) to even a slight departure from symmetry in autocatalytic rates. Furthermore, our preliminary explorations for more than two species suggest that such a phenomenon can occur for four species but not for three. This rather surprising observation suggests a wealth of open problems for future research.

Key words. Stochastic chemical reaction network, stochastic averaging, autocatalytic reactions, asymmetry, fragility, switching behavior

MSC codes. 92B05, 92C40, 92C42, 60J27, 37N25

[†]Department of Mathematics, University of California, San Diego, 9500 Gilman Drive, La Jolla CA 92093, USA (yif064@ucsd.edu, https://profiles.ucsd.edu/yi.fu).

[‡]Department of Mathematics and Statistics, University of Maryland, Baltimore County, 1000 Hilltop Circle, Baltimore MD 21250, USA (hwkang@umbc.edu, https://sites.google.com/umbc.edu/hwkang/).

[§] School of Mathematical Sciences, University of Nottingham, University Park, Nottingham NG7 2RD, United Kingdom (Wasiur.KhudaBukhsh@nottingham.ac.uk, https://www.wasiur.xyz/).

[¶]Department of Mathematics and Statistics, Concordia University, Montreal QC H3G 1M8, Canada (lea.popovic@concordia.ca, https://mypage.concordia.ca/mathstat/lpopovic/).

[‖]Division of Biostatistics, College of Public Health, and Department of Mathematics, The Ohio State University, 380E Cunz Hall, 1841 Neil Avenue, Columbus, OH 43210, USA (rempala.3@osu.edu, https://cph.osu.edu/people/grempala).

[#]Department of Mathematics, University of California, San Diego, 9500 Gilman Drive La Jolla CA 92093-0112, USA (rjwilliams@ucsd.edu, https://mathweb.ucsd.edu/~williams/).

1. Introduction. In biochemical and biological contexts, autocatalytic interactions—where a species promotes its own production, often through feedback with other components—play a central role. They capture nonlinear feedback loops capable of generating rich dynamical phenomena such as bistability, sustained oscillations, and stochastic switching between metastable states. These mechanisms are not only fundamental in chemical kinetics, where autocatalysis underlies classical models of pattern formation and self-organization [21, 45], but they also appear in biological settings ranging from gene regulation and signaling cascades [15, 25] to ecological interactions and evolutionary dynamics [7, 35]. They have also appeared in models of opinion dynamics in populations [31]. Accordingly, stochastic models of autocatalytic networks have attracted sustained interest across biology, chemistry, physics and social science, both as paradigms for understanding complex behavior and for exploring the interplay between noise and nonlinearity.

Recognizing these broad applications, it is important to understand how sensitive such models are to uncertainties in their parameters and inputs. In real-life scientific applications, data and parameter values are almost always subject to measurement error, noise, or uncertainty. Even small perturbations in parameters or slight deviations in data collection protocols may lead to substantially different conclusions—particularly in complex models or analyses with many tunable settings. This phenomenon has been observed, for example, in studies showing that different analysts working with the same dataset can arrive at strikingly different effect sizes depending on seemingly minor analytical choices [29]. Similarly, biological investigations reveal that efficient performance in complex systems often trades off with robustness, wherein small parameter changes can shift outcomes in significant ways [50]. Consequently, an essential component of robust scientific inference is the study of how sensitive results are to such small variations in inputs or analytic choices. By explicitly quantifying robustness (or lack thereof), one can identify critical parameters, improve the design of experiments, and provide more reliable guidance to policy, practice, or further empirical work. This general concern with robustness and sensitivity provides a natural motivation for studying this phenomena in autocatalytic systems, such as the classical Togashi–Kaneko (TK) model [48].

In its simplest two species form, the TK model describes the stochastic dynamics of the reactions

$$S_1 + S_2 \longrightarrow 2S_2, \quad S_1 + S_2 \longrightarrow 2S_1, \quad \emptyset \longrightarrow S_1 \longrightarrow \emptyset, \quad \emptyset \longrightarrow S_2 \longrightarrow \emptyset. \tag{1.1}$$

The first two reactions are autocatalytic, meaning that the presence of S_2 promotes the production of additional S_2, and likewise S_1 promotes its own replication, creating a self-amplifying feedback loop. The remaining reactions describe the production of each species from the environment and their eventual degradation, ensuring that the system operates as an open dynamical process. The model (1.1) has attracted considerable attention in recent times in the physics and biology literature [9, 48] because of its interesting switching dynamics, dubbed "discreteness induced transitions". The name refers to the fact that under certain conditions the system spends most of its time near states where one of the species is absent, switching rapidly from one such state to another. This dynamics is driven by a discrete change in species counts and is prominent when the rates of autocatalytic reactions are fast relative to the rates of production and degradation of species.

Many of the observations and analyses of the two species TK model and its extensions to higher dimensions consider symmetric autocatalytic reactions, meaning that the rate constants for autocatalysis are equal. Initial numerical observations of the mentioned switching behavior were made by Togashi and Kaneko [48, 49] for a model on four species that extends the system (1.1) by adding the following autocatalytic reactions (to form a loop), plus production and degradation reactions:

$$S_2 + S_3 \longrightarrow 2S_3,\ S_3 + S_4 \longrightarrow 2S_4,\ S_4 + S_1 \longrightarrow 2S_1,\ \emptyset \longrightarrow S_3 \longrightarrow \emptyset,\ \emptyset \longrightarrow S_4 \longrightarrow \emptyset.$$

Togashi and Kaneko assumed that all autocatalytic reactions have the same rate constants. This symmetric autocatalytic model on four species was further analytically explored by Biancalini et al. [7, 8], using a stochastic differential equation (SDE) based on volume rescaling to provide an approximation of the stationary distribution for species counts. An approximation using reflected SDEs (RSDE) of the same type of symmetric autocatalytic model with a loop of reactions on d species was provided by Fan et al. [24]. Bibbona et al. in [9] derived an exact expression for the stationary distribution for the symmetric d species TK model with a loop of autocatalytic reactions in the special case where there is a common production rate, and a common degradation rate that is equal to $d/(d-1)$ times the common autocatalytic rate constant. They also derived a stationary distribution for a generalization of the symmetric TK model on d species, in which autocatalytic reactions with the same rate constants are present among all pairs of distinct species and degradation rates are common. Their analytic results also use volume rescaling of rates to identify parameter regimes in which the stationary probability mass concentrates near the boundary states where one or more species is absent.

Beyond the symmetric case for autocatalytic rates, there are almost no studies investigating how autocatalytic asymmetries, meaning disparities between the autocatalytic rate constants, can affect the long-term behavior of the system. Saito and Kaneko in Sec. IV of [46] explore the effect of asymmetry in autocatalytic rates on the direction of the net chemical reaction flow in a two species TK model with added mutations. They approximated the stationary distribution of species counts using a heuristic model reduction, and observed that the direction of the net chemical reaction flow can be reversed as the volume parameter (rescaling reaction rates) decreases. In a more recent work [26], Gallinger and Popovic approximated the two species TK model with a version of this model where production and degradation reactions are replaced by mutations, and noted for their model that even slight departures from symmetry in the autocatalyic rates can produce qualitatively different stationary behavior. This sensitivity, which we call *model fragility*, is striking: even a very small parameter change—for example, slightly shifting from symmetric to asymmetric rate constants—can drastically alter the system's long-term dynamics. By fragility we mean lack of robustness, where minor variations in model parameters alter the qualitative structure of the stationary distribution or the dominant dynamical regime. Unlike classical phase transitions, fragility does not necessarily depend on tuning a single critical parameter or crossing a sharply defined threshold; rather, it reflects a broader non-robustness of the stochastic dynamics. This phenomenon is likely of practical importance, relevant for design of synthetic biological circuits and for understanding breakdowns in natural biochemical communication networks.

Model fragility in autocatalytic systems remains underexplored from an analytical perspective. The effect of asymmetry in production rates was explored numerically by Togashi and Kaneko in [49], noting the effect on the means of species for a range of different parameter regimes, as well as the change in their concentration as the total species count increased. Worsfold and Morris explored asymmetry in degradation rates in [54] using simulations and a Piecewise Deterministic Markov Process (PDMP) approximation of the switching in the growth and decay dynamics of the total population size due to concentrations of mass at boundary states. Saito and Kaneko, in Sec. IV of [46], introduced asymmetry in autocatalytic rates as well, observing the changes in the long term mean number of transitions between species (forming a chemical reaction flow), using a heuristic dimension reduction.

Despite these investigations, rigorous analysis of systems with autocatalytic asymmetry remains severely understudied. In this work, we develop a rigorous framework for analyzing fragility in a TK system that includes additional mutation reactions, with the aim of clarifying the mechanisms by which small parameter changes can destabilize autocatalytic dynamics. A novelty of our framework is a different parametrization of the reaction rates that separates the time scales between autocatalytic reaction dynamics and production/degradation dynamics. It avoids the usual rescaling the process in terms of the volume parameter, and hence allows us to obtain a stochastic approximation that: (a) provides the dynamics in time (rather than just the stationary distribution), and (b) is valid at small volumes (where the discreteness induced transitions are prominent). While our detailed analysis focuses on the two species case, we also illustrate the fragility phenomenon in some systems with more species. This raises further interesting questions about the differences between loop structured autocatalytic reactions and other TK-like models with more connected autocatalytic reaction networks analyzed by Bibbona et al. [9].

1.1. Our contributions. We consider a TK model with an arbitrary number of species, which incorporates additional mutation reactions and removes the assumption of symmetry in the autocatalytic rates. Fast changes in species counts in the two species case of our model connect naturally to the Moran model, a feature already noted for the model with symmetric autocatalytic rates in [8, 46] and significantly utilized in [26] for the analysis of their model with asymmetric autocatalytic rates.

Within our formulation we propose a comprehensive analytical and probabilistic approach to characterize fragility in such a TK system. In particular, we show that the separation of time scales between fast autocatalytic interactions with mutations and the slower production and degradation reactions gives rise to a stochastic averaging principle, which we prove rigorously. This averaging principle links the rapid fluctuations due to the autocatalytic reactions with mutations to the time evolution of the total population size. For the two species case, we derive explicit formulas for the stationary distribution of the fast dynamics and analyze how the mean number for each of the species varies with the population size, revealing that even slight asymmetry in the autocatalytic rates can lead to qualitatively different asymptotic behavior. We also investigate the regime where mutation rate constants are small compared with the autocatalytic rate constants, using stochastic singular perturbation techniques and diffusion approximations, clarifying the role of mutations in shaping both variability and stability of autocatalytic networks. Finally, we give some examples of behavior that can occur when there are more than two species. These preliminary explorations for multiple species point to a wealth of open questions for future research.

Our results advance the theoretical understanding of how asymmetry and small parameter perturbations influence the robustness of autocatalytic systems, with potential implications for more general biochemical networks.

1.2. Organization of the paper. The rest of the paper is organized as follows. In Section 2, our stochastic model of the TK system for d species with mutations is defined using a d-dimensional Continuous Time Markov Chain (CTMC). In Section 3, we consider a family of such models indexed by a small parameter $\varepsilon > 0$, where the autocatalytic and mutation reactions are considered to be fast (of order $1/\varepsilon$) and the production and degradation reactions are slower (of order one). The disparity in speed of reactions enables us to establish a rigorous stochastic averaging theorem (Theorem 3.1), characterizing how an average of the fast dynamics influences the degradation of the slower evolving total population size process. Steady-state means for the components of the fast process, with fixed total population size, feature prominently in the stochastic averaging theorem. In Section 4, for the case of two species, we provide an explicit formula for the stationary distribution of the fast process (with fixed total population size), derive the associated means and explore their behavior as functions of the total population size. In particular, we observe fragility — substantially different qualitative asymptotic behavior of some means (as the population size goes to infinity) in the symmetric versus the asymmetric regime. In Section 5, we further explore the behavior of these means when mutation rate constants are small compared with autocatalytic rate constants, using a singular perturbation method and a constrained Langevin (diffusion) approximation. In Section 6, we provide numerical results for some examples of our model with more than two species. Preliminary investigations suggest that fragility can occur for four species but not for three. This illustrates that there are many interesting questions to explore in the future for models with more than two species. Additional mathematical details are relegated to the appendices.

1.3. Notational conventions. Let $\mathbb{N}_0$ denote the set of non-negative integers and $\mathbb{R}$ denote the set of real numbers. Given a positive integer d, $\mathbb{N}_0^d$ (resp. $\mathbb{R}^d$) will denote the set of d-dimensional vectors whose entries are in $\mathbb{N}_0$ (resp. $\mathbb{R}$). We endow $\mathbb{N}_0$ and $\mathbb{N}_0^d$ with their discrete topologies and $\mathbb{R}$ and $\mathbb{R}^d$ with their usual Euclidean topologies. The σ-algebras on these spaces will be the usual Borel σ-algebras. Let $e_1, e_2, \ldots, e_d$ denote the standard unit basis vectors in $\mathbb{R}^d$, *i.e.*, $e_1 \coloneqq (1, 0, \ldots, , 0)$, $e_2 \coloneqq (0, 1, 0, \ldots, 0), \ldots, e_d \coloneqq (0, 0, \ldots, 1)$. We will use the notation $(i)_d$ to denote i (modulo d) — the remainder after integer division of i by d. For a real-valued function f of a real variable, we say $f(u) = O(u)$ as $u \to 0$ if $\frac{f(u)}{u}$ is bounded for all u in a sufficiently small neighborhood of zero. For a set U, we denote the indicator function of U by $1_U(\cdot)$ defined as $1_U(x) = 1$ if $x \in U$, and zero otherwise. For sets U and V in $\mathbb{R}^d$ for some $d \geq 1$, the space of V-valued continuous functions on the set U will be denoted by $C(U, V)$. The subset $C_b(U, V)$ of $C(U, V)$ consists of the bounded continuous functions from U into V. When $V = \mathbb{R}$, we simply write $C_b(U)$ for $C_b(U, \mathbb{R})$, and $C(U)$ for $C(U, \mathbb{R})$. When $U \subset \mathbb{R}$, the space of càdlàg (right continuous with finite left limits) functions from U into V is denoted by $D(U, V)$, which we will typically equip with the Skorokhod J_1-topology [10, 23, 34]. Given a measure ν on a measurable space $(W, \mathcal{W})$ and a function $f : W \mapsto \mathbb{R}$ that is integrable with respect to ν, we denote the integral of f with respect to ν by $\langle f, \nu \rangle$. Given a separable metric space E, let $M([0, \infty) \times E)$ denote the space of measures ν on $[0, \infty) \times E$ (endowed with the product Borel σ-algebra), such that $\nu([0, t] \times E) < \infty$ for all $t \geq 0$. Also define,

$$\bar{M}([0, \infty) \times E) \coloneqq \{\nu \in M([0, \infty) \times E) : \nu([0, t] \times E) = t \text{ for all } t \geq 0\}.$$

The space $M([0, \infty) \times E)$ is equipped with the topology induced by the metric

$$\rho(\nu, \nu') \coloneqq \int_0^\infty e^{-t} \left(1 \wedge \rho_t(\nu, \nu')\right) \mathrm{d}t,$$

where $\rho_t(\nu, \nu')$ is the Prokhorov metric distance between the restrictions of ν and ν' to $[0, t] \times E$.

2. Togashi–Kaneko stochastic model with mutations. In this section, we present our TK model with mutations as a Stochastic Chemical Reaction Network (SCRN). For a general introduction to SCRNs, we refer the reader to [4, 53].

Let $S_1, S_2, \ldots, S_d$ denote d chemical species. Consider the following system of chemical reactions:

$$S_i + S_{(i+1)_d} \xrightarrow{\kappa_{(i+1)_d}} 2S_{(i+1)_d}, \quad \emptyset \xrightarrow{\lambda_i} S_i \xrightarrow{\delta_i} \emptyset, \quad S_i \xrightarrow{\mu_{i,j}} S_j, \tag{2.1}$$

where $\kappa \coloneqq (\kappa_1, \kappa_2, \ldots, \kappa_d)$, $\lambda = (\lambda_1, \lambda_2, \ldots, \lambda_d)$, $\delta \coloneqq (\delta_1, \delta_2, \ldots, \delta_d)$, and $\mu \coloneqq (\mu_{i,j} : i \neq j \in \{1, 2, \ldots, d\})$ are the reaction rate constants for the model. Note that mutation reactions are not present in the original TK model. However, the stochastic averaging result we present in Section 3 takes a vastly simplified form if we include mutations since they eliminate absorbing states of a certain Markov chain that we will introduce later. We will also analyze the behavior of the system as the mutation rate constants approach zero. Note that the mutation reactions do not alter the total population size.

Let $X_1(t), X_2(t), \ldots, X_d(t)$ denote the species copy numbers of $S_1, S_2, \ldots, S_d$, respectively, at time t. Write $X(t) \coloneqq (X_1(t), X_2(t), \ldots, X_d(t))$ for $t \geq 0$. We assume the stochastic law of mass-action [4, 23] throughout this paper, and model the stochastic dynamics of the species copy numbers X using a CTMC. Following the random time change representation [4, 23], the trajectories of the components of X can be described via

$$\begin{aligned} X_i(t) = X_i(0) &- N_i^a\left(\int_0^t \kappa_{(i+1)_d} X_i(u) X_{(i+1)_d}(u) \mathrm{d}u\right) \\ &+ N_{i-1}^a\left(\int_0^t \kappa_i X_{i-1}(u) X_i(u) \mathrm{d}u\right) \\ &+ N_i^b\left(\int_0^t \lambda_i \mathrm{d}u\right) - N_i^d\left(\int_0^t \delta_i X_i(u) \mathrm{d}u\right) - \sum_{j\neq i} N_{i,j}^m\left(\int_0^t \mu_{i,j} X_i(u) \mathrm{d}u\right) \\ &+ \sum_{j \neq i} N_{j,i}^m\left(\int_0^t \mu_{j,i} X_j(u) \mathrm{d}u\right), \end{aligned} \tag{2.2}$$

for $t \geq 0$ and $i = 1, 2, \ldots, d$, where N_i^a, N_i^b, N_i^d, and $N_{i,j}^m$ for $i \neq j \in \{1, \ldots, d\}$, are independent (of each other and of $X(0)$) unit-rate Poisson processes. The superscripts a, b, d, m on the Poisson processes denote the reaction types: autocatalytic, production (birth), degradation, and mutation, respectively. Here, $\kappa_0 \coloneqq \kappa_d$, $X_0 \coloneqq X_d$, and $N_0^a \coloneqq N_d^a$. Let us define, for all $t \geq 0$,

$$Y(t) \coloneqq \|X(t)\|_1 = \sum_{i=1}^d X_i(t), \tag{2.3}$$

which keeps track of the total population size over all of the species $S_1, S_2, \ldots, S_d$ at time t. The trajectory of the process Y is described by

$$Y(t) = Y(0) + \sum_{i=1}^d N_i^b\left(\int_0^t \lambda_i \mathrm{d}u\right) - \sum_{i=1}^d N_i^d\left(\int_0^t \delta_i X_i(u) \mathrm{d}u\right), \tag{2.4}$$

which reduces to a simple birth-death process with constant birth rate and linear death rate when $\delta_i = \delta$ for all $i = 1, 2, \ldots, d$. For each $y \in \mathbb{N}_0$, define the sets

$$E_y \coloneqq \{x \in \mathbb{N}_0^d : \|x\|_1 = y\}, \quad \text{and } E \coloneqq \{(x, y) : x \in E_y, y \in \mathbb{N}_0\}. \tag{2.5}$$

The pair of processes (X, Y) satisfying (2.2) and (2.4) is itself a CTMC with infinitesimal generator Q given by

$$\begin{aligned} Qf(x,y) \coloneqq & \sum_{i=1}^{d} \kappa_{(i+1)_d} x_i x_{(i+1)_d} (f(x - e_i + e_{(i+1)_d}, y) - f(x,y)) \\ & + \sum_{i=1}^{d} \lambda_i (f(x + e_i, y+1) - f(x,y)) \\ & + \sum_{i=1}^{d} \delta_i x_i (f(x - e_i, y-1) - f(x,y)) \\ & + \sum_{i=1}^{d} \sum_{j \in \{1,2,\ldots,d\} \setminus \{i\}} \mu_{i,j} x_i (f(x - e_i + e_j, y) - f(x,y)), \end{aligned} \tag{2.6}$$

for bounded functions $f : E \mapsto \mathbb{R}$, and $x \coloneqq (x_1, x_2, \ldots, x_d)$, y such that $(x, y) \in E$.

3. Multiscale analysis. Several works use a scaling approach to describe approximately the dynamics of the TK model [48, 49, 46, 9, 24, 54]. Most of them use the volume of the system, with classical rescaling of mass-action rates by the volume to simplify the stochastic dynamics and examine how and when it leads to "discreteness induced transitions" (DITs) and concentration of the probability distribution for the species counts near the boundary states. Such dynamics is prominent in the parameter regime when the autocatalytic rates are fast relative to the production and degradation of the individual species, hence it requires having production and degradation rates be much smaller than the inverse of the size of the volume in order to separate time-scales. The drawback of such rescaling is that a good approximation requires a large volume, while the DITs are in fact observed predominantly for small volumes. It also allows one to approximate only the long time probability distribution of species counts, rather than of a stochastic process in time.

We analyze the CTMC model for the pair (X, Y) without scaling of the volume, and choose instead a scaling of the rates that directly separates the time-scale of the autocatalysis from the production and degradation by a factor of ε. This allows us to derive an approximation for the behavior of (X, Y) for all volumes and leads to a rigorous approximation for the stochastic dynamics of the total population size for all species in time. The long time distribution of individual species counts is contained within this dynamic process at each point in time.

3.1. Family of scaled models. We consider a family of models indexed by a small parameter $\varepsilon \in (0, \varepsilon_0)$ for some $\varepsilon_0 > 0$, where for each ε, $(X^\varepsilon, Y^\varepsilon)$ denotes a CTMC defined as in Section 2 with parameters δ_i, λ_i, κ_i, $\mu_{i,j}$, replaced by scaled parameters:

$$\delta_i^\varepsilon = \delta_i, \quad \lambda_i^\varepsilon = \lambda_i, \quad \kappa_i^\varepsilon = \varepsilon^{-1} \kappa_i, \quad \mu_{i,j}^\varepsilon = \varepsilon^{-1} \mu_{i,j}, \tag{3.1}$$

respectively, where $i \neq j \in \{1, 2, \ldots, d\}$, and $\delta_i, \lambda_i, \kappa_i, \mu_{i,j}$ are positive constants. As before, we write $\kappa \coloneqq (\kappa_1, \kappa_2, \ldots, \kappa_d)$, $\lambda = (\lambda_1, \lambda_2, \ldots, \lambda_d)$, $\delta \coloneqq (\delta_1, \delta_2, \ldots, \delta_d)$, and $\mu \coloneqq (\mu_{i,j} : i \neq j \in \{1, 2, \ldots, d\})$.

This scaling regime implies that the autocatalytic reactions and the mutation reactions are fast compared to the production and the degradation reactions. Consequently, the stochastic process X^ε is fast and has rapid jumps. On the other hand, the stochastic process Y^ε is slow. Intuitively, as ε becomes small, the effect of the autocatalytic and the mutation reactions can be averaged. The fast variables can be described by an ergodic CTMC when the slow variable is frozen to a particular value. Therefore, in the limit of $\varepsilon \to 0$, the fast variables can be approximated by their ergodic means, which will, of course, depend on the frozen value of the slow variable. In turn, the slow variable Y^ε will evolve according to a production-degradation process (a birth-death process in probability theory terminology) whose rates will involve the averaged quantities (the ergodic means of the fast variables). We will make this heuristic precise in the following section, where we will provide a rigorous stochastic averaging theorem, which enables us to describe a reduced model for Y^ε when ε is small.

Before presenting our stochastic averaging result, we remark that in the absence of mutation reactions, the fast process can be described by a CTMC with multiple absorbing states when the slow variable Y^ε is frozen at a given value y. The lack of a unique stationary distribution for the fast process leads to a technical challenge for identifying the limiting description of the slow process. The introduction of the mutation reactions helps us avoid this situation.

3.2. Stochastic averaging theorem. The idea of stochastic averaging dates back to the works of Khas'minskii [30, 38]. Since then, the method has been applied to a wide variety of applications such as stochastic approximation algorithms [13, 40], dynamical systems [52, 55], Partial Differential Equations (PDEs) [43], and queueing theory [5, 44]. In the context of SCRNs, early examples of the application of the stochastic averaging principle are [6, 18, 36, 37], where sufficient conditions are derived for various multiscale approximations in terms of the generators. Some recent contributions include [20, 27, 28, 41]. We will follow the approach of [39] in this paper.

In order to carry out the averaging, first note that, under our scaling regime, the generator $Q(\varepsilon)$ in (2.6) becomes

$$
\begin{aligned}
Q(\varepsilon)f(x,y) \coloneqq & \sum_{i=1}^{d} \varepsilon^{-1}\kappa_{(i+1)_d} x_i x_{(i+1)_d}(f(x-e_i+e_{(i+1)_d},y)-f(x,y)) \\
& + \sum_{i=1}^{d} \lambda_i (f(x+e_i,y+1)-f(x,y)) \\
& + \sum_{i=1}^{d} \delta_i x_i (f(x-e_i,y-1)-f(x,y)) \\
& + \sum_{i=1}^{d} \sum_{j\in\{1,2,\ldots,d\}\setminus\{i\}} \varepsilon^{-1}\mu_{i,j} x_i (f(x-e_i+e_j,y)-f(x,y)),
\end{aligned} \tag{3.2}
$$

for $x \in E_y$, $y \in \mathbb{N}_0$, and bounded functions $f : E \mapsto \mathbb{R}$. Now, for each fixed $y \in \mathbb{N}_0$, consider a CTMC on the state space E_y with infinitesimal generator given by

$$
\begin{aligned}
B_y g(x) &:= \sum_{i=1}^{d} \kappa_{(i+1)_d} x_i x_{(i+1)_d} (g(x - e_i + e_{(i+1)_d}) - g(x)) \\
&\quad + \sum_{i=1}^{d} \sum_{j \in \{1,2,\dots,d\} \setminus \{i\}} \mu_{i,j} x_i (g(x - e_i + e_j) - g(x)),
\end{aligned} \tag{3.3}
$$

for $x \in E_y$ and $g \in C_b(E_y)$. Since the CTMC is irreducible, and supported on a finite state space E_y, the Markov chain is ergodic. Let π_y denote the unique stationary distribution for this CTMC. We are now ready to state our main result.

THEOREM 3.1. *Assume $Y^\varepsilon(0) = y(0)$, a non-random non-negative integer, independent of ε. Then, as $\varepsilon \to 0$, the process Y^ε converges weakly to a stochastic process $\bar{Y}$ in the space $D([0,\infty), \mathbb{N}_0)$ equipped with the Skorokhod J_1-topology, where $\bar{Y}$ is a CTMC on the non-negative integers, $\mathbb{N}_0$, with generator G given by*

$$
Gh(y) := \left(\sum_{i=1}^{d} \lambda_i \right) (h(y+1) - h(y)) + \left(\sum_{i=1}^{d} \delta_i m_i^{\mu,\kappa}(y) \right) (h(y-1) - h(y)), \tag{3.4}
$$

for $h \in C_b(\mathbb{N}_0)$ with $\bar{Y}(0) = y(0)$, and $m_i^{\mu,\kappa}(y) := \langle x_i, \pi_y \rangle = \sum_{x \in E_y} x_i \pi_y(x)$ for $i = 1, 2, \dots, d$.

The proof for Theorem 3.1 is given in Appendix A. Note that $m_i^{\mu,\kappa}(0) = 0$ for $i = 1, 2, \dots, d$. Therefore, the limiting stochastic process $\bar{Y}$ is, in fact, a birth-death process. As alluded to at the beginning of Section 3, the effect of the autocatalytic and the mutation reactions is entirely averaged in the limit of $\varepsilon \to 0$. When the slow variable Y^ε is at level y, *i.e.*, $Y^\varepsilon(t) = y$ at some time t, the fast variables are approximated by their ergodic means $m_i^{\mu,\kappa}(y)$, which yields the generator G for the limiting process $\bar{Y}$. In the next two sections, we will analyze the stationary distribution π_y when $d = 2$.

4. Stationary distribution and means for the fast process with two species. The averaged slow process $\bar{Y}$ in (3.4) involves knowledge of the stationary distribution π_y of the fast process X^ε with $Y^\varepsilon = y$ frozen. This fast process is a CTMC on the state space E_y whose generator B_y is given in (3.3). Such a process has the same law as a system of reactions that only has autocatalytic reactions and mutations, and whose sum of all species counts is conserved to equal y. An explicit formula for its stationary distribution for general d is unknown (except in the case when $d = 2$ and autocatalytic rates are all equal, cf. [33] eq. (11)). When $d = 2$ and mutation rates satisfy $\mu_{ij} = \mu_j$, for all $i \neq j$, but the autocatalytic rates κ_j are not necessarily all equal, we will use an observation of Gallinger and Popovic [26] to relate the fast process on E_y to a process used in population genetics known as the Moran model with genic selection. For the remainder of this section and the next, we assume that $\mu_{ij} = \mu_j$, for all $i \neq j$.

The Moran model is used for describing a population of d species types undergoing mutations with parameters and genetic drift (reproduction by resampling) under genic selection. This population genetics model is different in general from the fast process for the original TK model in that its reproduction by resampling, which corresponds to autocatalytic reactions $S_i + S_j \longrightarrow 2S_j$, is assumed to occur between all pairs of species types (as in Sec. 4 model of [9]), rather than only for pairs in a loop, $j =$

$(i+1)_d$, as in the TK model. However, this is not an issue for models on two species, when the Moran model with genic selection and the fast TK process are equivalent. The stationary distribution for the population genetics model is known (as shown by Etheridge and Griffiths in [22]) to be a weighted version of the Dirichlet-Multinomial distribution on the $(d-1)$-dimensional simplex. When $d=2$ the Dirichlet-Multinomial is also known as the Beta-Binomial(y,α) distribution, and, by (14) in [22], we have that the stationary distribution π_y satisfies[1]

$$\pi_y(x_1) \propto \kappa_1^{x_1}\kappa_2^{y-x_1}\binom{y}{x_1}\frac{(\alpha_1)_{(x_1)}(\alpha_2)_{(y-x_1)}}{(|\alpha|)_{(y)}}, \quad x_1\in\{0,1,\dots,y\} \tag{4.1}$$

where the parameters of the Beta-Binomial are

$$\alpha_1=\frac{\mu_1}{\kappa_1},\ \alpha_2=\frac{\mu_2}{\kappa_2},\ |\alpha|=\alpha_1+\alpha_2, \tag{4.2}$$

$(\alpha)_{(x_1)}$ is the ascending factorial or the Pochhammer function, $(\alpha)_{(x_1)}=\alpha(\alpha+1)\dots(\alpha+x_1-1)$. The stationary distribution π_y can also be written in terms of Gamma functions, using $(\alpha)_{(x_1)}=\Gamma(\alpha+x_1)/\Gamma(\alpha)$, as

$$\pi_y(x_1)\propto\kappa_1^{x_1}\kappa_2^{y-x_1}\binom{y}{x_1}\frac{\Gamma(|\alpha|)}{\Gamma(|\alpha|+y)}\frac{\Gamma(\alpha_1+x_1)}{\Gamma(\alpha_1)}\frac{\Gamma(\alpha_2+y-x_1)}{\Gamma(\alpha_2)}, \quad x_1\in\{0,1,\dots,y\}.$$

4.1. Means under the stationary distribution for the fast process for two species. In order to calculate the means $m_i^{\mu,\kappa}(y)=\mathbb{E}_{\pi_y}[x_i],\ i=1,2$, we need to evaluate the normalizing constant for π_y:

$$u^{\mu,\kappa}(y)=\sum_{x_1=0}^{y}\kappa_1^{x_1}\kappa_2^{y-x_1}\binom{y}{x_1}\frac{\Gamma(|\alpha|)}{\Gamma(|\alpha|+y)}\frac{\Gamma(\alpha_1+x_1)}{\Gamma(\alpha_1)}\frac{\Gamma(\alpha_2+y-x_1)}{\Gamma(\alpha_2)},$$

which can be expressed as $u^{\mu,\kappa}(y)=\mathbb{E}_{\pi_y^{BB}}\left[\kappa_1^{\xi}\kappa_2^{y-\xi}\right]=\kappa_2^y\,\Psi_\xi(s)\big|_{s=\ln(\frac{\kappa_1}{\kappa_2})}$ using the Beta-Binomial(y,α) distribution π_y^{BB} for the random variable ξ:

$$\pi_y^{BB}(x_1):=\binom{y}{x_1}\frac{\Gamma(\alpha_1+\alpha_2)}{\Gamma(\alpha_1+\alpha_2+y)}\frac{\Gamma(\alpha_1+x_1)}{\Gamma(\alpha_1)}\frac{\Gamma(\alpha_2+y-x_1)}{\Gamma(\alpha_2)}, \quad x_1\in\{0,1,...,y\}, \tag{4.3}$$

whose moment generating function is denoted by $\Psi_\xi(s)=\mathbb{E}_{\pi_y^{BB}}[e^{s\xi}],\ s\in\mathbb{R}$.

When the rates of autocatalysis are equal, $\kappa_1=\kappa_2=\kappa$, the normalization factor is simply $u^{\mu,\kappa}(y)=\kappa^y\mathbb{E}_{\pi_y^{BB}}[1^\xi]=\kappa^y$, and the stationary distribution reduces to the Beta-Binomial(y,α) distribution π_y^{BB} for which the mean is easy to evaluate:

$$m_1^{\mu,\kappa}(y)=\mathbb{E}_{\pi_y^{BB}}[\xi]=y\frac{\alpha_1}{\alpha_1+\alpha_2}. \tag{4.4}$$

The stationary distribution in this symmetric case was derived in [33], its mean first passage times were computed in [46] and approximated by SDE methods in [7]. The same distribution (4.3) appears, perhaps surprisingly, as the stationary distribution,

[1]Here, when $d=2$, $\pi_y(x)=\pi_y(x_1,y-x_1)$ can be viewed as a function of $x_1\in\{0,1,\dots,y\}$, which we abbreviate as $\pi_y(x_1)$. We shall use this abbreviated notation in this section and the next one. The sign $\propto$ denotes that the expressions are equal up to a multiplication by a normalizing constant.

conditioned on total population size, for the original symmetric TK model on two species — see Theorem 3.1 of [9].

For asymmetric rates of autocatalysis, we use the known expression for the Beta-Binomial(y,α) moment generating function $\Psi_\xi(\cdot)$, in terms of the Gauss hypergeometric function ${}_2F_1(a,b,c;z)$ (cf. [19]):

$$\Psi_\xi(s) = {}_2F_1(-y,\alpha_1,\alpha_1+\alpha_2;1-e^s), \tag{4.5}$$

so that

$$u^{\mu,\kappa}(y) = \kappa_2^y \Psi_\xi(s)\big|_{s=\ln(\frac{\kappa_1}{\kappa_2})} = \kappa_2^y\, {}_2F_1\left(-y,\alpha_1,\alpha_1+\alpha_2;1-\frac{\kappa_1}{\kappa_2}\right)$$

and the mean $m_1^{\mu,\kappa}$ is then given by

$$m_1^{\mu,\kappa}(y) = \frac{\kappa_2^y}{u^{\mu,\kappa}(y)}\sum_{x_1=0}^{y} x_1\left(\frac{\kappa_1}{\kappa_2}\right)^{x_1}\pi_y^{BB}(x_1) = \frac{\partial_s\Psi_\xi(s)}{\Psi_\xi(s)}\Big|_{s=\ln(\frac{\kappa_1}{\kappa_2})}. \tag{4.6}$$

from which one also directly obtains the mean $m_2^{\mu,\kappa}(y) = y - m_1^{\mu,\kappa}(y)$ of $y - x_1 = x_2$.

The Gauss hypergeometric function, when evaluated in the first coordinate at a nonpositive integer $a = -y, y \in \mathbb{Z}_+$ reduces to the polynomial

$$ {}_2F_1(-y,b,c;z) = \sum_{n=0}^{y}(-1)^n\binom{y}{n}\frac{(b)_{(n)}}{(c)_{(n)}}z^n. \tag{4.7}$$

The derivative of the Gauss hypergeometric function with respect to the last variable is known (cf. Chapter 15 in [1]) to satisfy

$$\partial_z\, {}_2F_1(a,b,c;z) = {}_2F_1(a+1,b+1,c+1;z)\,\frac{ab}{c}$$

hence

$$\partial_s\, {}_2F_1(-y,\alpha_1,\alpha_1+\alpha_2;1-e^s) = {}_2F_1(-y+1,\alpha_1+1,\alpha_1+\alpha_2+1;1-e^s)\, ye^s\frac{\alpha_1}{\alpha_1+\alpha_2}$$

and we get an expression for the mean $m_1^{\mu,\kappa}$ in terms of these functions as

$$m_1^{\mu,\kappa}(y) = y\frac{\kappa_1}{\kappa_2}\frac{\alpha_1}{\alpha_1+\alpha_2}\,\frac{{}_2F_1(-y+1,\alpha_1+1,\alpha_1+\alpha_2+1;1-\frac{\kappa_1}{\kappa_2})}{{}_2F_1(-y,\alpha_1,\alpha_1+\alpha_2;1-\frac{\kappa_1}{\kappa_2})}. \tag{4.8}$$

Using asymptotics for the hypergeometric function ${}_2F_1$, we show in Corollary B.1 that

$$\lim_{y\to\infty} m_1^{\mu,\kappa}(y) = \frac{\mu_1}{\kappa_2-\kappa_1}. \tag{4.9}$$

for $0<\kappa_1<\kappa_2$ and $\mu_1,\mu_2>0$.

Note that, when autocatalytic rates are symmetric, $\kappa_1=\kappa_2$, then the mean $m_1^{\mu,\kappa}(y) = y\frac{\alpha_1}{\alpha_1+\alpha_2}$ grows linearly; whereas when autocatalytic rates are asymmetric for $0<\kappa_1<\kappa_2$, $m_1^{\mu,\kappa}(y)$ tends to a (positive) constant as $y\to\infty$. An intuitive explanation of this phenomenon is that: when the autocatalytic rates are symmetric their dynamics is unbiased creating an equal pull towards each boundary; this results in a stationary distribution that is evenly distributed near the two boundaries $x_1 = 0$ and $x_1 = y$ with the proportion determined by the asymmetry in the mutation rates directing the departures away from the boundaries; in the case of asymmetric autocatalytic rates, their bias pulls the dynamics towards one boundary, and this pull counterbalances the departures away from the boundaries for larger values of y due to the quadratic nature of autocatalysis; this results in a stationary distribution that after some value of y is predominantly concentrated in a very small neighborhood of the boundaries - for $\kappa_1 < \kappa_2$ this concentration ends up in a small neighborhood of $x_1 = 0$ with (4.9) determining the mean of the mass not on the boundary itself. This limiting value tends to $\frac{\mu_1}{\kappa_2}$ as $\kappa_1 \to 0$ with μ_1, κ_2 fixed, and it tends to zero as $\mu_1 \to 0$ with $\kappa_1 < \kappa_2$ fixed (see Figures 1 and 2 for a visual illustration).

The above analysis demonstrates fragility, i.e., markedly different qualitative behavior in the symmetric and asymmetric TK models with mutation for large population sizes.

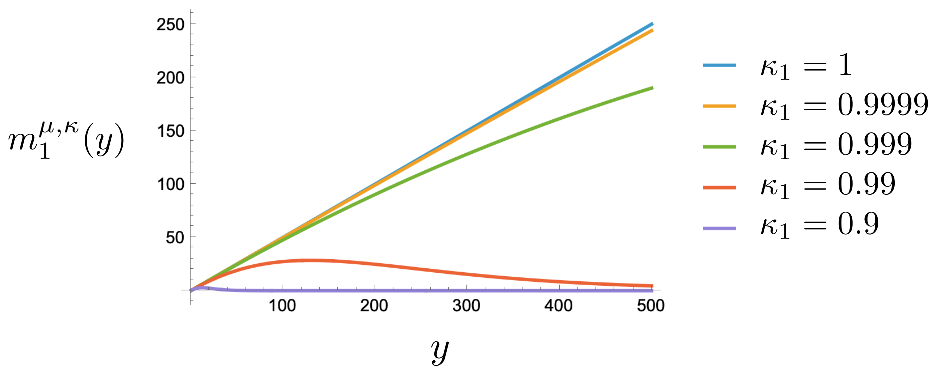

FIG. 1. ***Plot for $m_1^{\mu,\kappa}(\cdot)$ for the two species TK model with mutations.*** *Here, $\alpha_j = \frac{\mu_j}{\kappa_j}$, $j = 1, 2$, and we take $\mu_1 = \mu_2 = 0.01$, $\kappa_2 = 1$ and various κ_1. As y keeps increasing, the green curve ($\kappa_1 = 0.999$) and the orange curve ($\kappa_1 = 0.9999$) will eventually decrease to $\frac{0.01}{1-0.999} = 10$ and $\frac{0.01}{1-0.9999} = 100$, respectively, as predicted by* (4.9).

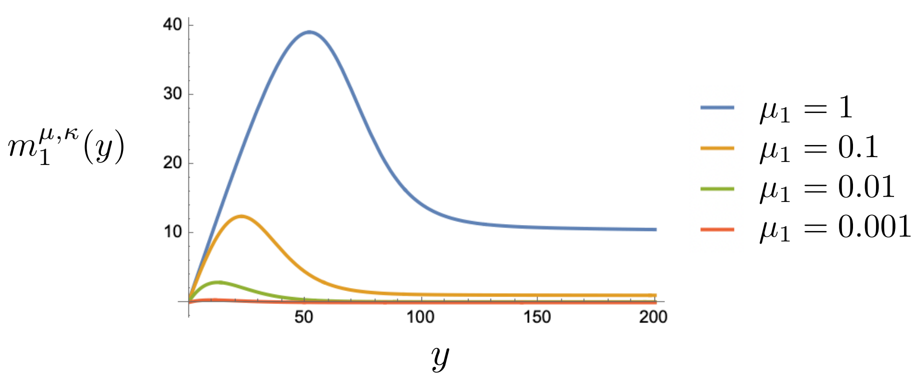

FIG. 2. ***Plot for $m_1^{\mu,\kappa}(\cdot)$ for the two species TK model with mutations.*** *Here, $\alpha_j = \frac{\mu_j}{\kappa_j}$, $j = 1, 2$, and we take $\mu_2 = 0.01$, $\kappa_1 = 0.9$, $\kappa_2 = 1$ and various μ_1. For each μ_1 value, the curve asymptotes towards $\frac{\mu_1}{\kappa_2 - \kappa_1} = 10\mu_1$, as in* (4.9).

4.2. Implication for the slow process. In this section, we explore how fragility in $m_1^{\mu,\kappa}(\cdot)$ can lead to different stationary behaviors of $\bar{Y}$ (whose generator is given in (3.4)) in the symmetric ($\kappa_1 = \kappa_2$) and the asymmetric ($\kappa_1 < \kappa_2$) cases.

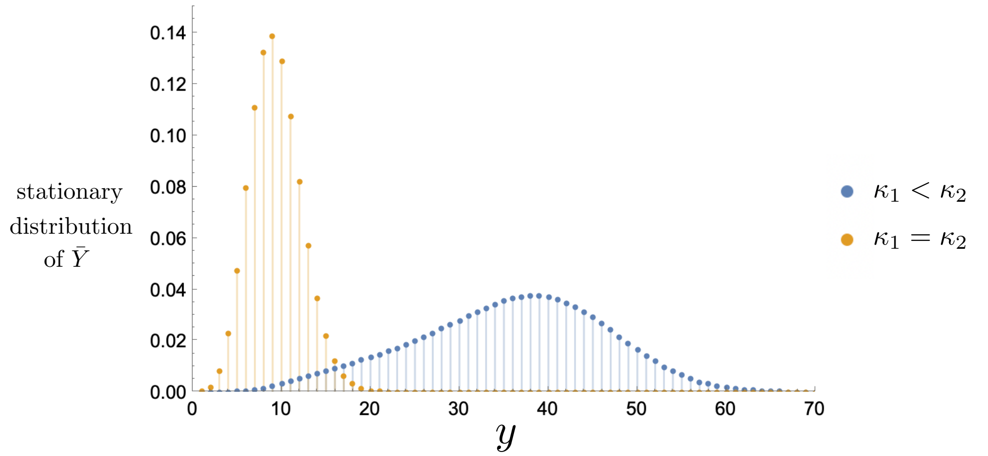

FIG. 3. ***Stationary distribution for $\bar{Y}$ whose generator is given in* (3.4) *for the two species TK model with mutations.*** *Here, we take $\mu_1 = \mu_2 = 0.01$, $\kappa_2 = 1$, $\lambda_1 = \lambda_2 = 2.3$, $\delta_1 = 1$, $\delta_2 = 0.1$, and we take $\kappa_1 = 0.9$ in the asymmetric case (and $\kappa_1 = 1$ in the symmetric case). The mean and the variance of the stationary distribution for the symmetric case are* 8.364 *and* 8.364*, respectively. The mean and the variance of the stationary distribution for the asymmetric case are* 34.592 *and* 114.366*, respectively. Note that for different values of $\kappa_1 = \kappa_2$ while keeping other rate constants fixed, we have the same stationary distribution for $\bar{Y}$ as what is shown in the orange distribution.*

In the symmetric case ($\kappa_1 = \kappa_2$), we have by (4.4) that

$$m_1^{\mu,\kappa}(y) = y\frac{\mu_1}{\mu_1 + \mu_2} \qquad \text{and} \qquad m_2^{\mu,\kappa}(y) = y\frac{\mu_2}{\mu_1 + \mu_2},$$

and so

$$\sum_{i=1}^{2} \delta_i m_i^{\mu,\kappa}(y) = y\frac{\delta_1\mu_1 + \delta_2\mu_2}{\mu_1 + \mu_2}.$$

Hence, $\bar{Y}$ is a birth-death process with constant birth rate and linear death rate (which models an M/M/∞ queue), and its stationary distribution has a Poisson distribution with mean $\frac{(\lambda_1+\lambda_2)(\mu_1+\mu_2)}{\delta_1\mu_1+\delta_2\mu_2}$. Note that the stationary distribution for $\bar{Y}$ is insensitive to the values of $\kappa_1 = \kappa_2$.

In the asymmetric case ($\kappa_1 < \kappa_2$), we have by (4.9) that

$$\sum_{i=1}^{2} \delta_i m_i^{\mu,\kappa}(y) \sim \delta_1 \frac{\mu_1}{\kappa_2 - \kappa_1} + \delta_2 \left(y - \frac{\mu_1}{\kappa_2 - \kappa_1} \right), \tag{4.10}$$

where $G_1(y) \sim G_2(y)$ means $\lim_{y\to\infty} \frac{G_1(y)}{G_2(y)} = 1$.

In Figure 3, we plot the stationary distribution for $\bar{Y}$ in both the symmetric case and the asymmetric case when δ_2 is small. Specifically, we take $\mu_1 = \mu_2 = 0.01$, $\lambda_1 = \lambda_2 = 2.3$, $\delta_1 = 1$ and $\delta_2 = 0.1$. When $\kappa_1 = \kappa_2 = 0.9$ or $\kappa_1 = \kappa_2 = 1$, the stationary distribution for $\bar{Y}$ is a Poisson distribution with mean $\frac{92}{11}$. This Poisson mean does not depend on the value of $\kappa_1 = \kappa_2$. On the other hand, when $\kappa_1 = 0.9$ and $\kappa_2 = 1$, the stationary distribution is no longer Poisson and deviates greatly from those in the symmetric cases (the case of $\kappa_1 = \kappa_2 = 0.9$ and the case of $\kappa_1 = \kappa_2 = 1$).

5. Further analysis of the stationary distribution of the fast process for two species. In this section, we explore small mutation limits for the stationary distribution for the fast process, π_y, and its first-coordinate mean $m_1^{\mu,\kappa}(y)$, which are introduced in Sections 3 and 4. In Section 5.1, we take a stochastic singular perturbation approach [14] to find such limits. In Section 5.2, we apply the Constrained Langevin Approximation (CLA) [3, 16, 42] and explore the limiting behavior of the density for the stationary distribution of the CLA. The two methods give comparable asymptotic results to describe π_y when the mutation rates μ_1 and μ_2 are small, and both approximations capture exactly (4.4) in the symmetric case ($\kappa_1 = \kappa_2$).

5.1. Limit of stationary distribution for the fast process as mutations vanish. Assume that $\tilde{\kappa} = \frac{\kappa_2}{\kappa_1}$ and $\tilde{\mu} = \frac{\mu_2}{\mu_1}$ are constants. As was done for the 1D model in Bruno et al. [14], when $\mu_1 > 0$ (with $\mu_2 = \tilde{\mu}\mu_1$) is small, the stationary distribution π_y for the fast process can be characterized using a power series expansion in terms of μ_1. Compared to the notation used there, $\mathrm{D}_{\mathrm{tot}}$ is replaced by y, D^{R} and D^{A} are replaced by S_1 and S_2, k_E^A/V and k_E^R/V are replaced by κ_1 and κ_2, $\delta_A = \delta + \bar{k}_E^A$ and $\delta_R = \delta + \bar{k}_E^R$ are replaced by μ_1 and μ_2, respectively. From [14], we have that as $\mu_1 \to 0$,

$$\pi_y(x_1) = \begin{cases} \frac{\tilde{\mu}\tilde{\kappa}^{y-1}}{1+\tilde{\mu}\tilde{\kappa}^{y-1}} + O(\mu_1) & \text{for } x_1 = 0 \\ \frac{y^2}{x_1(y-x_1)} \frac{\tilde{\mu}\tilde{\kappa}^{y-x-1}}{1+\tilde{\mu}\tilde{\kappa}^{y-1}}\mu_1 + O(\mu_1^2) & \text{for } x_1 \in \{1, \ldots, y-1\} \\ \frac{1}{1+\tilde{\mu}\tilde{\kappa}^{y-1}} + O(\mu_1) & \text{for } x_1 = y, \end{cases}$$

and then $m_1^{\mu,\kappa}(y)$ is approximately $\frac{y}{1+\tilde{\mu}\tilde{\kappa}^{y-1}}$ when μ_1 and $\mu_2 = \tilde{\mu}\mu_1$ are small. When $\kappa_1 = \kappa_2$, this approximation is the same as (4.4). On the other hand, when $\kappa_1 < \kappa_2$ and we let $y \to \infty$, the approximate value converges to zero, which is the same as the limit obtained by letting $\mu_1 \to 0$ in (4.9).

5.2. Constrained Langevin Approximation for the fast process. We approximate the stationary distribution π_y of the fast process, under a fixed total population size, using the Constrained Langevin Approximation (CLA) [3, 16, 42]. For fixed $y > 0$, suppose that X_1^y denotes the molecular count of species S_1, governed by the generator B_y defined in (3.3). We normalize the species count and scale the reaction rates as follows:

$$\tilde{Z}^y(t) = y^{-1}X_1^y(t), \quad \kappa_1 = y^{-1}\tilde{\kappa}_1, \quad \kappa_2 = y^{-1}\tilde{\kappa}_2.$$

The normalized count of species S_1 then evolves according to

$$\begin{aligned} \tilde{Z}^y(t) = \tilde{Z}^y(0) &+ \frac{1}{y}N_1\left(\int_0^t y\tilde{\kappa}_1\tilde{Z}^y(s)\left(1-\tilde{Z}^y(s)\right)\mathrm{d}s\right) \\ &- \frac{1}{y}N_2\left(\int_0^t y\tilde{\kappa}_2\tilde{Z}^y(s)\left(1-\tilde{Z}^y(s)\right)\mathrm{d}s\right) \\ &+ \frac{1}{y}N_3\left(\int_0^t y\mu_1\left(1-\tilde{Z}^y(s)\right)\mathrm{d}s\right) - \frac{1}{y}N_4\left(\int_0^t y\mu_2\tilde{Z}^y(s)\,\mathrm{d}s\right), \end{aligned}$$

for $t \geq 0$, where N_k, $k = 1, 2, 3, 4$, are independent unit rate Poisson processes.

The CLA for this process (see [16]) is a one-dimensional reflected diffusion process living on $[0, 1]$ satisfying

$$\begin{aligned} Z^y(t) = Z^y(0) &+ \int_0^t b(Z^y(s))\,\mathrm{d}s + \frac{1}{\sqrt{y}}\int_0^t \sigma(Z^y(s))\,\mathrm{d}W(s) \\ &+ \frac{1}{\sqrt{y}}\int_0^t \left(1_{\{0\}}\left(Z^y(s)\right) - 1_{\{1\}}\left(Z^y(s)\right)\right)\mathrm{d}L^y(s), \end{aligned} \tag{5.1}$$

for $t \geq 0$, where $Z^y(0) = \tilde{Z}^y(0)$, $W(\cdot)$ is a standard one-dimensional Brownian motion and $L^y(\cdot)$ is a continuous non-decreasing one-dimensional process that only increases when Z^y is at 0 or 1. Indeed, L^y is the local time of the process Z^y at the boundary points 0 and 1. The drift and diffusion coefficients are

$$b(z) = (\tilde{\kappa}_1 - \tilde{\kappa}_2)z(1-z) + \mu_1(1-z) - \mu_2 z,$$
$$\sigma^2(z) = (\tilde{\kappa}_1 + \tilde{\kappa}_2)z(1-z) + \mu_1(1-z) + \mu_2 z.$$

The stationary distribution for Z^y is given by

$$\rho(z) = \frac{C_y}{\sigma^2(z)} \exp\left(\int_0^z \frac{2b(u)y}{\sigma^2(u)}\,\mathrm{d}u\right), \qquad z \in [0,1], \tag{5.2}$$

where C_y is a normalization constant (see Example 2 in [3] for how to derive this). Using this expression, we analyze the behavior of $\rho(z)$ near the boundary points $z = 0$ and $z = 1$.

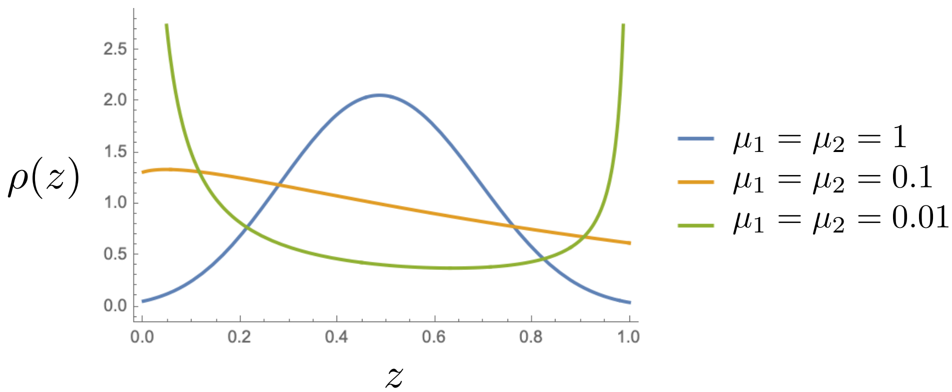

FIG. 4. ***Density $\rho(\cdot)$ of stationary distribution for the CLA for the two species TK model with mutations.*** *Here, we take $V = y = 10$, $\kappa_1 = 0.9$, $\kappa_2 = 1$, and various $\mu_1 = \mu_2$.*

We see from Figure 4 that when μ_1 and μ_2 are small, the density (5.2) of the stationary distribution for the CLA concentrates around 0 and 1, and then $m_1^{\mu,\kappa}(y)$ is approximately $\frac{y\rho(1)}{\rho(1)+\rho(0)} = y/(1 + \frac{\rho(0)}{\rho(1)})$. In the following proposition, we characterize the limit of $\rho(0)/\rho(1)$ as μ_1 and μ_2 go to zero proportionally. From this it can be seen that for $\kappa_1 = \kappa_2$, the approximation for $m_1^{\mu,\kappa}(y)$ in this limit is $y/(1+\tilde{\mu})$, which agrees with (4.4). For $\kappa_1 < \kappa_2$ ($\tilde{\kappa} > 1$), the approximation for $m_1^{\mu,\kappa}(y)$ in this limit is $y/(1 + \tilde{\mu}\exp(2y\frac{\tilde{\kappa}-1}{\tilde{\kappa}+1}))$, which tends to zero as $y \to \infty$ and is consistent with the limit as $\mu_1 \to 0$ of (4.9).

PROPOSITION 5.1. *Assume that $\tilde{\kappa} = \frac{\kappa_2}{\kappa_1} = \frac{\tilde{\kappa}_2}{\tilde{\kappa}_1}$ and $\tilde{\mu} = \frac{\mu_2}{\mu_1}$ are constants. Then, for $\mu = (\mu_1, \mu_2) = (\mu_1, \tilde{\mu}\mu_1)$, we have the following directional limit*

$$\lim_{\mu \to 0} \frac{\rho(0)}{\rho(1)} = \tilde{\mu} \exp\left(2y\frac{\tilde{\kappa}-1}{\tilde{\kappa}+1}\right). \tag{5.3}$$

Proof. In the following, the integrand is bounded in absolute value by 1 (for all $\mu_1 > 0$) and so we can use bounded convergence to take the limit inside the integral and obtain

$$\begin{aligned}\lim_{\mu\to 0}\int_0^1 \frac{b(z)}{\sigma^2(z)}\mathrm{d}z &= \int_0^1 \lim_{\mu_1 \to 0} \frac{(\tilde{\kappa}_1 - \tilde{\kappa}_2)z(1-z) + \mu_1(1-z) - \tilde{\mu}\mu_1 z}{(\tilde{\kappa}_1 + \tilde{\kappa}_2)z(1-z) + \mu_1(1-z) + \tilde{\mu}\mu_1 z}\mathrm{d}z \\ &= \frac{\tilde{\kappa}_1 - \tilde{\kappa}_2}{\tilde{\kappa}_1 + \tilde{\kappa}_2} = \frac{1-\tilde{\kappa}}{1+\tilde{\kappa}}.\end{aligned}$$

Combining this with (5.2) and $\frac{\sigma^2(1)}{\sigma^2(0)} = \frac{\mu_2}{\mu_1} = \tilde{\mu}$, yields the desired result. $\square$

6. Stationary distribution for the fast process with more than two species. In this section, we describe a few sample results emanating from our preliminary exploration of the behavior of the means $m_\bullet^{\mu,\kappa}(\cdot)$ for the fast process when there are three or more species. As far as we know, there are no existing closed-form formulae for the stationary distributions π_y when $d > 2$. Nevertheless, with the total population size fixed at y, the fast process is an irreducible, finite state continuous time Markov chain with infinitesimal generator B_y, and so it is possible to numerically find π_y by solving $\pi_y B_y = 0$ and normalizing.

Recall that $m_i^{\mu,\kappa}(y) = \sum_{x \in \mathbb{N}_0^d : \|x\|_1 = y} x_i \pi_y(x)$ for $i = 1, 2, \ldots, d$. In Figures 5 and 6, we plot $m_i^{\mu,\kappa}(\cdot)$ for the three species and four species TK models with mutations for some sample parameters. Here, for the three species case, each $m_i^{\mu,\kappa}(y)$ tends to grow approximately linearly as y increases, and we did not observe fragility. In the four species case, some $m_i^{\mu,\kappa}(y)$ seem to tend to grow linearly as y increases, while others seem to tend to have finite limits as y increases. This suggests that there may be some fragility in the four-species model. Based on similar preliminary samples, we believe that exploring fragility in multi-species TK-like models is an interesting topic for future study.

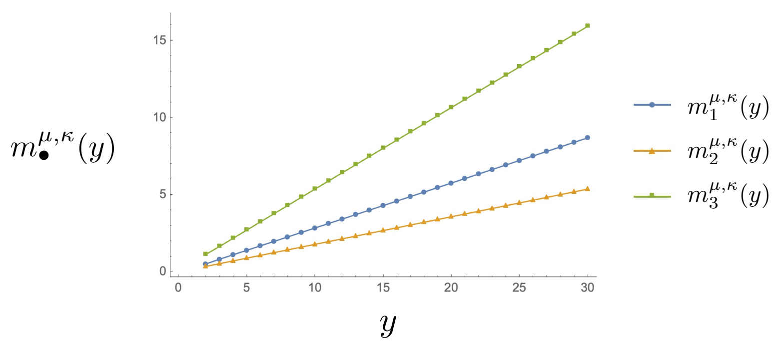

FIG. 5. ***Sample plot for*** $m_i^{\mu,\kappa}(\cdot)$***,*** $i = 1, 2, 3$***, for the three-species TK model with mutations.*** *Here, we take* $\kappa_1 = 1$, $\kappa_2 = 2$, $\kappa_3 = 3$, $\mu_{2,1} = \mu_{3,1} = 0.01$, $\mu_{1,2} = \mu_{3,2} = 0.02$ *and* $\mu_{1,3} = \mu_{2,3} = 0.03$.

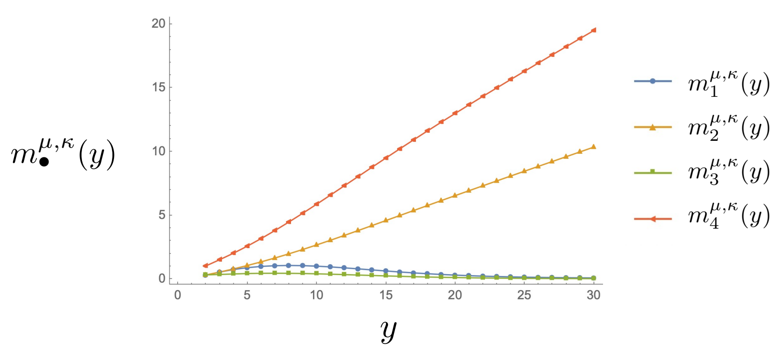

FIG. 6. ***Sample plot for*** $m_i^{\mu,\kappa}(\cdot)$***,*** $i = 1, 2, 3, 4$***, for the four-species TK model with mutations.*** *Here, we take* $\kappa_1 = 1$, $\kappa_2 = 2$, $\kappa_3 = 3$, $\kappa_4 = 4$, $\mu_{2,1} = \mu_{3,1} = \mu_{4,1} = 0.01$, $\mu_{1,2} = \mu_{3,2} = \mu_{4,2} = 0.02$, $\mu_{1,3} = \mu_{2,3} = \mu_{4,3} = 0.03$ *and* $\mu_{1,4} = \mu_{2,4} = \mu_{3,4} = 0.04$.

7. Discussion. In this paper, we provide a multiscale analysis of the d-species TK model with mutations allowing for asymmetry in the reaction rates. In particular, we prove a stochastic averaging principle, which allows us to average out the fast process and describe the dynamics of the slow process (keeping track of the total population size) in terms of certain summary statistics of the fast process. When the slow process is frozen at a given value, the dynamics of the fast process can be described by means of an ergodic continuous-time Markov chain. For the TK model with two species, we provide a thorough analysis of the stationary distribution of this Markov chain and its first moment to demonstrate fragility when there is asymmetry in the autocatalytic reactions. Preliminary numerical calculations for the TK model with more than two species suggest a nontrivial dependence of the fragility on the dimension (number of species) as well as the asymmetry. The clear appearance of fragility in dimension four despite no such indications in dimension three is intriguing, and leaves many open questions for research, which we plan to pursue in the future.

The notion of fragility also connects to a broader theme in the study of chemical reaction networks: stochastic and deterministic descriptions of the same system can differ markedly. As shown in [2], extinction events in the stochastic setting may have no counterpart in deterministic models, where boundary equilibria dominate. Deterministic approaches can sometimes capture switching behavior through stability analysis of equilibria [12], but many switching phenomena arise only stochastically [11]. Autocatalytic systems are particularly sensitive in this respect, as their stochastic and deterministic formulations may diverge sharply [51]. Several approximation methods have been proposed to study such dynamics [8, 24], underscoring their relevance not only in abstract theory but also in biology, where autocatalytic processes are widespread and even hypothesized to have contributed to the origin of life [32].

Appendix A. Proof of Theorem 3.1. Let us define the operator $A : C_b(\mathbb{N}_0) \mapsto C(E)$ as follows:

$$Af(x,y) := \left(\sum_{i=1}^{d} \lambda_i\right)(f(y+1)-f(y)) + \left(\sum_{i=1}^{d} \delta_i x_i\right)(f(y-1)-f(y)), \tag{A.1}$$

for $f \in C_b(\mathbb{N}_0)$ and $(x,y) \in E$. Note that here $f : \mathbb{N}_0 \mapsto \mathbb{R}$ and the function $Af : E \mapsto \mathbb{R}$, where the set E is defined in (2.5). Therefore, while f has only one argument, the function Af has two. The operator A will be useful in the proof of Theorem 3.1, which we now present.

Proof of Theorem 3.1. We will follow [39, Theorem 2.1]. Let $\{\mathcal{F}_t^\varepsilon, t \geq 0\}$ denote the filtration generated by $(X^\varepsilon, Y^\varepsilon)$, *i.e.*, $\mathcal{F}_t^\varepsilon := \sigma((X^\varepsilon(s), Y^\varepsilon(s)) \mid s \leq t)$ for $t \geq 0$. First note that the family of processes $\{Y^\varepsilon, \varepsilon \in (0, \varepsilon_0)\}$ satisfies the compact containment condition: for each $T > 0$, $a \in (0,1)$, there exists a compact set $K_{a,T} \subset \mathbb{N}_0$ such that

$$\inf_{\varepsilon} \mathbb{P}\left(Y^\varepsilon(t) \in K_{a,T}, t \leq T\right) \geq 1 - a.$$

To see that this holds, for fixed $T > 0$, note that for $0 \leq t \leq T$,

$$Y^\varepsilon(t) \leq y(0) + \sum_{i=1}^{d} N_i^b \left(\int_0^t \lambda_i \mathrm{d}u\right),$$

which implies that

$$\sup_{t\le T} Y^{\varepsilon}(t) \le y(0) + \sum_{i=1}^{d} N_i^b \left(\int_0^T \lambda_i \mathrm{d}u \right), \tag{A.2}$$

since sample paths of Poisson processes are nondecreasing. For $a \in (0,1)$, let $n_{a,T}$ be a sufficiently large positive integer that $\mathbb{P}\left(Z_T \le n_{a,T}\right) \ge 1-a$, where the random variable $Z_T \sim \text{Poisson}(T\sum_{i=1}^d \lambda_i)$. Then, we can choose $K_{a,T} = \{0, 1, \ldots, n_{a,T} + y(0)\}$ so that

$$\inf_{\varepsilon} \mathbb{P}\left(Y^{\varepsilon}(t) \in K_{a,T}, t \le T\right) \ge \inf_{\varepsilon} \mathbb{P}\left(Z_T \le n_{a,T}\right) \ge 1-a. \tag{A.3}$$

Since $Y^{\varepsilon} = \|X^{\varepsilon}\|_1$, the above also immediately implies that the collection $\{X^{\varepsilon}(t), t \in [0,T], \varepsilon \in (0,\varepsilon_0)\}$ is relatively compact as a collection of $\mathbb{N}_0^d$-valued random variables for each $T > 0$. In fact, the collection of random measures $\{\Gamma^{\varepsilon}, \varepsilon \in (0,\varepsilon_0)\}$, where Γ^{ε} is defined on $[0,\infty) \times \mathbb{N}_0^d$ such that

$$\Gamma^{\varepsilon}([0,t] \times U) := \int_0^t 1_U \left(X^{\varepsilon}(s)\right) \mathrm{d}s, \quad t \ge 0, U \subset \mathbb{N}_0^d,$$

is relatively compact as a collection of $\bar{M}([0,\infty) \times \mathbb{N}_0^d)$-valued random elements. This follows from Lemma 1.3 in [39] and the fact that for any $T > 0$, $a \in (0,1)$, with $K_{a,T}$ as in (A.3) and $\tilde{K}_{a,T} = \{x \in \mathbb{N}_0^d : \sum_{i=1}^d x_i \in K_{a,T}\}$, we have using (A.3) that

$$\begin{aligned}\mathbb{E}\left[\Gamma^{\varepsilon}([0,T] \times \tilde{K}_{a,T})\right] &= \int_0^T \mathbb{P}\left(X^{\varepsilon}(s) \in \tilde{K}_{a,T}\right) \mathrm{d}s \\ &\ge T\,\mathbb{P}\left(\sup_{t \le T} Y^{\varepsilon}(t) \in K_{a,T}\right) \ge T(1-a).\end{aligned}$$

To show relative compactness of the collection $\{Y^{\varepsilon}, \varepsilon \in (0,\varepsilon_0)\}$, given the compact containment proved above, it suffices to prove (see Theorems 3.9.1 and 3.9.4 of [23]) that for every $f \in C_b(\mathbb{N}_0)$, $\varepsilon \in (0,\varepsilon_0)$, the stochastic process

$$M_f^{\varepsilon}(t) := f(Y^{\varepsilon}(t)) - f(Y^{\varepsilon}(0)) - \int_0^t Af(X^{\varepsilon}(s), Y^{\varepsilon}(s))\mathrm{d}s, \quad t \ge 0, \tag{A.4}$$

is an $\{\mathcal{F}_t^{\varepsilon}\}$-martingale, and for each $T > 0$,

$$\sup_{\varepsilon} \mathbb{E}\left[\int_0^T |Af(X^{\varepsilon}(s), Y^{\varepsilon}(s))|^2 \mathrm{d}s\right] < \infty, \tag{A.5}$$

where A is the operator defined in (A.1). To see these properties, fix $f \in C_b(\mathbb{N}_0)$, $\varepsilon \in (0,\varepsilon_0)$ and $T > 0$. Define the function $h_f \in C_b(E)$ by $h_f(x,y) := f(y)$ for $(x,y) \in E$, so that h_f does not depend on the first coordinate. Then, for all $(x,y) \in E$,

$$Q(\varepsilon)h_f(x,y) = \sum_{i=1}^{d} \lambda_i (f(y+1) - f(y)) + \sum_{i=1}^{d} \delta_i x_i (f(y-1) - f(y)) = Af(x,y),$$

i.e., the two functions $Q(\varepsilon)h_f$ and Af are identical. Since $Q(\varepsilon)$ is the infinitesimal generator for the Markov process $(X^{\varepsilon}, Y^{\varepsilon})$, then $\{h_f(X^{\varepsilon}(t), Y^{\varepsilon}(t)) - h_f(X^{\varepsilon}(0), Y^{\varepsilon}(0)) -$

$\int_0^t Q(\varepsilon)h_f(X^\varepsilon(s),Y^\varepsilon(s))\mathrm{d}s, t \geq 0\}$ is a local $\{\mathcal{F}_t^\varepsilon\}$-martingale (see [4, 23]), and since this is the same as M_f^ε, defined in (A.4), the local martingale property also holds for M_f^ε. Next, we observe that by (A.1)-(A.2),

$$\begin{aligned}
\mathbb{E}\left[\int_0^T |Af(X^\varepsilon(s),Y^\varepsilon(s))|^2 \mathrm{d}s\right] &\leq 8|f|_\infty^2 T\,\mathbb{E}\left[\left(\sum_{i=1}^d \lambda_i\right)^2 + \left(\max_{i=1}^d \delta_i\right)^2\left(\sup_{t\leq T} Y^\varepsilon(t)\right)^2\right] \\
&\leq 8|f|_\infty^2 T\left(C_1^2 + C_2^2\mathbb{E}\left[(y(0)+Z_T)^2\right]\right) \\
&\leq 8|f|_\infty^2 T\left(C_1^2 + 2C_2^2(y(0)^2 + C_1T + C_1^2T^2)\right), \qquad \text{(A.6)}
\end{aligned}$$

where $|f|_\infty = \sup_{y\in\mathbb{N}_0}|f(y)|$, $C_1 = \sum_{i=1}^d \lambda_i, C_2 = \max_{i=1}^d \delta_i$ and we used the fact that $\mathbb{E}\left[Z_T^2\right] = C_1T + C_1^2T^2$. Combining this with the boundedness of f and the Cauchy-Schwarz inequality, we can see that for each $t \geq 0$, $\{M_f^\varepsilon(t\wedge\tau_k)\}_{k=1}^\infty$ is L^2-bounded and therefore uniformly integrable for any sequence of stopping times $\{\tau_k\}_{k=1}^\infty$. It follows that M_f^ε is in fact an $\{\mathcal{F}_t^\varepsilon\}$-martingale, not just a local martingale (see Proposition 1.8 of [17]). Furthermore, we see from the above inequality, where the last line is independent of ε, that (A.5) holds. Hence, $\{Y^\varepsilon, \varepsilon \in (0,\varepsilon_0)\}$ is relatively compact.

Combining all of the above, we have that the collection of random elements $\{(Y^\varepsilon,\Gamma^\varepsilon), \varepsilon \in (0,\varepsilon_0)\}$, which take values in $D([0,\infty),\mathbb{N}_0)\times\bar{M}([0,\infty)\times\mathbb{N}_0^d)$, is relatively compact. By the same reasoning as in the proof of [39, Theorem 2.1], for any weak limit point $(\bar{Y},\Gamma)$ of $\{(Y^\varepsilon,\Gamma^\varepsilon), \varepsilon\in(0,\varepsilon)\}$ along a sequence of ε tending to 0, we have that the process

$$f(\bar{Y}(t)) - f(\bar{Y}(0)) - \int_{[0,t]\times\mathbb{N}_0^d} Af(x,\bar{Y}(s))\Gamma(\mathrm{d}s\times\mathrm{d}x), \quad t\geq 0, \qquad \text{(A.7)}$$

is an $\{\mathcal{F}_t\}$-martingale for all $f \in C_b(\mathbb{N}_0)$, where we can take the filtration $\{\mathcal{F}_t, t\geq 0\}$ to be the natural filtration generated by $(\bar{Y},\Gamma)$, *i.e.*, $\mathcal{F}_t := \sigma(\bar{Y}(s),\Gamma([0,s]\times U) \mid s \leq t, U\subset\mathbb{N}_0^d)$.

Now, define the operator $B : C_b(\mathbb{N}_0^d) \mapsto C(E)$ as $Bg(x,y) := B_y g(x)$, where B_y is the operator defined in (3.3). Setting $h_g \in C_b(E)$ as $h_g(x,y) = g(x)$ for $g \in C_b(\mathbb{N}_0^d)$, we see that the process M_g^ε defined for $t \geq 0$ by

$$\begin{aligned}
M_g^\varepsilon(t) &= h_g(X^\varepsilon(t),Y^\varepsilon(t)) - h_g(X^\varepsilon(0),Y^\varepsilon(0)) - \int_0^t Q(\varepsilon)h_g(X^\varepsilon(s),Y^\varepsilon(s))\mathrm{d}s \\
&= g(X^\varepsilon(t)) - g(X^\varepsilon(0)) - \int_0^t \varepsilon^{-1}Bg(X^\varepsilon(s),Y^\varepsilon(s))\mathrm{d}s - \Delta_g^\varepsilon(t),
\end{aligned}$$

is a local $\{\mathcal{F}_t^\varepsilon\}$-martingale, where $\Delta_g^\varepsilon(t)$ is given by

$$\begin{aligned}
\Delta_g^\varepsilon(t) = &\int_0^t \sum_{i=1}^d \lambda_i(g(X^\varepsilon(s)+e_i) - g(X^\varepsilon(s)))\mathrm{d}s \\
&+ \int_0^t \sum_{i=1}^d \delta_i X_i^\varepsilon(s)(g(X^\varepsilon(s)-e_i) - g(X^\varepsilon(s)))\mathrm{d}s.
\end{aligned}$$

By similar reasoning to that surrounding (A.6), we can show that for each $T > 0$, $\mathbb{E}\left[\int_0^T |Q(\varepsilon)h_g(X^\varepsilon(s),Y^\varepsilon(s))|^2\mathrm{d}s\right] < \infty$ and conclude that M_g^ε is in fact an $\{\mathcal{F}_t^\varepsilon\}$-

martingale, not just a local martingale. Also, for $T > 0$,

$$\lim_{\varepsilon \to 0} \mathbb{E}\left[\sup_{t \le T} \varepsilon |\Delta_g^\varepsilon(t)|\right] \le CT \lim_{\varepsilon \to 0} \mathbb{E}\left[\sup_{t \le T} \varepsilon Y^\varepsilon(t)\right] \le \lim_{\varepsilon \to 0} \varepsilon CT \left(y(0) + T \sum_{i=1}^d \lambda_i\right) = 0,$$

for a positive constant C. Therefore, by multiplying M_g^ε with ε and letting $\varepsilon \to 0$, using the weak convergence of $(Y^\varepsilon, \Gamma^\varepsilon)$ to $(\bar{Y}, \Gamma)$ along a subsequence of ε tending to zero, and the boundedness of the function g, we obtain that

$$\int_{[0,t] \times \mathbb{N}_0^d} Bg(x, \bar{Y}(s)) \Gamma(\mathrm{d}s \times \mathrm{d}x), \quad t \ge 0,$$

is a continuous, locally of bounded variation $\{\mathcal{F}_t\}$-martingale, and hence, is constant. That is, for all $g \in C_b(\mathbb{N}_0^d)$, we have

$$\int_{[0,t] \times \mathbb{N}_0^d} Bg(x, \bar{Y}(s)) \Gamma(\mathrm{d}s \times \mathrm{d}x) = 0, \quad \text{for all } t \ge 0$$

almost surely. Following [39, Example 2.3], since the CTMC generated by B_y is ergodic and admits a unique stationary distribution π_y, which is a probability measure on $\mathbb{N}_0^d$ concentrated on E_y for each $y \in \mathbb{N}_0$, it follows that *almost surely*

$$\Gamma(\mathrm{d}s \times \mathrm{d}x) = \mathrm{d}s \times \pi_{\bar{Y}(s)}(\mathrm{d}x), \quad \text{on } [0, \infty) \times \mathbb{N}_0^d.$$

Then, from (A.7), we may conclude that the limit point $\bar{Y}$ is a solution to the martingale problem for the generator

$$\begin{aligned} \sum_{x \in E_y} Af(x, y)\pi_y(x) &= \left(\sum_{i=1}^d \lambda_i\right)(f(y+1) - f(y)) \\ &\quad + \left(\sum_{i=1}^d \delta_i m_i^{\mu,\kappa}(y)\right)(f(y-1) - f(y)) \\ &= Gf(y), \end{aligned} \qquad \square$$

where $f \in C_b(\mathbb{N}_0)$, $m_i^{\mu,\kappa}(y) = \sum_{x \in E_y} x_i \pi_y(x)$ and the operator G is the generator in (3.4). Because the process $\bar{Y}$ is upper bounded by a pure production (birth) process with production rate $\sum_{i=1}^d \lambda_i$, the solution to the martingale problem can be shown to be unique up to stopping times $T_k := \inf\{t > 0 : \bar{Y}(t) = k\}$ for $k \in \mathbb{N}_0$, and since $\lim_{k \to \infty} T_k = \infty$ *almost surely*, the uniqueness extends to the situation without the stopping times. From this uniqueness, it follows that $\{Y^\varepsilon, \varepsilon \in (0, \varepsilon_0)\}$ converges weakly to the unique solution of the martingale problem. This completes the proof.

Appendix B. Asymptotics for the hypergeometric function and derivation of (4.9)**.**

The hypergeometric function is given for $0 < z < 1$, $c \neq 0, -1, -2, \ldots$, by the Gauss power series

$${}_2F_1(a, b, c; z) := \sum_{n=0}^\infty \frac{(a)_{(n)}(b)_{(n)}}{(c)_{(n)}} \frac{z^n}{n!},$$

where $(\alpha)_{(n)}$ is the rising factorial or the (ascending) Pochhammer function, which is such that $(\alpha)_{(0)} = 1$ and for $n \geq 1$, $(\alpha)_{(n)} = \alpha(\alpha+1)\dots(\alpha+n-1)$. First, note that

$$ {}_2F_1(-a,b,c;z) = {}_2F_1(b,-a,c;z) $$

by symmetry with respect to the first two arguments in the ${}_2F_1$ function.

Lemma B.1. Fix $0 < b < c$ and $0 < z < 1$. Then,

$$ \lim_{a\to\infty} a\frac{{}_2F_1(-a+1,b+1,c+1;z)}{{}_2F_1(-a,b,c;z)} = \frac{c}{z}. \tag{B.1} $$

Proof of Lemma B.1. For this proof, we use asymptotics of ${}_2F_1$, following Temme [47]. The function ${}_2F_1$ has an integral representation of the form

$$ {}_2F_1(-a,b,c;z) = \frac{\Gamma(c)}{\Gamma(b)\Gamma(c-b)}\int_0^1 t^{b-1}(1-t)^{c-b-1}(1-zt)^a dt, $$

for $c > b > 0, a > 0, 0 < z < 1$ (see (28.3.39) in [47]), which can be transformed into the following (see (28.3.51) in [47]):

$$ {}_2F_1(-a,b,c;z) = {}_2F_1(b,-a,c;z) = (1-z)^{c-b-1}\left(\frac{\ln(1-z)}{-z}\right)^{c-1} F_a(b,c,w) $$

where $w = (a+1)\ln(1-z)$ and

$$ F_a(b,c,w) = \frac{\Gamma(c)}{\Gamma(b)\Gamma(c-b)}\int_0^1 f(s)s^{b-1}(1-s)^{c-b-1}e^{ws}ds, \tag{B.2} $$

with $\phi(u) = \left(\frac{e^u-1}{u}\right)^{b-1}$, $\psi(u) = \left(\frac{e^u-1}{u}\right)^{c-b-1}$ for $u \neq 0$ and $f(s) = \phi(s\ln(1-z))\psi((s-1)\ln(1-z))$ for $s \in (0,1)$. Extending ϕ, ψ and f by continuity to be defined at 0 and since there is a power series expansion for $\frac{e^u-1}{u}$ near $u = 0$, there is a power series expansion for

$$ g(s) = (1-s)^{c-b-1}f(s) \tag{B.3} $$

for all s in a sufficiently small neighborhood of $s = 0$, given by $g(s) = \sum_{n=0}^\infty \beta_n^{b,c,z}s^n$ where $\beta_n^{b,c,z} = \frac{\frac{d^n g}{ds^n}(0)}{n!}$. Then, by Watson's Lemma (a version of Laplace's method), as $w \to -\infty$,

$$ F_a(b,c,w) \sim \frac{\Gamma(c)}{\Gamma(c-b)}\sum_{n=0}^\infty \beta_n^{b,c,z}\frac{(b)_{(n)}}{(-w)^{n+b}}, \tag{B.4} $$

where $G_1(w) \sim G_2(w)$ means $\lim_{w\to-\infty}\frac{G_1(w)}{G_2(w)} = 1$. Note that $\beta_0^{b,c,z} = f(0) = \phi(0)\psi(-\ln(1-z)) = \left(\frac{-z}{(1-z)\ln(1-z)}\right)^{c-b-1}$.

Applying these asymptotics to the quantities in (B.1) where $0 < z < 1$, $c > b > 0$, we obtain for $a > 1$,

$$ a\frac{{}_2F_1(-a+1,b+1,c+1;z)}{{}_2F_1(-a,b,c;z)} = a\frac{(1-z)^{c-b-1}\left(\frac{\ln(1-z)}{-z}\right)^{c} F_a(b+1,c+1,a\ln(1-z))}{(1-z)^{c-b-1}\left(\frac{\ln(1-z)}{-z}\right)^{c-1} F_a(b,c,(a+1)\ln(1-z))} $$

$$= a\left(\frac{\ln(1-z)}{-z}\right)\frac{F_a(b+1,c+1,a\ln(1-z))}{F_a(b,c,(a+1)\ln(1-z))}$$

$$\sim a\left(\frac{\ln(1-z)}{-z}\right)\frac{\frac{\Gamma(c+1)}{\Gamma(c-b)}\sum_{n=0}^{\infty}\beta_n^{b+1,c+1,z}\frac{((b+1))_{(n)}}{(-a\ln(1-z))^{n+b+1}}}{\frac{\Gamma(c)}{\Gamma(c-b)}\sum_{n=0}^{\infty}\beta_n^{b,c,z}\frac{(b)_{(n)}}{(-(a+1)\ln(1-z))^{n+b}}}$$

$$\sim a\left(\frac{\ln(1-z)}{-z}\right)\frac{c\Gamma(c)\left(\beta_0^{b+1,c+1,z}\frac{((b+1))_{(0)}}{(-a\ln(1-z))^{b+1}}\right)}{\Gamma(c)\left(\beta_0^{b,c,z}\frac{(b)_{(0)}}{(-(a+1)\ln(1-z))^{b}}\right)}$$

$$\sim \frac{c}{z}\left(\frac{a+1}{a}\right)^b,$$

since $\beta_0^{b+1,c+1,z} = \beta_0^{b,c,z}$, and where $H_1(a) \sim H_2(a)$ means $\lim_{a\to\infty}\frac{H_1(a)}{H_2(a)} = 1$. The desired result in Lemma B.1 follows directly by taking the limit as $a \to \infty$ in the right member above. ∎

Remark B.1. One way to derive (B.4) is to apply Watson's Lemma (a version of Laplace's asymptotic method), which states that if $0 \le T \le \infty$, $h(s) = s^\lambda g(s)$ for $s \in (0,T)$ where $\lambda > -1$, $g(s)$ is infinitely differentiable for all sufficiently small $s \ge 0$ with $g(0) \neq 0$ and $\int_0^T |h(s)|\mathrm{d}s < \infty$, then

$$\int_0^T e^{-\gamma s}h(s)\mathrm{d}s \sim \sum_{n=0}^{\infty}\frac{\frac{\mathrm{d}^n g}{\mathrm{d}s^n}(0)}{n!}\frac{\Gamma(\lambda+n+1)}{\gamma^{\lambda+n+1}}, \qquad \text{as} \qquad \gamma\to\infty.$$

To obtain (B.4), we take $T = 1$, $\lambda = b-1$, $\gamma = -w$ and $g(\cdot)$ as in (B.3).

Corollary B.1. Fix $0 < \kappa_1 < \kappa_2$ and $\mu_1, \mu_2 > 0$. Then,

$$\lim_{y\to\infty} m_1^{\mu,\kappa}(y) = \frac{\mu_1}{\kappa_2-\kappa_1}.$$

Proof of Corollary B.1. We apply Lemma B.1 with $a = y$, $b = \alpha_1 = \frac{\mu_1}{\kappa_1}$, $c = \alpha_1+\alpha_2 = \frac{\mu_1}{\kappa_1}+\frac{\mu_2}{\kappa_2}$, $z = 1-\frac{\kappa_1}{\kappa_2}$ to obtain:

$$\begin{aligned}\lim_{y\to\infty} m_1^{\mu,\kappa}(y) &= \frac{\kappa_1}{\kappa_2}\frac{\alpha_1}{\alpha_1+\alpha_2}\lim_{y\to\infty} y\frac{{}_2F_1(-y+1,\alpha_1+1,\alpha_1+\alpha_2+1;1-\frac{\kappa_1}{\kappa_2})}{{}_2F_1(-y,\alpha_1,\alpha_1+\alpha_2;1-\frac{\kappa_1}{\kappa_2})}\\ &= \frac{\kappa_1}{\kappa_2}\frac{\alpha_1}{\alpha_1+\alpha_2}\lim_{a\to\infty} a\frac{{}_2F_1(-a+1,b+1,c+1;z)}{{}_2F_1(-a,b,c;z)}\\ &= \frac{\kappa_1}{\kappa_2}\frac{\alpha_1}{\alpha_1+\alpha_2}\frac{\alpha_1+\alpha_2}{1-\frac{\kappa_1}{\kappa_2}} = \frac{\mu_1}{\kappa_2-\kappa_1}.\end{aligned}$$

Acknowledgments. The authors gratefully acknowledge the hospitality and generous support of the American Institute of Mathematics (AIM) through their AIM SQuarREs program. We thank Felipe Campos for some numerical simulations of a two-dimensional constrained Langevin approximation considered in a preliminary stage of our work. The authors also wish to acknowledge the AIM Workshop on "Limits and Control of Stochastic Reaction Networks" (organized by Andrea Agazzi and Daniele Cappelletti, July 26–30, 2021), which ultimately led to the formation of our AIM SQuaRE. We thank the workshop organizers and participants for their early discussions on the Togashi–Kaneko model, including those who joined some follow-up conversations, but ultimately were not part of the SQuaRE itself.

Declarations.

Authorship and Contributor Statement. All authors have made substantial intellectual contributions to the study conception, execution, and design of the work. All authors contributed equally and have read and approved the final manuscript.
Conflicts of Interest. The authors declare there are no conflicts of interest.
Data and Code Availability Statement. Mathematica code used to produce the figures is available in the Supplementary Information.
Funding. W. R. KhudaBukhsh was supported by the Engineering and Physical Sciences Research Council (EPSRC) [grant number EP/Y027795/1] of the UK Research and Innovation (UKRI). R. J. Williams was supported in part by NSF Grant DMS-2153866 and the Charles Lee Powell Foundation. G. A. Rempała was partially supported by the HEALMOD initiative at The Ohio State University. L. Popovic was funded by the Discovery Grants program from the Natural Sciences and Engineering Research Council of Canada (NSERC, RGPIN-2024-04774). This work is an outcome of a SQuaRE project funded by the American Institute of Mathematics (AIM). The authors gratefully acknowledge the hospitality and support of AIM.

REFERENCES

[1] M. Abramowitz and I. A. Stegun, eds., *Handbook of Mathematical Functions*, vol. 55 of Applied Mathematics Series, National Bureau of Standards, tenth printing, 1972.

[2] D. F. Anderson and D. Cappelletti, *Discrepancies between extinction events and boundary equilibria in reaction networks*, J. Math. Biol., 79 (2019), pp. 1253–1277, https://doi.org/10.1007/s00285-019-01394-9.

[3] D. F. Anderson, D. J. Higham, S. C. Leite, and R. J. Williams, *On constrained langevin equations and (bio)chemical reaction networks*, Multiscale Modeling & Simulation, 17 (2019), p. 1–30, https://doi.org/10.1137/18m1190999.

[4] D. F. Anderson and T. G. Kurtz, *Stochastic analysis of biochemical systems*, vol. 1.2 of Mathematical Biosciences Institute Lecture Series. Stochastics in Biological Systems, Springer, Cham; MBI Mathematical Biosciences Institute, Ohio State University, Columbus, OH, 2015, https://doi.org/10.1007/978-3-319-16895-1.

[5] V. V. Anisimov, *Switching processes in queueing models*, Applied Stochastic Methods Series, ISTE, London; John Wiley & Sons, Inc., Hoboken, NJ, 2008, https://doi.org/10.1002/9780470611340.

[6] K. Ball, T. G. Kurtz, L. Popovic, and G. A. Rempala, *Asymptotic analysis of multiscale approximations to reaction networks*, Annals of Applied Probability, 16 (2006), pp. 1925–1961.

[7] T. Biancalani, L. Dyson, and A. J. McKane, *Noise-induced bistable states and their mean switching time in foraging colonies*, Physical Review Letters, 112 (2014), https://doi.org/10.1103/physrevlett.112.038101.

[8] T. Biancalani, T. Rogers, and A. J. McKane, *Noise-induced metastability in biochemical networks*, Physical Review E, 86 (2012), https://doi.org/10.1103/physreve.86.010106.

[9] E. Bibbona, J. Kim, and C. Wiuf, *Stationary distributions of systems with discreteness-induced transitions*, Journal of The Royal Society Interface, 17 (2020), p. 20200243, https://doi.org/10.1098/rsif.2020.0243.

[10] P. Billingsley, *Convergence of probability measures*, Wiley Series in Probability and Statistics: Probability and Statistics, John Wiley & Sons, Inc., New York, second ed., 1999, https://doi.org/10.1002/9780470316962. A Wiley-Interscience Publication.

[11] L. M. Bishop and H. Qian, *Stochastic bistability and bifurcation in a mesoscopic signaling system with autocatalytic kinase*, Biophysical Journal, 98 (2010), p. 1–11, https://doi.org/10.1016/j.bpj.2009.09.055.

[12] A. Blokhuis, P. F. Stadler, and N. Vassena, *Stoichiometric recipes for periodic oscillations in reaction networks*, 2025, https://doi.org/10.48550/ARXIV.2508.15273.

[13] V. S. Borkar, *Stochastic approximation*, Cambridge University Press, Cambridge; Hindustan Book Agency, New Delhi, 2008. A dynamical systems viewpoint.

[14] S. Bruno, F. A. Campos, Y. Fu, D. Del Vecchio, and R. J. Williams, *Analysis of singularly perturbed stochastic chemical reaction networks motivated by applications to epigenetic cell*

memory, SIAM Journal on Applied Dynamical Systems, 23 (2024), pp. 2695–2731.

[15] D. V. D. Bruno S, Williams RJ, *Epigenetic cell memory: The gene's inner chromatin modification circuit*, PLoS Comput. Biol., 18 (2022), https://doi.org/https://doi.org/10.1371/journal.pcbi.1009961.

[16] F. A. Campos and R. J. Williams, *Error bounds for one-dimensional constrained Langevin approximations for nearly density-dependent Markov chains*, Adv. in Appl. Probab., 56 (2024), pp. 825–867, https://doi.org/10.1017/apr.2024.25.

[17] K. L. Chung and R. J. Williams, *Introduction to Stochastic Integration*, Birkhäuser, New York NY, second ed., 1990.

[18] A. Crudu, A. Debussche, A. Muller, and O. Radulescu, *Convergence of stochastic gene networks to hybrid piecewise deterministic processes*, The Annals of Applied Probability, 22 (2012), https://doi.org/10.1214/11-aap814.

[19] J. M. Dickey, *Multiple hypergeometric functions: Probabilistic interpretations and statistical uses*, Journal of the American Statistical Association, 78 (1983), pp. 628–637.

[20] T. Enger and P. Pfaffelhuber, *A unified framework for limit results in chemical reaction networks on multiple time-scales*, Electronic Journal of Probability, 28 (2023), https://doi.org/10.1214/22-ejp897.

[21] I. R. Epstein and J. A. Pojman, *An Introduction to Nonlinear Chemical Dynamics: Oscillations, Waves, Patterns, and Chaos*, Oxford University Press, New York, 1998, https://global.oup.com/academic/product/an-introduction-to-nonlinear-chemical-dynamics-9780195096705.

[22] A. Etheridge and R. Griffiths, *A coalescent dual process in a Moran model with genic selection*, Theoretical Population Biology, 75 (2009), p. 320–330, https://doi.org/10.1016/j.tpb.2009.03.004.

[23] S. N. Ethier and T. G. Kurtz, *Markov Processes*, Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics, John Wiley & Sons, Inc., New York, 1986, https://doi.org/10.1002/9780470316658. Characterization and convergence.

[24] W.-T. Fan, Y. Yang, and C. Yuan, *Constrained Langevin approximation for the Togashi-Kaneko model of autocatalytic reactions*, Math. Biosci. Eng., 20 (2023), pp. 4322–4352, https://doi.org/10.3934/mbe.2023201.

[25] J. E. Ferrell, *Feedback loops and reciprocal regulation: recurring motifs in the systems biology of the cell cycle*, Current Opinion in Cell Biology, 25 (2013), pp. 676–686, https://doi.org/10.1016/j.ceb.2013.07.007.

[26] C. Gallinger and L. Popovic, *Asymmetric autocatalytic reactions and their stationary distribution*, Royal Society Open Science, 11 (2024), https://doi.org/10.1098/rsos.231878.

[27] A. Ganguly and W. R. KhudaBukhsh, *Functional limit theorems and parameter inference for multiscale stochastic models of enzyme kinetics*, 2024, https://doi.org/10.48550/ARXIV.2409.06565.

[28] A. Ganguly and W. R. KhudaBukhsh, *Asymptotic analysis of the total quasi-steady state approximation for the michaelis–menten enzyme kinetic reactions*, 2025, https://doi.org/10.48550/ARXIV.2503.20145.

[29] E. Gould, H. S. Fraser, T. H. Parker, S. Nakagawa, S. C. Griffith, P. A. Vesk, F. Fidler, D. G. Hamilton, R. N. Abbey-Lee, J. K. Abbott, L. A. Aguirre, C. Alcaraz, I. Aloni, D. Altschul, K. Arekar, J. W. Atkins, J. Atkinson, C. M. Baker, M. Barrett, K. Bell, and . . . , *Same data, different analysts: variation in effect sizes due to analytical decisions in ecology and evolutionary biology*, BMC Biology, 23 (2025), p. 35, https://doi.org/10.1186/s12915-024-02101-x.

[30] R. Z. Has'minskii, *On stochastic processes defined by differential equations with a small parameter*, Theory of Probability & Its Applications, 11 (1966), p. 211–228, https://doi.org/10.1137/1111018.

[31] F. Herrerías-Azcué and T. Galla, *Consensus and diversity in multistate noisy voter models*, Physical Review E, 100 (2019), p. 022304.

[32] W. Hordijk and M. Steel, *Autocatalytic networks at the basis of life's origin and organization*, Life, 8 (2018), p. 62, https://doi.org/10.3390/life8040062.

[33] B. Houchmandzadeh and M. Vallade, *Exact results for a noise-induced bistable system*, Physical Review E, 91 (2015), https://doi.org/10.1103/physreve.91.022115.

[34] J. Jacod and A. N. Shiryaev, *Limit theorems for stochastic processes*, vol. 288 of Grundlehren der mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], Springer-Verlag, Berlin, second ed., 2003, https://doi.org/10.1007/978-3-662-05265-5.

[35] J. Jhawar, R. G. Morris, U. Amith-Kumar, M. Danny Raj, T. Rogers, H. Rajendran, and V. Guttal, *Noise-induced schooling of fish*, Nature Physics, 16 (2020), pp. 488–493.

[36] H.-W. Kang and T. G. Kurtz, *Separation of time-scales and model reduction for stochastic*

reaction networks, Annals of Applied Probability, 23 (2013), pp. 529–583.
[37] H.-W. Kang, T. G. Kurtz, and L. Popovic, *Central limit theorems and diffusion approximations for multiscale Markov chain models*, Annals of Applied Probability, 24 (2014), pp. 721–759.
[38] R. Z. Khas'minskii, *A limit theorem for the solutions of differential equations with random right-hand sides*, Theory of Probability & Its Applications, 11 (1966), p. 390–406, https://doi.org/10.1137/1111038.
[39] T. G. Kurtz, *Averaging for martingale problems and stochastic approximation*, in Applied stochastic analysis (New Brunswick, NJ, 1991), vol. 177 of Lect. Notes Control Inf. Sci., Springer, Berlin, 1992, pp. 186–209, https://doi.org/10.1007/BFb0007058.
[40] H. Kushner, *Stochastic approximation: a survey*, WIREs Computational Statistics, 2 (2009), p. 87–96, https://doi.org/10.1002/wics.57.
[41] L. Laurence and P. Robert, *Analysis of stochastic chemical reaction networks with a hierarchy of timescales*, J. Stat. Phys., 192 (2025), pp. Paper No. 39, 31, https://doi.org/10.1007/s10955-025-03428-7.
[42] S. C. Leite and R. J. Williams, *A constrained langevin approximation for chemical reaction networks*, The Annals of Applied Probability, 29 (2019), https://doi.org/10.1214/18-aap1421.
[43] E. Pardoux and A. Y. Veretennikov, *Averaging of backward stochastic differential equations, with application to semi-linear PDE's*, Stochastics Stochastics Rep., 60 (1997), pp. 255–270, https://doi.org/10.1080/17442509708834109.
[44] O. Perry and W. Whitt, *A fluid limit for an overloaded X model via a stochastic averaging principle*, Math. Oper. Res., 38 (2013), pp. 294–349, https://doi.org/10.1287/moor.1120.0572.
[45] I. Prigogine and G. Nicolis, *Self-Organization in Nonequilibrium Systems: From Dissipative Structures to Order through Fluctuations*, Wiley, New York, 1977, https://archive.org/details/selforganization0000ncol.
[46] N. Saito and K. Kaneko, *Theoretical analysis of discreteness-induced transition in autocatalytic reaction dynamics*, Physical Review E, 91 (2015), https://doi.org/10.1103/physreve.91.022707.
[47] N. M. Temme, *Asymptotic Methods for Integrals (Series in Analysis; Volume 6)*, World Scientific Publishing Co. Pte. Ltd, 2014, https://doi.org/https://doi.org/10.1142/9195#t=toc.
[48] Y. Togashi and K. Kaneko, *Transitions induced by the discreteness of molecules in a small autocatalytic system*, Phys. Rev. Lett., 86 (2001), pp. 2459–2462, https://doi.org/10.1103/PhysRevLett.86.2459.
[49] Y. Togashi and K. Kaneko, *Alteration of chemical concentrations through discreteness-induced transitions in small autocatalytic systems*, Journal of the Physical Society of Japan, 72 (2003), p. 62–68, https://doi.org/10.1143/jpsj.72.62.
[50] C. Trivellin, P. Rugbjerg, and L. Olsson, *Performance and robustness analysis reveals phenotypic trade-offs in yeast*, Life Science Alliance, 7 (2024), p. e202302215, https://doi.org/10.26508/lsa.202302215.
[51] M. Vellela and H. Qian, *A quasistationary analysis of a stochastic chemical reaction: Keizer's paradox*, Bulletin of Mathematical Biology, 69 (2007), p. 1727–1746, https://doi.org/10.1007/s11538-006-9188-3.
[52] A. Y. Veretennikov, *Large deviations in averaging principle for stochastic differential equation systems (noncompact case)*, Stochastics Stochastics Rep., 48 (1994), pp. 83–96, https://doi.org/10.1080/17442509408833899.
[53] D. J. Wilkinson, *Stochastic modelling for systems biology*, Chapman & Hall/CRC Mathematical and Computational Biology Series, Chapman & Hall/CRC, Boca Raton, FL, 2006.
[54] J. R. Worsfold and R. G. Morris, *Stochastically bistable growth and decay in the Togashi-Kaneko model*, Phys. Rev. E, 112 (2025), p. Paper No. 014131, https://doi.org/10.1103/yf5b-mljb.
[55] Y. Xu, B. Pei, and R. Guo, *Stochastic averaging for slow-fast dynamical systems with fractional Brownian motion*, Discrete Contin. Dyn. Syst. Ser. B, 20 (2015), pp. 2257–2267, https://doi.org/10.3934/dcdsb.2015.20.2257.