

TOM KURTZ’S WORK ON REPRESENTATION AND APPROXIMATION OF STOCHASTIC (BIO)CHEMICAL REACTION NETWORKS

BY RUTH J. WILLIAMS^{1,a}

¹*University of California, San Diego, arjwilliams@ucsd.edu*

In the 1970s, Tom Kurtz established foundational results on the rigorous approximation of continuous time Markov chains used in modeling chemical reaction networks. These results involved approximation of the stochastic processes over compact time intervals, and included functional laws of large numbers (hydrodynamic/fluid limits), functional central limit theorems (linear noise diffusion approximations), and Langevin-type diffusion approximations (in the strong or pathwise sense). Later on, in the mid 2000s, inspired by the use of biochemical reaction networks as models in systems and synthetic biology, with various collaborators, Tom Kurtz expanded on his earlier results, allowing for multiple time and size scales. This short note summarizes some of Tom Kurtz’s contributions in this arena, which have had an enduring impact on the modeling and analysis of stochastic (bio)chemical reaction networks.

1. Introduction. A substantial part of Tom Kurtz’s work involved the development of general theory and methods for rigorously approximating complex Markov process models by more tractable models. These abstract developments were often motivated by and accompanied by applications in the biological, physical and/or engineering sciences. This blend of theory and applications is nicely illustrated in his early CBMS-NSF Lectures Notes on “Approximation of Population Processes” [19], which blossomed into the more comprehensive and often-cited monograph by Ethier and Kurtz [7] —“Markov Processes: Characterization and Convergence.”

One of the early applications of Tom Kurtz’s theories was to stochastic models of systems of chemical reactions. Leveraging theorems on convergence of stochastic processes developed in his early works [12, 13, 14], in the 1970s, Kurtz published a series of papers, especially [15, 16, 17], which established the fundamental theory of functional laws of large numbers (hydrodynamic/fluid limits), functional central limit theorems (linear noise diffusion approximations), and Langevin-type diffusion approximations (in the strong or pathwise sense) for density dependent continuous time Markov chains of which stochastic models of chemical reaction networks are a prime example. These approximations were valid when a scaling parameter, taken in the chemical reaction context to be a measure of volume of the reacting species, was large. In contrast to prior developments, these works provided mathematically rigorous process level approximations rather than heuristic approximations or approximations of the distribution at a single point in time.

In the early period of Tom Kurtz’s career (late 1960s and 1970s), interest in chemical reaction network models was motivated by applications in chemistry, physics and chemical engineering, especially in modeling the dynamics of chemical stirred tank reactors, where all species were present in large numbers. More recently, revolutionary developments in gene sequencing and editing, and advances in experimental measurements at the cellular and molecular level, have driven interest in stochastic modeling of *biochemical* reaction networks, especially in the fields of systems and synthetic biology. In this context, some species

MSC2020 subject classifications. Primary 60J28, 60J60, 60F17, 92C45; secondary 60H10, 92C40.

Keywords and phrases. Chemical reaction networks, reaction rate equation, diffusion approximation, linear noise approximation, Langevin approximation, multiscaling, functional limit theorems.

are only present in small numbers and stochasticity can fundamentally alter cellular dynamics through mechanisms such as noise-induced switching, extinction phenomena, and variability in gene expression. These new applications have driven further developments for stochastic models of chemical reaction networks, where Tom Kurtz was once again at the forefront. With collaborators, he established rigorous multiscaling limits, providing the mathematical foundations for understanding dynamic stochastic effects across multiple temporal and size scales.

In this short note, I will summarize some of the highlights of Tom Kurtz's work on stochastic chemical reaction networks (SCRNs)¹.

2. Stochastic Chemical Reaction Networks. In this section, we provide background on Stochastic Chemical Reaction Networks (SCRNs). The reader is referred to Anderson and Kurtz [4] for a more extensive introduction to this subject. Here $\mathbb{Z}$ is the set of integers, $\mathbb{Z}_+ = \{0, 1, 2, \dots\}$ is the set of non-negative integers, $\mathbb{R}$ is the set of real numbers, and $\mathbb{R}_+$ is the set of non-negative real numbers. For $d \geq 1$, $\mathbb{Z}_+^d$ (resp. $\mathbb{R}^d$, $\mathbb{R}_+^d$) is the set of d -dimensional vectors with entries in $\mathbb{Z}_+$ (resp. $\mathbb{R}$, $\mathbb{R}_+$). For a vector $v \in \mathbb{R}^d$, $|v| = \sum_{i=1}^d |v_i|$, the L^1 -norm of v .

A **Stochastic Chemical Reaction Network (SCRN)** is a continuous time Markov chain X , with state space $\mathcal{X} \subset \mathbb{Z}_+^d$, that keeps track of the number of copies of each of finitely many chemical **species** $S_1, \dots, S_d$ subject to finitely many chemical **reactions** $r \in \mathcal{R}$. In particular, for $i = 1, \dots, d$, $t \geq 0$, the quantity $X_i(t)$ is the number of copies of species S_i in the system at time t . When a reaction $r \in \mathcal{R}$ occurs, $v_{r,i}^-$ copies of species S_i are consumed and $v_{r,i}^+$ copies of species S_i are produced for $i = 1, \dots, d$, where $v_r^- = (v_{r,1}^-, v_{r,2}^-, \dots, v_{r,d}^-)$, $v_r^+ = (v_{r,1}^+, v_{r,2}^+, \dots, v_{r,d}^+) \in \mathbb{Z}_+^d$. The net change in the state of X when reaction r occurs is given by the **reaction vector** $\hat{v}_r = v_r^+ - v_r^-$, which is assumed to be non-zero. The reaction r is usually written as

$$(1) \quad \sum_{i=1}^d v_{r,i}^- S_i \longrightarrow \sum_{i=1}^d v_{r,i}^+ S_i.$$

The rate at which reaction r occurs is given by a **propensity** function $\Lambda_r : \mathcal{X} \rightarrow \mathbb{R}_+$, where $\Lambda_r(x) = 0$ if $x + \hat{v}_r \notin \mathcal{X}$. A common form for a propensity function is the following, associated with **mass action kinetics**, where the rate of the reaction is proportional to the quantities of the species being consumed by the reaction:

$$(2) \quad \Lambda_r(x) = \kappa_r \prod_{i=1}^d (x_i)_{v_{r,i}^-}, x \in \mathcal{X},$$

where $\{\kappa_r\}_{r \in \mathcal{R}}$ are positive constants and for $m \in \mathbb{Z}_+$, the quantity $(m)_l$ is the falling factorial, i.e., $(m)_0 := 1$ and $(m)_l := m(m-1) \cdots (m-l+1)$. For example, for the reaction system

$$(3) \quad \emptyset \rightarrow S_1, \quad 2S_1 \rightarrow S_2, \quad S_1 + S_2 \rightarrow S_3, \quad S_3 \rightarrow \emptyset,$$

where $\emptyset$ denotes an external source or sink, the propensity functions under mass action kinetics would be $\Lambda_1(x) = \kappa_1$, $\Lambda_2(x) = \kappa_2 x_1(x_1 - 1)$, $\Lambda_3(x) = \kappa_3 x_1 x_2$, $\Lambda_4(x) = \kappa_4 x_3$. Let $\mathcal{V} := \{\hat{v}_r : r \in \mathcal{R}\}$, $n := |\mathcal{V}|$, and enumerate the members of $\mathcal{V}$ as $\{v_1, \dots, v_n\}$. Some

¹For brevity and readability, this exposition has focussed on SCRNs. However, the same Markov chain models, with reinterpretation of chemical species as other types of entities, can be used to model other applications such as epidemic and evolutionary dynamics. The reader is referred to Ethier and Kurtz [7], especially Chapter 11, for further reading.

reactions may have the same reaction vector and so we can combine propensities for such to get the overall intensity for the jump of the Markov chain for each element of $\mathcal{V}$:

$$(4) \quad \Upsilon_j(x) := \sum_{r: \bar{v}_r = v_j} \Lambda_r(x), \quad j = 1, \dots, n, \quad x \in \mathcal{X}.$$

The infinitesimal generator for X is then given by

$$(5) \quad Q_{x,y} = \begin{cases} \Upsilon_j(x) & \text{if } y - x = v_j \text{ for some } 1 \leq j \leq n, \\ -\sum_{j=1}^n \Upsilon_j(x) & \text{if } y = x, \\ 0 & \text{otherwise.} \end{cases}$$

A representation² for X , introduced by Kurtz for density dependent Markov chains [16], which is particularly useful for establishing strong approximations, is the following³

$$(6) \quad X(t) = X(0) + \sum_{j=1}^n v_j N_j \left(\int_0^t \Upsilon_j(X(s)) ds \right), \quad t \geq 0,$$

where $N_1, N_2, \dots, N_n$ are independent, unit rate Poisson processes. This *random time change representation* has become one of the central tools in modern theory and computation for stochastic chemical reaction networks. Beyond its role in establishing rigorous approximation theorems, it underlies stochastic simulation algorithms (including those for hybrid systems), pathwise coupling constructions, and statistical analysis of reaction network time series. For proving weak approximations to SCRNs (and other Markov processes), Kurtz developed a general treatment, using martingale problems to characterize and prove convergence⁴. In particular, the SCRN X can be characterized in law using its initial distribution and the fact that for any bounded function f on $\mathcal{X}$ with compact support,

$$(7) \quad f(X(t)) - \int_0^t Qf(X(s)) ds, \quad t \geq 0,$$

is a martingale with respect to the filtration generated by X .

3. Classical Scaling. For chemical reactions, typically the propensity functions depend on a measure of the volume of reacting species, which we denote by V , and the quantity of interest is the normalized volume fraction (or concentration) for each species, $\bar{x}^V = x/V$. Accordingly, it is frequently assumed that the overall intensity function Υ_j^V for change vector v_j , given by (4), scales with the volume V such that

$$(8) \quad \frac{1}{V} \Upsilon_j^V(V\bar{x}^V) = \lambda_j(\bar{x}^V) + \frac{1}{V} \epsilon_j^V(\bar{x}^V),$$

²For simplicity, here we assume that X does not explode in finite time.

³Volkonskii [25] introduced the idea of representing a Markov process as a random time change of another Markov process. Helms [8] generalized this to multiple random time changes. In proving strong approximations for density dependent Markov chains, Kurtz [16] represented the Markov chains as random time changes of Poisson processes. Subsequently, Kurtz [18] developed an extensive theory for multiparameter random time changes of Markov processes, and explored the relationship between solutions of random time change problems and martingale problems.

⁴In very early work, Kurtz [12] used operator semigroups to characterize and prove weak (distributional) convergence of Markov processes. However, following the development by Stroock and Varadhan [22, 23] of the theory of martingale problems for characterizing diffusion processes, Kurtz transitioned to increasingly using martingale characterizations and ultimately developed a general treatment for Markov processes (see for example, Chapter 4 of Ethier and Kurtz [7]).

where on each compact set, $\lambda_j(\cdot)$ is uniformly Lipschitz continuous, and $\{\epsilon_j^V(\cdot), V \in \mathfrak{V}\}$ is uniformly bounded for a given countable increasing set of volumes, $\mathfrak{V}$, that tends to infinity. In particular, this holds for the classical scaling frequently assumed for mass action kinetics, where the proportionality constant in (2) satisfies $\kappa_r^V = k_r V^{1-|v_r^-|}$ and k_r does not depend on V . Adjoining a superscript V to indicate the dependence of the process X on V , the concentration process is defined by $\bar{X}^V = \frac{1}{V} X^V$ and under (8) can be represented as

$$(9) \quad \bar{X}^V(t) = \bar{X}^V(0) + \frac{1}{V} \sum_{j=1}^n v_j N_j \left(V \int_0^t \left(\lambda_j(\bar{X}^V(s)) + \frac{1}{V} \epsilon_j^V(\bar{X}^V(s)) \right) ds \right), \quad t \geq 0.$$

Then, $\{\bar{X}^V, V \in \mathfrak{V}\}$ is a **density dependent family**⁵ of Markov chains in the sense of [17]⁶.

4. Approximations. Consider a sequence of concentration processes $\{\bar{X}^V, V \in \mathfrak{V}\}$ associated with SCRN satisfying (8). We can assume these are all defined on a single probability space. Suppose that the state spaces $\bar{\mathcal{X}}^V = \frac{1}{V} \mathcal{X}^V$ for $\bar{X}^V$ are all contained in a set $\bar{\mathcal{X}} \subset \mathbb{R}_+^d$ and the $\lambda_j, j = 1, \dots, n$ are defined and locally Lipschitz continuous there. To keep the exposition simple, we will assume that $\bar{\mathcal{X}}$ is compact and so the λ_j are uniformly Lipschitz continuous there⁷. Let $\mu, \sigma : \bar{\mathcal{X}} \rightarrow \mathbb{R}^d$ be defined for $x \in \bar{\mathcal{X}}$ by

$$(10) \quad \mu(x) := \sum_{j=1}^n v_j \lambda_j(x), \quad \sigma(x) := \left(\sum_{j=1}^n \lambda_j(x) v_j v_j^\top \right)^{1/2},$$

where the superscript of $\top$ denotes transpose and the square root is the unique non-negative definite matrix square root. For $T > 0$ and any right continuous function $x : [0, T] \rightarrow \mathbb{R}^d$ let $\|x\|_T = \sup_{t \in [0, T]} |x(t)|$.

4.1. Fluid Limit (Functional Law of Large Numbers). Given $x_0 \in \bar{\mathcal{X}}$, by the uniform Lipschitz continuity of μ (inherited from the λ_j), there is a unique solution $\bar{X}$ of the following differential equation (usually called the reaction rate equation) with initial condition x_0 :

$$(11) \quad \frac{d\bar{X}(t)}{dt} = \mu(\bar{X}(t)), \quad \bar{X}(0) = x_0.$$

The following theorem provides a functional law of large numbers limit (often also called a fluid or hydrodynamic limit) for $\bar{X}^V$ as $V \rightarrow \infty$. In [13], Kurtz proved a version of this with convergence in probability. Theorem 2.2 in [17] implies the result stated below, which has almost sure convergence.

⁵In Kurtz's 1976 paper [16], density dependent Markov chains have $\epsilon_j^V \equiv 0$. However, in [17], he allows for a non-zero perturbation $\frac{1}{V} \epsilon_j^V$ as in (8), that vanishes as $V \rightarrow \infty$. Some authors use the term *nearly* density dependent for such perturbations, keeping density dependent for the stricter sense of [16]. The generalization allowed by (8) is important for chemical reaction networks, since reactions in which more than one copy of a given species is consumed only satisfy (8) with $\epsilon_j^V \neq 0$.

⁶The setup in [16, 17] for density dependent Markov chains allows for the equivalent of countably many reactions, under some additional conditions. The focus here on finitely many reactions simplifies the exposition and is natural for chemical reaction networks.

⁷For the approximations of SCRN developed by Kurtz in [15, 16, 17] and also reported in Chapter 11 of [7], the λ_j are assumed to be uniformly Lipschitz continuous, or locally Lipschitz continuous with at most linear growth as $|x| \rightarrow \infty$ for $x \in \bar{\mathcal{X}}$. Such assumptions are typically satisfied for chemical reaction networks when $\bar{\mathcal{X}}$ is compact. However, if $\bar{\mathcal{X}}$ is unbounded, they are a little too stringent. For example, for the simple reaction system in (3) with mass action kinetics, for $x \in \bar{\mathcal{X}} = \mathbb{R}_+^d$, $\lambda_2(x) = k_2 x_1^2$ and $\lambda_3(x) = k_3 x_1 x_2$ grow quadratically as $|x| \rightarrow \infty$. Nevertheless, by introducing stopping times $\tau_K = \inf\{t \geq 0 : |\bar{X}^V(t)| \geq K\}$, similar results to those stated here can be seen to hold for the stopped processes $\bar{X}^V(\cdot \wedge \tau_K)$ for each $K > 0$ when $\bar{\mathcal{X}}$ is not compact and the λ_j are locally Lipschitz continuous.

THEOREM 1. *Suppose that almost surely, $\lim_{V \rightarrow \infty} \bar{X}^V(0) = x_0 \in \bar{\mathcal{X}}$. Then, for each $T > 0$, we have that almost surely*

$$(12) \quad \lim_{V \rightarrow \infty} \|\bar{X}^V(\cdot) - \bar{X}(\cdot)\|_T = 0.$$

4.2. Linear Noise Approximation (Functional Central Limit Theorem). It is natural to ask about the fluctuations of $\bar{X}^V$ about the deterministic fluid limit $\bar{X}$. Using a power series expansion in the Kolmogorov forward equation (also called the chemical master equation), van Kampen [24] developed an approximation for the distribution at time t of such fluctuations, which are of order $V^{-1/2}$. Kurtz [14] developed a full process level approximation for $V^{1/2}(\bar{X}^V(\cdot) - \bar{X}(\cdot))$, which is a diffusion process with coefficients driven by $\bar{X}$. This is a functional central limit theorem, often called the Linear Noise Approximation (LNA).

Let x_0 and $\bar{X}$ be as in (11). For each $V \in \mathfrak{V}$ and $t \geq 0$, let

$$(13) \quad Y^V(t) = \sqrt{V} (\bar{X}^V(t) - \bar{X}(t)).$$

Assuming that μ has continuous first-order partial derivatives on $\bar{\mathcal{X}}$, let Y denote the diffusion process satisfying

$$(14) \quad Y(t) = \int_0^t \partial\mu(\bar{X}(s))Y(s)ds + \int_0^t \sigma(\bar{X}(s))dW(s), \quad t \geq 0,$$

where for $x \in \bar{\mathcal{X}}$, $\partial\mu(x)$ is the matrix whose ik entry is $\frac{\partial\mu_i(x)}{\partial x_k}$, and W is a standard d -dimensional Brownian motion. The well posedness of (14) comes from the fact that it is linear in Y and $\bar{X}$ is already determined. An equivalent distributional version of Y can be obtained by solving the equation:

$$(15) \quad Y(t) = \int_0^t \partial\mu(\bar{X}(s))Y(s)ds + \sum_{j=1}^n v_j \int_0^t \sqrt{\lambda_j(\bar{X}(s))}dW_j(s),$$

where the W_j , $j = 1, \dots, n$ are independent standard one-dimensional Brownian motions.

The processes Y^V and Y are stochastic processes taking values in $D([0, \infty), \mathbb{R}^d)$, the space of right continuous paths having finite left limits in $\mathbb{R}^d$, and Y in fact has continuous paths. We endow this space with the usual Skorokhod J_1 -topology. Convergence in distribution of Y^V to Y as $V \rightarrow \infty$ means weak convergence of the associated probability laws induced on $D([0, \infty), \mathbb{R}^d)$. By Theorem 3.1 in [14] we have the following.

THEOREM 2. *Suppose that μ has continuous first-order partial derivatives on $\bar{\mathcal{X}}$ and assume that $\sqrt{V}(\bar{X}^V(0) - x_0)$ converges in distribution to the zero vector. Then, Y^V converges in distribution to Y .*

This theorem yields the linear noise diffusion approximation:

$$(16) \quad \bar{X}^V(t) \approx \bar{X}(t) + \frac{1}{\sqrt{V}}Y(t), \quad t \geq 0.$$

4.3. Langevin Diffusion Approximation: Strong Approximation. The LNA is frequently used to study local fluctuations about a steady-state of the fluid model. However, since it linearizes fluctuations, it typically fails to capture global nonlinear behavior well. Also, the right hand side of (16) is an unconstrained diffusion process and can typically take negative values, which violates the natural nonnegativity of chemical concentrations. An alternative

diffusion approximation, which is more suited to capturing nonlinearities is the Langevin approximation, Z^V for $\bar{X}^V$, which solves the (chemical) Langevin equation:

$$(17) \quad Z^V(t) = \bar{X}^V(0) + \int_0^t \mu(Z^V(s))ds + \frac{1}{\sqrt{V}} \sum_{j=1}^n v_j \int_0^t \sqrt{\lambda_j(Z^V(s))} dW_j(s),$$

where W_j , $j = 1, \dots, n$ are independent standard one-dimensional Brownian motions. Note here that rather than the drift and dispersion coefficients depending on $\bar{X}$, as in the linear noise approximation, the coefficients here depend on the approximation itself, Z^V . Indeed, equation (17) is usually only well posed until the first time some component of Z^V reaches zero (i.e., the boundary of the positive orthant in $\mathbb{R}_+^d$ is reached). This can be seen even with a simple example such as the following.

Example. Consider a system with a reversible reaction involving two species: $S_1 \xrightleftharpoons[k_2]{k_1} S_2$, under mass action kinetics and classical scaling. The fluid approximation $\bar{X}$ satisfies

$$\frac{d\bar{X}_1}{dt} = k_2\bar{X}_2(t) - k_1\bar{X}_1(t), \quad \frac{d\bar{X}_2}{dt} = k_1\bar{X}_1(t) - k_2\bar{X}_2(t).$$

The linear noise approximation is given by $\bar{X}^V \approx \bar{X} + \frac{1}{\sqrt{V}}Y$, where in differential form,

$$dY_2(t) = -dY_1(t) = k_1Y_1(t) - k_2Y_2(t) + \sqrt{k_1\bar{X}_1(t)}dW_1(t) - \sqrt{k_2\bar{X}_2(t)}dW_2(t).$$

The Langevin approximation is given by $\bar{X}^V(\cdot) \approx Z^V(\cdot)$, where

$$(18) \quad dZ_2^V(t) = -dZ_1^V(t) \\ = (k_1Z_1^V(t) - k_2Z_2^V(t))dt + \frac{1}{\sqrt{V}} \left(\sqrt{k_1Z_1^V(t)}dW_1(t) - \sqrt{k_2Z_2^V(t)}dW_2(t) \right).$$

As can be seen in (18), just after Z_2^V becomes zero, the stochastic term involving dW_1 for Z_2^V would tend to fluctuate both positively and negatively. However, if Z_2^V were to become negative, then the square root of Z_2^V is not well defined (in the last term in (18)). A similar issue occurs when Z_1^V becomes zero. Consequently, solutions of (18) are only well defined until the first time Z^V reaches the boundary of the positive two-dimensional quadrant.

In the mid 1970s, Kórmlos, Major and Tusnády [11] developed theory which enabled one to build on the same probability space, a Brownian motion and a Poisson process such that the pathwise distance between them over compact time intervals is of order $\log T$, where $T \geq 2$ is the length of the time interval. Building on this theory, Kurtz [17] developed strong (pathwise) diffusion approximations of the form (17) for density dependent Markov chains. This theory assumes, besides non-negativity and uniform Lipschitz continuity of the λ_j , that $\sqrt{\lambda_j}$ is uniformly Lipschitz continuous for each j , to ensure existence and uniqueness of solutions of (17). As can be seen in the above example, the last assumption usually breaks down for SCRN with mass action kinetics as the boundary of the orthant is approached, where only Hölder continuity holds in general. Nevertheless, with mass action kinetics in a compact set, a unique solution of (17) can be constructed up until the first time that the solution reaches the boundary of the positive orthant and the theory of Kurtz can be applied to obtain a strong pathwise approximation up until that time. The next theorem follows from applying Theorem 3.3 in [17] with stopping times. Here we assume that Z^V is a solution of (17) up until the first time τ that it reaches the boundary of the positive orthant (for concreteness, we hold Z^V constant thereafter) and λ_j is uniformly Lipschitz continuous wherever Z^V and $\bar{X}^V$ live. Note that in contrast to the limit results in Theorems 1 and 2, the next theorem is an approximation result for a finite V . If one were to take the limit as $V \rightarrow \infty$, $\bar{X}^V$ and Z^V both converge to the deterministic limit $\bar{X}$.

THEOREM 3. *For each $T > 0$, $V \geq 2$, there is a random variable γ_T^V whose distribution is independent of V such that $E[\exp(\eta\gamma_T^V)] < \infty$ for some $\eta > 0$, and*

$$(19) \quad \|\bar{X}^V(\cdot) - Z^V(\cdot)\|_{\tau \wedge T} \leq \gamma_T^V \frac{\log V}{V}.$$

Given that the LNA and the Langevin approximation are both diffusion approximations to the same chemical reaction network, one might wonder about the relationship between the two. From Section 4 of [17], under further conditions, including that μ has bounded first and second partial derivatives, it can be seen that an estimate like (19) holds with $\bar{X} + \frac{1}{\sqrt{V}}Y$ in place of $\bar{X}^V$, i.e., the error estimate for the difference between the approximations is of the same order as the error between $\bar{X}^V$ and Z^V . Nevertheless, as mentioned previously, the Langevin approximation is usually better at capturing nonlinearities in the dynamics.

Various mechanisms have been suggested for continuing the Langevin approximation beyond time τ . For example, replacing λ_j by $|\lambda_j|$ in (17), although such a diffusion, if well defined, will typically stray outside of the positive orthant. A natural extension that lives in the orthant is a jump-diffusion, which behaves like the Langevin approximation in the interior of the orthant and jumps back into the orthant when the boundary is reached. Such jump diffusions can be simulated, however, their analysis and comparison with the original Markov chain is made difficult by the presence of the boundary jumps. Leite and Williams [21] showed that a reflected diffusion process (called a Constrained Langevin Approximation—CLA) can be obtained as a limit in distribution of such jump diffusions with jump sizes shrinking to zero and proved that the diffusion is well posed, assuming there are inflows and outflows for each species (although their rates can be arbitrarily small). Campos and Williams [6] provided error estimates of order $\log V/V$ for a strong approximation between such a CLA and a nearly density dependent Markov chain in one-dimension. The CLA in d -dimensions is a solution $\tilde{Z}^V$ of the stochastic differential equation:

$$\tilde{Z}^V(t) = \bar{X}^V(0) + \int_0^t \mu(\tilde{Z}^V(s))ds + \frac{1}{\sqrt{V}} \int_0^t \sigma(\tilde{Z}^V(s))dW(s) + \frac{1}{\sqrt{V}} \int_0^t \gamma(\tilde{Z}^V(s))dL^V(s),$$

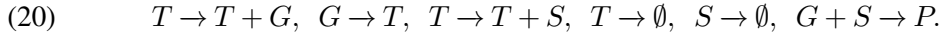
where W is a standard d -dimensional Brownian motion and L^V is a continuous, non-decreasing, one-dimensional (local time) process which only increases when $\tilde{Z}^V$ is on the boundary of the orthant and $\gamma(z) = \mu(z)/|\mu(z)|$, for z on the boundary of the orthant.

5. Multiscaling and Biochemical Networks. In 2003–04, the NSF Institute for Mathematics and its Applications (IMA) at the University of Minnesota held a program on “Probability and Statistics in Complex Systems: Genomics, Networks and Financial Engineering.” Tom Kurtz was a key member of the scientific program committee and participated extensively in the year-long program. Inspired by the growing use of mathematical models for biochemical reaction networks in systems and synthetic biology, a working group self-organized to study this topic. Tom Kurtz and some other longer term participants in the IMA program were members. The group considered various models, including deterministic chemical reaction networks and related algebraic structures, and continuous time Markov chain stochastic models. In many models of biological interest, some species are only present in small numbers, whilst others are in abundance, and reaction rate constants can vary by several orders of magnitude. This spurred interest within the working group in developing approximations for the complex stochastic models that took multiple time and size scales into account, and reduced the dimensionality of the microscopic models. Using an example of intracellular viral infection as a test bed, a subgroup, consisting of Karen Ball, Tom Kurtz, Lea Popovic and Greg Rempała, set out to develop a multiscale approximation, which ultimately resulted in

the paper [5]. Subsequently, Kang and Kurtz [9] developed a systematic approach to separating time-scales and proved associated limit theorems for SCRN. Their analysis involved functional laws of large numbers, as well as averaging theorems such as those in [20], relating ergodic behavior of fast components to the dynamics of slower components. The results in [9] are in a sense a generalization of the fluid limit results in §4.1 to the more complex situation of SCRN with multiple scalings of time, reaction rates and species count normalizations. In a later paper, in a multiscaling generalization of some aspects of §4.2 and §4.3, Kang, Kurtz and Popovic [10] developed a systematic approach to proving functional central limit theorems and diffusion approximations that capture fluctuations around deterministic limits developed in [9].

In [5, 9, 10], there is a scaling parameter⁸, V . However, it no longer always represents a volume measure of the system. Rather, it is a large parameter suitable for the approximation, e.g., V might be the order of magnitude of the most abundant species in the system. The components of the species count process X^V are normalized by powers of V , that could depend on the individual components, such that $\bar{X}_i^V = V^{-\alpha_i} X_i^V$, where $\alpha_i \geq 0$ is such that $\bar{X}_i^V$ is of order one, for $i = 1, \dots, d$. In addition to this multiscaling in size of components, another set of exponents $\{\beta_r, r \in \mathcal{R}\} \subset \mathbb{R}$ are introduced to take into account possible large variations in the reaction rate constants from reaction to reaction, e.g., with mass action kinetics, $\kappa_r^V = k_r V^{\beta_r}$ so that the renormalized rate constants, k_r , are of order one. Finally, different time scales are considered, i.e., the time in $\bar{X}^V$ is scaled by a factor V^γ for at least two different values of $\gamma \in \mathbb{R}$. Below, the key example from [5] (which was also studied in [10]) is briefly summarized, to illustrate how multiscaling can arise and be studied.

The reaction system for the intracellular viral infection model studied in [5, 10] can be written as follows. Here T , G and S denote the viral nucleic acids — classified as template or genome, and the viral structural protein, respectively, and P denotes the viral product.



The reactions are numbered sequentially as $1, 2, \dots, 6$, from left to right, and the associated rate constants with mass action kinetics are κ_i , $i = 1, 2, \dots, 6$. Let $X_1(t), X_2(t), X_3(t)$ denote the number of molecules of species T, G, S , respectively, in the system at time t . Using reaction rate constants gleaned from the literature, in [5, 10] the following scaling parameters⁹ are used: $\alpha_1 = 0, \alpha_2 = 2/3, \alpha_3 = 1, \beta_1 = 0, \beta_2 = -2/3, \beta_3 = 1, \beta_4 = 0, \beta_5 = 0, \beta_6 = -5/3, k_1 = 1, k_2 = 2.5, k_3 = 1, k_4 = 0.25, k_5 = 2, k_6 = 0.75$. Two time scales are considered, the original one and scaling time by V^γ with $\gamma = 2/3$. With the latter time scaling, $\bar{X}_i^V(\cdot) = V^{-\alpha_i} X_i^V(V^\gamma \cdot)$, $i = 1, 2, 3$, can be represented by

$$\begin{aligned} \bar{X}_1^V(t) &= \bar{X}_1^V(0) + N_2 \left(\int_0^t V^{2/3} 2.5 \bar{X}_2^V(s) ds \right) - N_4 \left(\int_0^t V^{2/3} 0.25 \bar{X}_1^V(s) ds \right), \\ \bar{X}_2^V(t) &= \bar{X}_2^V(0) + V^{-2/3} N_1 \left(\int_0^t V^{2/3} \bar{X}_1^V(s) ds \right) - V^{-2/3} N_2 \left(\int_0^t V^{2/3} 2.5 \bar{X}_2^V(s) ds \right) \\ &\quad - V^{-2/3} N_6 \left(\int_0^t V^{2/3} 0.75 \bar{X}_2^V(s) \bar{X}_3^V(s) ds \right), \\ \bar{X}_3^V(t) &= \bar{X}_3^V(0) + V^{-1} N_3 \left(\int_0^t V^{5/3} \bar{X}_1^V(s) ds \right) - V^{-1} N_5 \left(\int_0^t V^{5/3} 2 \bar{X}_3^V(s) ds \right) \\ &\quad - V^{-1} N_6 \left(\int_0^t V^{2/3} 0.75 \bar{X}_2^V(s) \bar{X}_3^V(s) ds \right), \end{aligned}$$

⁸Called N in [5, 9, 10], but denoted by V here.

⁹As pointed out in [5], suitable scaling exponents need not be unique and determining appropriate ones is part art and part science (based on experimental measurements of rate constants, and experimental or simulated observations of system behavior).



FIG 1. Tom Kurtz introduced Ruth Williams' Kolmogorov Lecture on Reflected Diffusions and Biochemical Reaction Networks at the World Congress in Probability and Statistics, Toronto, 2016. Photo Credit: Ankit Gupta.

where $N_1, N_2, N_3, N_4, N_5, N_6$ are independent, unit rate Poisson processes. Assuming that $\lim_{V \rightarrow \infty} \bar{X}_2^V(0) = \bar{X}_2(0) \in (0, \infty)$ a.s., intuitively, one can see that the components $\bar{X}_1^V, \bar{X}_3^V$ are evolving on a faster time scale than $\bar{X}_2^V$. It is shown in [5, 10] that the effect of $(\bar{X}_1^V, \bar{X}_3^V)$ on the behavior of $\bar{X}_2^V$ is through stochastic averaging. In particular, $\int_0^t \bar{X}_3^V(s) ds \approx \frac{1}{2} \int_0^t \bar{X}_1^V(s) ds \approx 5 \int_0^t \bar{X}_2^V(s) ds$ and $\int_0^t \bar{X}_2^V(s) \bar{X}_3^V(s) ds \approx 5 \int_0^t (\bar{X}_2^V(s))^2 ds$. Let $\bar{X}_2$ satisfy

$$(21) \quad \bar{X}_2(t) = \bar{X}_2(0) + \int_0^t (7.5\bar{X}_2(s) - 3.75(\bar{X}_2(s))^2) ds,$$

and assuming $U(0)$ is a limit in distribution of $V^{1/3}(\bar{X}_2^V(0) - \bar{X}_2(0))$, let U satisfy

$$\begin{aligned} U(t) = U(0) + \int_0^t 7.5(1 - \bar{X}_2(s))U(s) ds \\ + \int_0^t \sqrt{72.5\bar{X}_2(s) - 48.75(\bar{X}_2(s))^2 + 11.25(\bar{X}_2(s))^3} dW(s), \end{aligned}$$

where W is a standard one-dimensional Brownian motion, independent of $U(0)$. Then, it is shown in [5, 10] that for each $T \geq 0$, as $V \rightarrow \infty$,

$$(22) \quad \|\bar{X}_2^V(\cdot) - \bar{X}_2(\cdot)\|_T \Rightarrow 0 \quad \text{and} \quad V^{1/3}(\bar{X}_2^V - \bar{X}_2) \Rightarrow U,$$

where $\Rightarrow$ denotes convergence in distribution. The interested reader is referred to [5, 10] for the many technical details and auxiliary results omitted here. Those works also have other examples and with [9] provide extensive general tools for establishing various multiscaling limits and approximations for SCRNs. Tom Kurtz has other works on analysis of SCRNs not discussed here. For a taste of these, the reader is referred to [1, 2, 3].

6. Afterword. On a personal note, I am grateful to Tom Kurtz for his wonderful mathematical contributions, leadership in the mathematics and probability communities, and friendship over the years.

His books and papers have been invaluable to me and my collaborators. We have followed behind his giant footsteps, utilizing his work to prove limit theorems and investigate properties of reflected diffusion and other approximations for complex stochastic networks.

Tom's professional leadership, especially in the probability community has been priceless. His deep commitment to the advancement of probability as a field, his caring support for probabilists, especially junior researchers, and his willingness to step up to help whenever needed, were signature traits.

Most of all, I valued his kind and caring attitude and friendship over the years. I still remember vividly the first time I went to the Midwest Probability Colloquium (while still a Ph.D. student at Stanford) and how he took special care to make sure I felt included.

Tom is deeply missed, but fortunately his mathematics and example carry on.

REFERENCES

- [1] Anderson, D. F., Cappelletti, D. and Kurtz, T. G. (2017). Finite time distributions of stochastically modeled chemical systems with absolute concentration robustness. *SIAM J. Appl. Dynam. Syst.* **16**, 1309–1339.
- [2] Anderson, D. F., Cappelletti, D., Kurtz, T. G. and Koyama, M. (2018). Non-explosivity of stochastically modeled reaction networks that are complex balanced. *Bull. Math. Biol.* **100**, 2561–2579.
- [3] Anderson, D. F., Craciun, G. and Kurtz, T. G. (2010). Product-form stationary distributions for deficiency zero chemical reaction networks. *Bull. Math. Bio.* **72**, 1947–1970.
- [4] Anderson, D. F. and Kurtz, T. G. (2015). *Stochastic Analysis of Biochemical Systems*. Math. Biosci. Inst. Lect. Series, Stochastics in Biological Systems, Springer, Cham; MBI Math. Biosc. Inst., Ohio State University, Columbus, OH, 2015.
- [5] Ball, K., Kurtz, T. G., Popovic, L. and Rempała, G. (2006). Asymptotic analysis of multiscale approximations to reaction networks. *Ann. Appl. Probab.* **16**, 1925–1961.
- [6] Campos, F. P., and Williams, R. J. (2024). Error bounds for one-dimensional constrained Langevin approximations for nearly density dependent Markov chains. *Adv. Appl. Prob.* **56**, 825–867.
- [7] Ethier, S. N. and Kurtz, T. G. (1986). *Markov Processes: Characterization and Convergence*. Wiley, N.Y.
- [8] Helms, L. L. (1974). Ergodic properties of several interacting Poisson particles. *Adv. Math.* **12**, 32–57.
- [9] Kang, H.-W. and Kurtz, T. G. (2013). Separation of time-scales and model reduction for stochastic reaction networks. *Ann. Appl. Probab.* **23**, 529–583.
- [10] Kang, H.-W., Kurtz, T. G. and Popovic, L. (2014). Central limit theorems and diffusion approximations for multiscale Markov chain models. *Ann. Appl. Probab.* **24**, 721–759.
- [11] Komlós, J., Major, P. and Tusnády, G. (1975). An approximation of partial sums of independent random variables and the sample distribution function, I. *Z. Wahrsch. Verw. Gebiete*. **32**, 111–131.
- [12] Kurtz, T. G. (1967). *Convergence of Operator Semigroups with Applications to Markov Processes*. Ph.D. dissertation, Stanford University.
- [13] Kurtz, T. G. (1970). Solutions of ordinary differential equations as limits of pure jump Markov processes. *J. Appl. Probab.* **7**, 49–58.
- [14] Kurtz, T. G. (1971). Limit theorems for sequences of jump Markov processes approximating ordinary differential processes. *J. Appl. Probab.* **8**, 344–356.
- [15] Kurtz, T. G. (1972). The relationship between stochastic and deterministic models for chemical reactions. *J. Chem. Phys.* **57**, 2976–2978.
- [16] Kurtz, T.G. (1976). Limit theorems and diffusion approximations for density dependent Markov chains. In: Wets, R.J.B. (eds) *Stochastic Systems: Modeling, Identification and Optimization, I. Mathematical Programming Studies*, vol 5. Springer, Berlin-Heidelberg. <https://doi.org/10.1007/BFb0120765>.
- [17] Kurtz, T. G. (1978). Strong approximation theorems for density dependent Markov chains. *Stochastic Processes and their Applications*. **6**, 223–240.
- [18] Kurtz, T. G. (1980). Representations of Markov processes as multiparameter time changes. *Ann. Prob.* **8**, 682–715.
- [19] Kurtz, T. G. (1981). *Approximation of Population Processes*. CBMS-NSF Regional Conference Series in Applied Mathematics, SIAM, Philadelphia.
- [20] Kurtz, T. G. (1992). Averaging for martingale problems and stochastic approximation. In *Applied Stochastic Analysis. Lecture Notes in Control and Inform. Sci.* **77**, Springer, pp. 186–209.
- [21] Leite, S. C. and Williams, R. J. (2019). A constrained Langevin approximation for chemical reaction networks. *Ann. Appl. Probab.* **29**, 1541–1608.
- [22] Stroock, D. W., and Varadhan, S. R. S. (1969). Diffusion processes with continuous coefficients, I. *Comm. Pure Appl. Math.* **22**, 345–400.
- [23] Stroock, D. W., and Varadhan, S. R. S. (1969). Diffusion processes with continuous coefficients, II. *Comm. Pure Appl. Math.* **22**, 479–530.
- [24] Van Kampen, N. G. (1961). A power series expansion of the master equation. *Canadian Journal of Physics*. **39**, 551.
- [25] Volkonskii V. A. (1958). Random substitution of time in strong Markov processes. *Theor. Probability Appl.* **3**, 310–326.