

Received February 1, 2013, accepted September 2, 2013, date of publication September 16, 2013, date of current version September 24, 2013.

Digital Object Identifier 10.1109/ACCESS.2013.2281668

Approximating Extremely Large Networks via Continuum Limits

YANG ZHANG¹, EDWIN K. P. CHONG¹ (Fellow, IEEE), JAN HANNIG², AND DONALD ESTEP³

¹Department of Electrical and Computer Engineering, Colorado State University, Ft. Collins, CO 80523-1373, USA

²Department of Statistics and Operation Research, The University of North Carolina at Chapel Hill, Chapel Hill, NC 27599-3260, USA

³Department of Mathematics and Department of Statistics, Colorado State University, Fort Collins, CO 80523-1373, USA

Corresponding author: Y. Zhang (yang.zhang@colostate.edu)

The work of Y. Zhang and E. Chong was supported in part by the National Science Foundation (NSF) under Grant ECCS-0700559 and ONR under Grant N00014-08-1-110. The work of J. Hannig was supported in part by NSF under Grants 1007543 and 1016441. The work of D. Estep was supported in part by the Defense Threat Reduction Agency under Grant HDTRA1-09-1-0036, the Department of Energy under Grants DE-FG02-04ER25620, DE-FG02-05ER25699, DE-FC02-07ER54909, DE-SC0001724, DE-SC0005304, and INL00120133, the Lawrence Livermore National Laboratory under Grants B573139, B584647, B590495, the National Aeronautics and Space Administration under Grant NNG04GH63G, the National Institutes of Health under Grant 5R01GM096192-02, the National Science Foundation under Grants DMS-0107832, DMS-0715135, DGE-0221595003, MSPA-CSE-0434354, ECCS-0700559, DMS-1016268, and DMS-FRG-1065046, and the Idaho National Laboratory under Grants 00069249 and 00115474.

ABSTRACT This paper is concerned with modeling of networks with an extremely large number of components using partial differential equations (PDEs). This modeling method is based on the convergence of a sequence of underlying Markov chains of the network indexed by N , the number of components in the network. As N goes to infinity, the sequence converges to a continuum limit, which is the solution of a certain PDE. We provide sufficient conditions for the convergence and characterize the rate of convergence. As an application, we model large wireless sensor networks by PDEs. While traditional Monte Carlo simulation for extremely large networks is practically infeasible, PDEs can be solved with reasonable computation overhead using well-established mathematical tools.

INDEX TERMS Modeling, partial differential equations, Markov processes, network modeling.

I. INTRODUCTION

This paper is concerned with modeling of networks involving an extremely large number of components. The conventional way to study large networks is by computer modeling and simulation [1]. The approach involves representing the network in computer software and then applying a numerical simulation method to study how the network behaves. Typically, each individual component is explicitly represented as a separate entity. As we are confronted with larger and larger networks, the number of its components that have to be represented increases, and this significantly lengthens the time it takes to write, manage, and run computer simulation programs. Simulating large networks typically requires expensive, highly sophisticated supercomputers involving large parallel computing hardware with specialized software. It is not uncommon for a simulation run to take days or weeks, even on a large supercomputer. The larger the network, the longer it takes. The computational overhead

associated with direct simulation thus severely limits the size and complexity of networks that can be studied in this fashion.

Our recent papers [2]–[5] address this problem by using *continuum modeling* to capture the global characteristics of large networks. In large networks, we are often more interested in the global characteristics of an entire network than in a particular individual component. Continuum models do away with the need to represent each individual component of a large network as a separate entity, and consider the behavior of the components on the scale of the aggregate rather than of the individual. Similar to treating water as a continuous fluid instead of a large number of individual molecules, continuum modeling treats the large number of communicating components (or nodes) in a network collectively as a continuum. The continuum modeling strategies in [3]–[5] use partial differential equations (PDEs) to approximate large sensor or cellular networks modeled by a certain class of Markov chains. The PDE model represents the global characteristics of the net-

work, while the individual characteristics of the components enter the model through the form and the parameters of the PDE.

PDEs are well-suited to the modeling of continuum behavior. Although uncommon in modeling networks, they are common in modeling many physical phenomena, including heat, sound, electromagnetism, and fluid flow. There are well-established mathematical tools to solve PDEs, such as the finite element method [6] and the finite difference method [7], incorporated into computer software packages such as Matlab and Comsol. We can use these tools to greatly reduce computation time. As a result, the effort to run the PDE models in a computer no longer suffers from the curse of sheer size. (In fact, as we will show, the larger the network, the closer the PDE approximates it.) Continuum modeling thus provides a powerful way to deal with the number of components in large networks. This, in turn, would make it possible to carry out—with reasonable computational burden even for extremely large systems—network performance evaluation and prototyping, network design, systematic parameter studies, and optimization of network characteristics.

The work in this paper is motivated by the continuum modeling strategies in the papers [3]–[5] mentioned above, and by the need for a rigorous description of the heuristic limiting process underlying the construction of their PDE models. We analyze the convergence of a class of Markov chains to their continuum limits, which are the solutions of certain PDEs. We consider a general Markov chain model in an abstract setting instead of that of any particular network model. We do this for two reasons: first, our network modeling results involve a class of Markov chains modeling a variety of communication networks; second, similar Markov chain models akin to ours arise in several other contexts. For example, a very recent paper [8] on human crowd modeling derives a limiting PDE in a fashion similar to our approach.

In the convergence analysis, we show that a sequence of Markov chains indexed by N , the number of components in the system that they model, converges in a certain sense to its continuum limit, which is the solution of a time-dependent PDE, as N goes to ∞ . The PDE solution describes the global spatio-temporal behavior of the model in the limit of large system size. We apply this abstract result to the modeling of a large wireless sensor network by approximating a particular global aspect of the network states (queue length) by a nonlinear convection-diffusion-reaction PDE. This network model includes the network example discussed in [3] as a special case.

A. RELATED LITERATURE

The modeling and analysis of stochastic systems such as networks is a large field of research, and much of the previous contributions share goals with the work in this paper.

In the field of direct numerical simulation approaches, many efforts have been made to accelerate the simulation. For example, parallel simulation techniques have been developed to exploit the computation power of multiprocessor and/or

cluster platforms [9]–[12]; new mechanisms for executing the simulation have been designed to improve the efficiency of event scheduling in event-driven simulations (see, e.g., [13], [14]); and fluid simulations, in contrast to traditional packet-level ones, have been used to simplify the network model by treating network traffic (*not* nodes) as continuous flows rather than discrete packets [15]–[18]. However, as the number of nodes in the network grows extremely large, computer-based simulations involving individual nodes eventually become practically infeasible. For the remainder of this subsection, we review some existing results on analysis of stochastic networks that do *not* depend on direct numerical simulation.

Our convergence analysis in this paper uses Kushner's ordinary differential equation (ODE) method [19]. This method essentially studies a “smoothing” limit as a certain “averaging” parameter goes to ∞ , but not a “large-system” limit as the number of components in the system goes to ∞ . In contrast, the limiting process analyzed in this paper involves two steps: the first similar to that in Kushner's ODE method, and the second a “large-system” limit. (We provide more details about the two-step procedure later in Sec. I-D.) In other words, while Kushner's method deals with a fixed state space, we treat a sequence of state spaces $\{\mathbb{R}^N\}$ indexed by increasing N , where N is the number of components in the system.

Kushner's ODE method is closely related to the line of research called stochastic approximation, started by Robbins and Monro [20] and Kiefer and Wolfowitz [21] in the early 1950s, which studies stochastic processes similar to those addressed by Kushner's ODE method, and has been widely used in many areas (see, e.g., [22], [23], for surveys). Among the numerous following efforts, several ODE methods including that of Kushner were first developed in the 1970s (see, e.g., [24], [25]) and extensively studied thereafter (see, e.g., [26]–[28]), many times addressing problems outside of the category of stochastic approximation (see, e.g. [19]).

The general subject of the approximation of Markov chains (or equivalently, the convergence of sequences of Markov chains to certain limits) goes beyond the scope of ODE methods or stochastic approximation, and there are results on the convergence of different models to different limits. A huge class of Markov chains (discrete-time or continuous-time) that model various systems, phenomena, and abstract problems, hence having in general very different forms from ours, have been shown to converge either to ODE solutions [29]–[31] (and more generally, abstract Cauchy problems [32]), or to stochastic processes like diffusion processes [19], [33]. These results use methods different from Kushner's, but share with it the principle idea of “averaging out” of the randomness of the Markov chain. Their deeper connection lies in weak convergence theory [19], [33], [34] and methods to prove such convergence that they have in common: the operator semigroup convergence theorem [35], the martingale characterization method [36], and identification of the limit as the solution to a stochastic differential

equation [19], [33]. The reader is referred to [19], [33] and the references therein for additional information on these methods.

Similar to Kushner's ODE method, all the convergence results discussed above differ from our approach in the sense that they essentially study only the single-step “smoothing” limit as an “averaging” parameter goes to ∞ , but do not have our second-step convergence to the “large-system” limit (PDE) (as $N \rightarrow \infty$). There are systems in which the “averaging” parameter represents some “size” of the system (e.g., population in the epidemic model, the metapopulations model, and the searching parasites problem [31], [32]). However, it is still the case that the convergence requires a fixed dimension of the state space of the Markov chain, like the case of Kushner's ODE convergence, and does not apply to the “large-system” limit in our second step. For example, in the epidemic model, the Markov chain represents the number of people in a population in two states: infected and uninfected, and the large population limit is studied. This is a single-step limit and the state space of the Markov chain is always in \mathbb{R}^2 . Notice that in these cases, the Markov chains model the number or proportion of the components in different states in the system, and unlike our model, the indexing or locations of the components are either unimportant or ignored. In contrast, in our case, the spatial index of nodes is addressed throughout.

In fact, a variety of other approximation methods for large systems, in general built on different ideas from the aforementioned ones, take a similar direction: they study the number or proportion of components in a certain state (or some related performance parameters), thus ignoring their order or the difference in their *spatial* locations. For example, the famous work of Gupta and Kumar [37], followed by many others (e.g., [38]–[41]), derives scaling laws of network performance parameters (e.g., throughput); and many efforts based on mean field theory [42]–[51] or on the theory of large deviations [52]–[56] study the convergence with regard to the so-called empirical (or occupancy) measure or distribution, which essentially represents the proportion of components in certain states, to a deterministic function, as the number of components grows large, treating the components as exchangeable in terms of their order or spatial indices. These approaches differ from our work at least in the sense that they only study the statistical instead of the spatio-temporal characteristics of the system. As a result of treating the components without regard to their locations, when the limits obtained by these approaches are in fact differential equations, they are usually ODEs instead of PDEs. Note that the statistical parameters studied in these works correspond to some deterministic quantities easily obtained from our deterministic limits that directly approximate the state of the systems. For example, the proportion of nodes with, say, empty queues in our network model can be directly calculated from the limiting PDE solution (in addition, their locations are directly observ-

able); and the instantaneous throughput can be obtained by integrating the PDE solution at a certain time over the spatial domain.

Of course, there do exist numerous continuum models in a wide spectrum of areas such as physics, chemistry, ecology, economics, transportation, and sociology (e.g., [8], [57]–[63]), many of which use PDEs to formulate spatio-temporal phenomena and approximate quantities such as the probability density of particle velocity in thermodynamic systems, the concentration of reactants in chemical reactions, the population density in animal swarms, the wealth of a company in consumption-investment processes, the car density on highways, and the density of people in human crowds. All these works differ from the work presented here both by the properties of the system being studied and the analytic approaches. In addition, most of them study distributions of limiting processes that are random, while our limiting functions themselves are deterministic. We especially emphasize the difference between our results and those of the mathematical physics of hydrodynamics [64]–[69], because the latter have a similar style by deducing macroscopic behavior from microscopic interactions of individual particles, and in some special cases result in similar PDEs. However, they use an entirely different approach, which usually requires different assumptions on the systems such as translation invariant transition probabilities, conservation of the number of particles, and particular distributions of the initial state; and their limiting PDE is not the direct approximation of system state, but the density of some associated probability measure.

There is a vast literature on the convergence of a large variety of network models different from ours, to essentially two kinds of limits: the fluid limit (or functional law of large numbers approximation) [70]–[79] and the diffusion approximation (or functional central limit theorem approximation), under the so-called fluid and diffusion scalings, respectively, with the latter limit mostly studied in networks in heavy traffic [80]–[91]. (Some papers study both limits [92]–[94].) Unlike our work, this field of research focuses primarily on networks with a fixed number of nodes.

Our work is to be distinguished from approaches where the model is constructed to be a continuum representation from the start. For example, many papers treat nodes as a continuum by considering only the average density of nodes [95]–[102]; and others model network traffic as a continuum by capturing certain average characteristics of the data packet traffic, with the averaging being over possibly different time scales [103]–[105]. The latter shares a similar idea with fluid simulations discussed at the beginning of this section.

B. MARKOV CHAIN MODEL

We first describe our model in full generality. Consider N points $V_N = \{v_N(1), \dots, v_N(N)\}$ in a compact, convex Euclidean domain \mathcal{D} representing a spatial region. We assume that these points form a *uniform* grid, though the model generalizes to nonuniform spacing of points under certain

conditions (see Sec. IV for discussion). We refer to these N points in \mathcal{D} as *grid* points.

We consider a discrete-time Markov chain

$$X_{N,M}(k) = [X_{N,M}(k, 1), \dots, X_{N,M}(k, N)]^\top \in \mathbb{R}^N$$

(the superscript \top represents transpose) whose evolution is described by the stochastic difference equation

$$X_{N,M}(k+1) = X_{N,M}(k) + F_N(X_{N,M}(k)/M, U_N(k)). \quad (1)$$

Here, $X_{N,M}(k, n)$ is the real-valued state associated with the grid point $v_N(n)$ at time k , where $n = 1, \dots, N$ is a *spatial* index and $k = 0, 1, \dots$ is a *temporal* index; $U_N(k)$ are i.i.d. random vectors that do not depend on the state $X_{N,M}(k)$; M is an “averaging” parameter (explained later); and F_N is a given function.

Treating N and M as indices that grow, the equation (1) defines a doubly indexed family $X_{N,M}(k)$ of Markov chains indexed by both N and M . (We will later take M to be a function of N , and treat this family as a sequence $X_N(k)$ of the *single* index N .) Below we give a concrete example of a system described by (1).

C. A STOCHASTIC NETWORK MODEL

In this subsection we demonstrate the various objects in the abstract Markov chain model analyzed in this paper on a prototypical example. We begin by describing a stochastic model of a wireless sensor network.

Consider a network of N wireless sensor nodes uniformly placed over the domain \mathcal{D} . That is, the N nodes are located on the grid points $V_N = \{v_N(1), \dots, v_N(N)\}$ described above. We label the node at $v_N(n)$ by n , where $n = 1, \dots, N$. The sensor nodes generate, according to a probability distribution, data messages that need to be communicated to the destination nodes located on the boundary of the domain, which represent specialized devices that collect the sensor data. The sensor nodes also serve as relays for routing messages to the destination nodes. Each sensor node has the capacity to store messages in a *queue*, and is capable of either transmitting or receiving messages to or from its *immediate* neighbors. (Generalization to further ranges of transmission can be found in our paper [106].) At each time instant $k = 0, 1, \dots$, each sensor node probabilistically decides to be a transmitter or receiver, but not both. This simplified rule of transmission allows for a relatively simple representation. We illustrate such a network over a two-dimensional domain in Fig. 1(a).

In this network, communication between nodes is interference-limited because all nodes share the same wireless channel. We assume a simple collision protocol: a transmission from a transmitter to a neighboring receiver is successful if and only if none of the other neighbors of the receiver is a transmitter, as illustrated in Fig. 1(b). We assume that in a successful transmission, *one* message is transmitted from the transmitter to the receiver.

We assume that the probability that a node decides to be a transmitter is a function of its normalized queue length (normalized by an “averaging” parameter M). That is, at

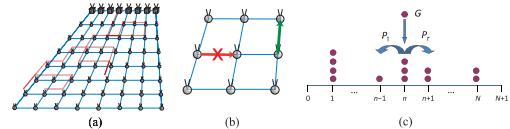


FIGURE 1. (a) An illustration of a wireless sensor network over a two-dimensional domain. Destination nodes are located at the far edge. We show the possible path of a message originating from a node located in the left-front region. (b) An illustration of the collision protocol: reception at a node fails when one of its other neighbors transmits (regardless of the intended receiver). (c) An illustration of the time evolution of the queues in the one-dimensional network model.

time k , node n decides to be a transmitter with probability $W(n, X_{N,M}(k, n)/M)$, where $X_{N,M}(k, n)$ is the *queue length* of node n at time k , and W is a given function.

In this section, for the sake of explanation, we simplify the problem even further and consider a one-dimensional domain (a two-dimensional example will be given in Sec. II-E.3). Here, N sensor nodes are equidistributed in an interval $\mathcal{D} \subset \mathbb{R}$ and labeled by $n = 1, \dots, N$. The destination nodes are located on the boundary of \mathcal{D} , labeled by $n = 0$ and $n = N + 1$.

We assume that if node n is a transmitter at a certain time instant, it randomly chooses to transmit one message to the right or the left immediate neighbor with probability $P_r(n)$ and $P_l(n)$, respectively, where $P_r(n) + P_l(n) \leq 1$. In contrast to strict equality, the inequality here allows for a more general stochastic model of transmission: after a sensor node randomly decides to transmit over the wireless channel, there is still a positive probability that the message is not transferred to its intended receiver (what might be called an “outage”).

The special destination nodes at the boundaries of the domain do not have queues; they simply receive any message transmitted to them and never themselves transmit anything. We illustrate the time evolution of the queues in the network in Fig. 1(c).

The queue lengths

$$X_{N,M}(k) = [X_{N,M}(k, 1), \dots, X_{N,M}(k, N)]^\top \in \mathbb{R}^N$$

form a Markov chain network model given by (1), where

$$U_N(k) = [Q(k, 1), \dots, Q(k, N), T(k, 1), \dots, T(k, N), G(k, 1), \dots, G(k, N)]^\top$$

is a random vector comprising independent random variables: $Q(k, n)$ are uniform random variables on $[0, 1]$ used to determine if the node is a transmitter or not; $T(k, n)$ are ternary random variables used to determine the direction a message is passed, which take values R , L , and S (representing transmitting to the right, the left, and neither, respectively) with probabilities $P_r(n)$, $P_l(n)$, and $1 - (P_r(n) + P_l(n))$, respectively; and $G(k, n)$ are the number of messages generated at node n at time k . We model $G(k, n)$ by independent Poisson random variables with mean $g(n)$.

For a generic $x = [x_1, \dots, x_N]^\top \in \mathbb{R}^N$, the n th component of $F_N(x, U_N(k))$, where $n = 1, \dots, N$, is

$$\begin{cases} 1 + G(k, n), & \text{if} \\ & Q(k, x_{n-1}) < W(n-1, x_{n-1}), T(k, n-1) = R, \\ & Q(k, x_n) > W(n, x_n), \\ & Q(k, x_{n+1}) > W(n+1, x_{n+1}); \\ & \text{or} \\ & Q(k, x_{n+1}) < W(n+1, x_{n+1}), T(k, n+1) = L, \\ & Q(k, x_n) > W(n, x_n), \\ & Q(k, x_{n-1}) > W(n-1, x_{n-1}) \\ -1 + G(k, n), & \text{if} \\ & Q(k, x_n) < W(n, x_n), T(k, n) = L, \\ & Q(k, x_{n-1}) > W(n-1, x_{n-1}), \\ & Q(k, x_{n-2}) > W(n-2, x_{n-2}); \\ & \text{or} \\ & Q(k, x_n) < W(n, x_n), T(k, n) = R, \\ & Q(k, x_{n+1}) > W(n+1, x_{n+1}), \\ & Q(k, x_{n+2}) > W(n+2, x_{n+2}) \\ G(k, n), & \text{otherwise,} \end{cases} \quad (2)$$

where x_n with $n \leq 0$ or $n \geq N+1$ are defined to be zero; and W is the function that specifies the probability that a node decides to be a transmitter, as defined earlier. Here, the three possible values of F_N correspond to the three events that at time k , node n successfully receives one message, successfully transmits one message, and does neither of the above, respectively. The inequalities and equations on the right describe conditions under which these three events occur: for example, $Q(k, x_{n-1}) < W(n-1, x_{n-1})$ corresponds to the choice of node $n-1$ to be a transmitter at time k , $T(k, n-1) = R$ corresponds to its choice to transmit to the right, $Q(k, x_n) > W(n, x_n)$ corresponds to the choice of node n to be a receiver at time k , and so on.

We simplify the situation further by assuming that $W(n, y) = \min(1, y)$. (We use this assumption throughout the paper.) With the collision protocol described earlier, this provides the analog of a network with backpressure routing [107].

After presenting the main results of the paper, we will revisit this network model in Sec. II-E and present a PDE that approximates its global behavior as an application of the main results.

D. OVERVIEW OF RESULTS IN THIS PAPER

In this subsection, we provide a brief description of the main results in Sec. II.

The Markov chain model (1) is related to a deterministic difference equation. We set

$$f_N(x) = EF_N(x, U_N(k)), \quad x \in \mathbb{R}^N, \quad (3)$$

and define $x_{N,M}(k) = [x_{N,M}(k, 1), \dots, x_{N,M}(k, N)]^\top \in \mathbb{R}^N$ by

$$\begin{aligned} x_{N,M}(k+1) &= x_{N,M}(k) + \frac{1}{M} f_N(x_{N,M}(k)), \\ x_{N,M}(0) &= \frac{x_{N,M}(0)}{M} \text{ a.s.} \end{aligned} \quad (4)$$

(“a.s.” is short for “almost surely”).

Example 1: For the one-dimensional Markov chain network model introduced in Sec. I-C, it follows from (2) (with the particular choice of $W(n, y) = \min(1, y)$) that for $x = [x_1, \dots, x_N]^\top \in [0, 1]^N$, the n th component of $f_N(x)$ in its corresponding deterministic difference equation (4), where $n = 1, \dots, N$, is (after some tedious algebra, as described in [3])

$$\begin{aligned} &(1-x_n)[P_r(n-1)x_{n-1}(1-x_{n+1}) \\ &+ P_l(n+1)x_{n+1}(1-x_{n-1})] \\ &- x_n[P_r(n)(1-x_{n+1})(1-x_{n+2}) \\ &+ P_l(n)(1-x_{n-1})(1-x_{n-2})] + g(n), \end{aligned} \quad (5)$$

where x_n with $n \leq 0$ or $n \geq N+1$ are defined to be zero.

We analyze the convergence of the Markov chain to the solution of a PDE using a two-step procedure. The first step depends heavily on the relation between $X_{N,M}(k)$ and $x_{N,M}(k)$. We show that for each N , as $M \rightarrow \infty$, the difference between $X_{N,M}(k)/M$ and $x_{N,M}(k)$ vanishes, by proving that they both converge in a certain sense to the solution of the same ODE. The basic idea of this convergence is that as the “fluctuation size” of the system decreases and the “fluctuation rate” of the system increases, the stochastic system converges to a deterministic “small-fast-fluctuation” limit, which can be characterized as the solution of a particular ODE. In our case, the smallness of the fluctuation size and largeness of the fluctuation rate is quantified by the “averaging” parameter M . We use a weak convergence theorem of Kushner [19] to prove this convergence.

In the second step, we treat M as a function of N , written M_N (therefore treating $X_{N,M_N}(k)$ and $x_{N,M_N}(k)$ as sequences of the *single* index N , written $X_N(k)$ and $x_N(k)$, respectively), and show that for *any sequence* $\{M_N\}$ of N , as $N \rightarrow \infty$, $x_N(k)$ converges to the solution of a certain PDE (and we show how to construct the PDE). This is essentially a convergence analysis on the approximating error between $x_N(k)$ and the PDE solution. We stress that this is different from the numerical analysis on classical finite difference schemes (see, e.g., [7], [108], [109]), because our difference equation (4), which originates from particular system models, differs from those designed specifically for the purpose of numerically solving differential equations. The difficulty in our convergence analysis arises from both the different form of (4) and the fact that it is in general nonlinear. We provide not only sufficient conditions for the convergence, but also a practical criterion for verifying such conditions otherwise difficult to check.

Finally, based on these two steps, we show that as N and M_N go to ∞ in a *dependent* way, the continuous-time-space extension (explained later) of the normalized Markov chain $X_N(k)/M_N$ converges to the PDE solution. We also characterize the rate of convergence. We note that special caution is needed for specifying the details of this dependence between the two indices N and M of the doubly indexed family $X_{N,M}(k)$ of Markov chains in the limiting process.

E. OUTLINE OF THE PAPER

The remainder of the paper is organized as follows. In Sec. II, we present the main theoretical results and apply the results to the wireless sensor network introduced above, and present some numerical experiments. In Sec. III, we present the proofs of the main results. Finally, we conclude the paper and discuss future work in Sec. IV.

II. MAIN RESULTS AND APPLICATIONS

A. CONSTRUCTION OF THE LIMITING PDE

We begin with the construction of the PDE whose solution describes the limiting behavior of the abstract Markov chain model.

For each N and the grid points $V_N = \{v_N(1), \dots, v_N(N)\} \subset \mathcal{D}$ as introduced in Sec. I-B, we denote the distance between any two neighboring grid points by ds_N . For any continuous function $w : \mathcal{D} \rightarrow \mathbb{R}$, let y_N be the vector in \mathbb{R}^N composed of the values of w at the grid points $v_N(n)$; i.e., $y_N = [w(v_N(1)), \dots, w(v_N(N))]^\top$. Given a point $s \in \mathcal{D}$, we let $\{s_N\} \subset \mathcal{D}$ be any sequence of grid points $s_N \in V_N$ such that as $N \rightarrow \infty$, $s_N \rightarrow s$. Let $f_N(y_N, s_N)$ be the component of the vector $f_N(y_N)$ corresponding to the location s_N ; i.e., if $s_N = v_N(n) \in V_N$, then $f_N(y_N, s_N)$ is the n th component of $f_N(y_N)$.

In order to obtain a limiting PDE, we have to make certain technical assumptions on the asymptotic behavior of the sequence of functions $\{f_N\}$ that insure that $f_N(y_N, s_N)$ is asymptotically close to an expression that looks like the right-hand side of a time-dependent PDE. Such conditions are familiar in the context of PDE limits of Brownian motion. Checking these conditions often amounts to a simple algebraic exercise. We provide a concrete example (the network model) in Sec. II-E where f_N satisfies these assumptions.

We assume that there exist sequences $\{\delta_N\}$, $\{\beta_N\}$, $\{\gamma_N\}$, and $\{\rho_N\}$, functions f and h , and a constant $c < \infty$, such that as $N \rightarrow \infty$, $\delta_N \rightarrow 0$, $\delta_N/\beta_N \rightarrow 0$, $\gamma_N \rightarrow 0$, $\rho_N \rightarrow 0$, and:

- Given s in the interior of \mathcal{D} , there exists a sequence of functions $\{\phi_N\} : \mathcal{D} \rightarrow \mathbb{R}$ such that

$$f_N(y_N, s_N)/\delta_N = f(s_N, w(s_N), \nabla w(s_N), \nabla^2 w(s_N)) + \phi_N(s_N), \quad (6)$$

for any sequence of grid points $s_N \rightarrow s$, and for N sufficiently large, $|\phi_N(s_N)| \leq c\gamma_N$; and

- Given s on the boundary of \mathcal{D} , there exists a sequence of functions $\{\varphi_N\} : \mathcal{D} \rightarrow \mathbb{R}$ such that

$$f_N(y_N, s_N)/\beta_N = h(s_N, w(s_N), \nabla w(s_N), \nabla^2 w(s_N)) + \varphi_N(s_N), \quad (7)$$

for any sequence of grid points $s_N \rightarrow s$, and for N sufficiently large, $|\varphi_N(s_N)| \leq c\rho_N$.

Here, $\nabla^i w$ represents all the i th order derivatives of w , where $i = 1, 2$.

Fix $T > 0$ for the rest of this section. Assume that there exists a unique function $z : [0, T] \times \mathcal{D} \rightarrow \mathbb{R}$ that solves the

limiting PDE

$$\dot{z}(t, s) = f(s, z(t, s), \nabla z(t, s), \nabla^2 z(t, s)), \quad (8)$$

with boundary condition

$$h(s, z(t, s), \nabla z(t, s), \nabla^2 z(t, s)) = 0 \quad (9)$$

and initial condition $z(0, s) = z_0(s)$.

Recall that $x_{N,M}(k)$ is defined by (4). Suppose that we associate the discrete time k with points on the real line spaced apart by a distance proportional to δ_N . Then, the technical assumptions (6) and (7) imply that $x_{N,M}(k)$ is, in a certain sense, close to the solution of the limiting PDE (8) with boundary condition (9). Below we develop this argument rigorously.

Establishing existence and uniqueness for the resulting nonlinear models is a difficult problem in theoretical analysis of PDEs in general. The techniques are heavily dependent on the particular form of f . Therefore, as is common with numerical analysis, we assume that this has been established. Later, we apply the general theory to the modeling of networks of particular characteristics. The resulting limiting PDE is a nonlinear reaction-convection-diffusion problem. Existence and uniqueness for such problems for “small” data and short times can be established under general conditions. Key ingredients are coercivity, which will hold as long as z is bounded away from 1, and diffusion dominance, which will also hold as long as z is bounded above.

B. CONTINUOUS TIME-SPACE EXTENSION OF THE MARKOV CHAIN

Next we define the continuous time-space extension of the Markov chain $X_{N,M}(k)$.

For each N and M , define

$$dt_{N,M} = \frac{\delta_N}{M}, t_{N,M}(k) = k dt_{N,M}, K_{N,M} = \left\lfloor \frac{T}{dt_{N,M}} \right\rfloor, \text{ and} \\ \tilde{T}_N = \frac{T}{\delta_N}. \quad (10)$$

First, we construct the continuous-time extension $X_{N,M}^{(o)}(\tilde{t})$ of $X_{N,M}(k)$, as the piecewise-constant time interpolant with interval length $1/M$ and normalized by M :

$$X_{N,M}^{(o)}(\tilde{t}) = X_{N,M}(\lfloor M\tilde{t} \rfloor)/M, \quad \tilde{t} \in [0, \tilde{T}_N]. \quad (11)$$

Similarly, define the continuous-time extension $x_{N,M}^{(o)}(\tilde{t})$ of $x_{N,M}(k)$ by

$$x_{N,M}^{(o)}(\tilde{t}) = x_{N,M}(\lfloor M\tilde{t} \rfloor), \quad \tilde{t} \in [0, \tilde{T}_N]. \quad (12)$$

Let $X_{N,M}^{(p)}(t, s)$, where $(t, s) \in [0, T] \times \mathcal{D}$, be the continuous-space extension of $X_{N,M}^{(o)}(\tilde{t})$ (with $\tilde{t} \in [0, \tilde{T}_N]$) by piecewise-constant space extensions on \mathcal{D} and with time scaled by δ_N so that the time-interval length is $\delta_N/M := dt_{N,M}$. By piecewise-constant space extension of $X_{N,M}^{(o)}$, we mean the piecewise-constant function on \mathcal{D} such that the value of this function at each point in \mathcal{D} is the value of the

component of the vector $X_{N,M}^{(o)}$ corresponding to the grid point that is “closest to the left” (taken one component at a time). Then $X_{N,M}^{(p)}(t, s)$ is the continuous time-space extension of $X_{N,M}(k)$, and for each t , $X_{N,M}^{(p)}(t, \cdot)$ is a real-valued function defined on \mathcal{D} . We illustrate in Fig. 2.

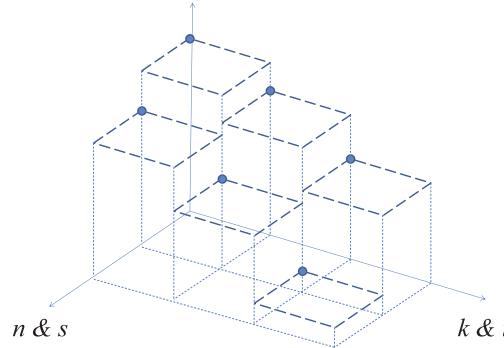


FIGURE 2. An illustration of $X_{N,M}(k)$ and $X_{N,M}^{(p)}(t, s)$ in one dimension, represented by solid dots and dashed-line rectangles, respectively.

The function $X_{N,M}^{(p)}(t, s)$ with $(t, s) \in [0, T] \times \mathcal{D}$ is in the space $D^{\mathcal{D}}[0, T]$ of functions from $[0, T] \times \mathcal{D}$ to \mathbb{R} that are Càdlàg with respect to the time component, i.e., right-continuous at each $t \in [0, T]$, and have left-hand limits at each $t \in (0, T]$. Denote the norm $\|\cdot\|^{(p)}$ on $D^{\mathcal{D}}[0, T]$ such that for $x \in D^{\mathcal{D}}[0, T]$,

$$\|x\|^{(p)} = \sup_{t \in [0, T]} \int_{\mathcal{D}} |x(t, s)| ds. \quad (13)$$

C. MAIN RESULTS FOR CONTINUUM LIMIT OF THE ABSTRACT MARKOV CHAIN MODEL

In this subsection, we present the main theorem, Theorem 1, which states that under some conditions, the continuous-time-space extension $X_{N,M}^{(p)}$ of the Markov chain $X_{N,M}(k)$ converges to the solution z of the limiting PDE (8) in the norm defined by (13), as N and M go to ∞ in a *dependent* way. By this we mean that we set M to be a function of N , written M_N , such that $M_N \rightarrow \infty$ as $N \rightarrow \infty$. Then we can treat $X_{N,M_N}(k)$, $x_{N,M_N}(k)$, $X_{N,M_N}^{(p)}$, dt_{N,M_N} , t_{N,M_N} , and K_{N,M_N} all as sequences of the *single* index N , written $X_N(k)$, $x_N(k)$, $X_N^{(p)}$, dt_N , t_N , and K_N respectively. We apply such changes of notation throughout the rest of the paper whenever M is treated as a function of N .

Define $z_N(k, n) = z(t_N(k), v_N(n))$ and $z_N(k) = [z_N(k, 1), \dots, z_N(k, N)]^\top \in \mathbb{R}^N$. Define the truncation error

$$u_N(k, n) = \frac{f_N(z_N(k), n)}{\delta_N} - \frac{z_N(k+1, n) - z_N(k, n)}{dt_N}, \quad (14)$$

and $u_N(k) = [u_N(k, 1), \dots, u_N(k, N)]^\top \in \mathbb{R}^N$. Define

$$\varepsilon_N(k, n) = x_N(k, n) - z_N(k, n), \quad (15)$$

and $\varepsilon_N(k) = [\varepsilon_N(k, 1), \dots, \varepsilon_N(k, N)]^\top \in \mathbb{R}^N$.

By (4), (10), (14), and (15), we have that

$$\begin{aligned} \varepsilon_N(k+1) \\ = \varepsilon_N(k) + \frac{1}{M_N}(f_N(x_N(k)) - f_N(z_N(k))) + dt_N u_N(k) \\ = \varepsilon_N(k) + \frac{1}{M_N}(f_N(z_N(k) + \varepsilon_N(k)) - f_N(z_N(k))) \\ + dt_N u_N(k). \end{aligned} \quad (16)$$

Let $\varepsilon_N = [\varepsilon_N(1)^\top, \dots, \varepsilon_N(K_N)^\top]^\top$ and $u_N = [u_N(0)^\top, \dots, u_N(K_N-1)^\top]^\top$ denote vectors in the $(K_N N)$ -dimensional vector space $\mathbb{R}^{K_N N}$. Assume that

$$\varepsilon_N(0) = 0. \quad (17)$$

Then by (16), for fixed z , there exists a function $H_N : \mathbb{R}^{K_N N} \rightarrow \mathbb{R}^{K_N N}$ such that

$$\varepsilon_N = H_N(u_N). \quad (18)$$

Define the vector norm $\|\cdot\|^{(N)}$ on $\mathbb{R}^{K_N N}$ such that for $x = [x(1)^\top, \dots, x(K_N)^\top]^\top \in \mathbb{R}^{K_N N}$, where $x(k) = [x(k, 1), \dots, x(k, N)]^\top \in \mathbb{R}^N$,

$$\|x\|^{(N)} = ds_N \max_{k=1, \dots, K_N} \sum_{n=1}^N |x(k, n)|. \quad (19)$$

Define

$$\mu_N = \lim_{\alpha \rightarrow 0} \sup_{\|u\|^{(N)} \leq \alpha} \frac{\|H_N(u)\|^{(N)}}{\|u\|^{(N)}}. \quad (20)$$

We now present the main theorem.

Theorem 1: (Main Theorem) Assume that:

- T1.1. there exist a sequence $\{\xi_N\}$ and $c_1 < \infty$ such that as $N \rightarrow \infty$, $\xi_N \rightarrow 0$, and for N sufficiently large, $\|u_N\|^{(N)} < c_1 \xi_N$;
- T1.2. for each N , there exists an identically distributed sequence $\{\lambda_N(k)\}$ of integrable random variables such that for each k and x , $|F_N(x, U_N(k))| \leq \lambda_N(k)$ a.s.;
- T1.3. for each N , the function $F_N(x, U_N(k))$ is continuous in x a.s.;
- T1.4. for each N , the ODE $\dot{y} = f_N(y)$ has a unique solution on $[0, \tilde{T}_N]$ for any initial condition $y(0)$, where \tilde{T}_N is as defined by (10);
- T1.5. z is Lipschitz continuous on $[0, T] \times \mathcal{D}$;
- T1.6. for each N , (17) holds; and
- T1.7. the sequence $\{\mu_N\}$ is bounded.

Then a.s., there exist $c_0 < \infty$, N_0 , and $\hat{M}_1 < \hat{M}_2 < \hat{M}_3, \dots$ such that for each $N \geq N_0$ and each $M_N \geq \hat{M}_N$,

$$\|X_N^{(p)} - z\|^{(p)} < c_0 \max\{\xi_N, ds_N\}.$$

This theorem states that as N and M_N go to ∞ in a dependent way, $X_N^{(p)}$ converges to z in $\|\cdot\|^{(p)}$ a.s. We prove this in Sec. III-C.

D. SUFFICIENT CONDITIONS ON f_N FOR THE BOUNDEDNESS OF $\{\mu_N\}$

The key assumption of Theorem 1 is that the sequence $\{\mu_N\}$ is bounded (Assumption T1.7). We present in the following theorem a result that gives specific sufficient conditions on f_N that guarantee that $\{\mu_N\}$ is bounded. This provides a practical criterion to verify this key assumption otherwise difficult to check.

Consider fixed z . We assume that $f_N \in \mathcal{C}^1$ and denote the jacobian matrix of f_N at x by $Df_N(x)$. Define for each N and for $k = 0, \dots, K_N - 1$,

$$A_N(k) = I_N + \frac{1}{M_N} Df_N(z_N(k)), \quad (21)$$

where I_N is the identity matrix in $\mathbb{R}^{N \times N}$.

We denote the 1-norm on \mathbb{R}^N and its induced norm on $\mathbb{R}^{N \times N}$ both by $\|\cdot\|_1^{(N)}$; i.e., for a vector $x = [x_1, \dots, x_N]^\top \in \mathbb{R}^N$,

$$\|x\|_1^{(N)} = \sum_{n=1}^N |x_n|, \quad (22)$$

and for a matrix $A \in \mathbb{R}^{N \times N}$ with a_{ij} being its (i, j) th component,

$$\|A\|_1^{(N)} = \max_{j=1, \dots, N} \sum_{i=1}^N |a_{ij}|. \quad (23)$$

We then have

Theorem 2: (Sufficient condition for key assumption)

Assume that:

- T2.1. for each N , (17) holds;
- T2.2. for each N , $f_N \in \mathcal{C}^1$; and
- T2.3. there exists $c < \infty$ such that for N sufficiently large and for $k = 1, \dots, K_N - 1$, $\|A_N(k)\|_1^{(N)} \leq 1 + c dt_N$, where $\|\cdot\|_1^{(N)}$ is defined by (23).

Then $\{\mu_N\}$ is bounded.

We prove this in Sec. III-D.

In Sec. III-E, we will show that these sufficient conditions hold for the network model described in Sec. I-C, and use this theorem to prove the convergence of its underlying Markov chain to a PDE.

E. APPLICATION TO NETWORK MODELS

In this subsection, we apply the main results to show how the Markov chain modeling the network introduced in Sec. I-C can be approximated by the solution of a PDE. This approximation was heuristically developed in [3].

We first deal with the one-dimensional network model. Its corresponding stochastic and deterministic difference equations (1) and (4) were specified by (2) and (5), respectively.

For this model we set δ_N (introduced in Sec. II-A) to be ds_N^2 . Then

$$dt_{N,M} := \delta_N/M = ds_N^2/M.$$

Assume that

$$P_l(n) = p_l(v_N(n)) \text{ and } P_r(n) = p_r(v_N(n)), \quad (24)$$

where $p_l(s)$ and $p_r(s)$ are real-valued functions defined on \mathcal{D} such that

$$p_l(s) = b(s) + c_l(s)ds_N \text{ and } p_r(s) = b(s) + c_r(s)ds_N.$$

Let $c = c_l - c_r$. The values $b(s)$ and $c(s)$ correspond to diffusion and convection quantities in the limiting PDE. Because $p_l(s) + p_r(s) \leq 1$, it is necessary that $b(s) \leq 1/2$. In order to guarantee that the number of messages entering the system from outside over finite time intervals remains finite throughout the limiting process, we set $g(n) = M g_p(v_N(n)) dt_N$, where $g_p : \mathcal{D} \rightarrow \mathbb{R}$ is called the message generation rate. Assume that b, c_l, c_r , and g_p are in \mathcal{C}^1 . Further assume that $x_{N,M}(k) \in [0, 1]^N$ for each k . Then f_N is in \mathcal{C}^1 .

We have assumed above that the probabilities P_l and P_r of the direction of transmission are the values of the continuous functions p_l and p_r at the grid points, respectively. This may correspond to stochastic routing schemes where nodes in close vicinity behave similarly based on some local information that they share; or to those with an underlying network-wide directional configuration that are continuous in space, designed to relay messages to destination nodes at known locations. On the other hand, the results can be extended to situations with certain levels of discontinuity, as discussed in Sec. IV.

By these assumptions and definitions, it follows from (5) that the function f in (8) for this network model is:

$$\begin{aligned} f(s, z(t, s), \nabla z(t, s), \nabla^2 z(t, s)) \\ = b(s) \frac{\partial}{\partial s} ((1 - z(t, s))(1 + 3z(t, s))z_s(t, s)) \\ + 2(1 - z(t, s))z_s(t, s)b_s(s) \\ + z(t, s)(1 - z(t, s))^2 b_{ss}(s) \\ + \frac{\partial}{\partial s} (c(s)z(t, s)(1 - z(t, s))^2) + g_p(s). \end{aligned} \quad (25)$$

Here, a single subscript s represents first derivative and a double subscript ss represents second derivative.

Note that the computations needed to obtain (25) (and later, (26), (48), and (49)) require tedious but elementary algebraic manipulations. For this purpose, we found it helpful to use the symbolic tools in Matlab.

Based on the behavior of nodes $n = 1$ and $n = N$ next to the destination nodes, we derive the boundary condition (9) of the PDE of this network. For example, the node $n = 1$ receives messages only from the right and encounters no interference when transmitting to the left. Replacing x_n with $n \leq 0$ or $n \geq N + 1$ by 0, it follows that the 1st component of $f_N(x)$ is

$$(1 - x_n)P_l(n+1)x_{n+1} \\ - x_n[P_l(n) + P_r(n)(1 - x_{n+1})(1 - x_{n+2})] + g(n).$$

Similarly, the N th component of $f_N(x)$ is

$$(1 - x_n)P_r(n-1)x_{n-1} \\ - x_n[P_r(n) + P_l(n)(1 - x_{n-1})(1 - x_{n-2})] + g(n).$$

Set β_N , defined in Sec. II-A, to be 1. Then from each of the above two functions we get the function h in (9) for the one-dimensional network:

$$\begin{aligned} h(s, z(t, s), \nabla z(t, s), \nabla^2 z(t, s)) \\ = -b(s)z(s)^3 + b(s)z(s)^2 - b(s)z(s). \end{aligned} \quad (26)$$

Note that the function h is the limit of $f_N(y_N, s_N)/\beta_N$, not $f_N(y_N, s_N)/\delta_N$ (whose limit is f). Solving $h = 0$ for real z , we have the boundary condition $z(t, s) = 0$.

Let z be the solution of the PDE (8) with f specified by (25) and with boundary condition $z(t, s) = 0$ and initial condition $z(0, s) = z_0(s)$. Assume that (17) holds. As in Sec. II-C, we treat M as a sequence of N , written M_N . In the following theorem we show the convergence of the Markov chain modeling the one-dimensional network to the PDE solution.

Theorem 3: (Convergence of network model) For the one-dimensional network model, a.s., there exist $c_0 < \infty$, N_0 , and $\hat{M}_1 < \hat{M}_2 < \hat{M}_3, \dots$ such that for each $N \geq N_0$ and each $M_N \geq \hat{M}_N$, $\|X_N^{(p)} - z\|^{(p)} < c_0 ds_N$.

We prove this in Sec. III-E.

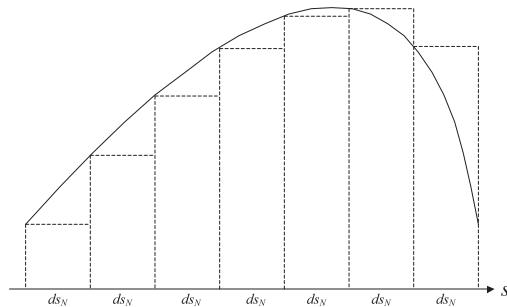


FIGURE 3. The PDE solution at a fixed time that approximates the normalized queue lengths of the network.

1) INTERPRETATION OF LIMITING PDE

Now we make some remarks on how to interpret a given limiting PDE. First, for fixed N and M , the normalized queue length of node n at time k , is approximated by the value of the PDE solution z at the corresponding point in $[0, T] \times \mathcal{D}$; i.e., $\frac{X_{N,M}(k,n)}{M} \approx z(t_{N,M}(k), v_N(n))$.

Second, we discuss how to interpret $C(t_o) := \int_{\mathcal{D}} z(t_o, s) ds$, the area below the curve $z(t_o, s)$ for fixed $t_o \in [0, T]$. Let $k_o = \lfloor t_o/dt_{N,M} \rfloor$. Then we have that $z(t_o, v_N(n))ds_N \approx \frac{X_{N,M}(k_o,n)}{M}ds_N$, the area of the n th rectangle in Fig. 3. Therefore

$$C(t_o) \approx \sum_{n=1}^N z(t_o, v_N(n))ds_N \approx \sum_{n=1}^N \frac{X_{N,M}(k_o,n)}{M}ds_N,$$

the sum of all rectangles. If we assume that all messages in the queue have roughly the same bits, and think of ds_N as the “coverage” of each node, then the area under any segment of the curve measures a kind of “data-coverage product” of the nodes covered by the segment, in the unit of “bit·meter.” As $N \rightarrow \infty$, the total normalized queue length

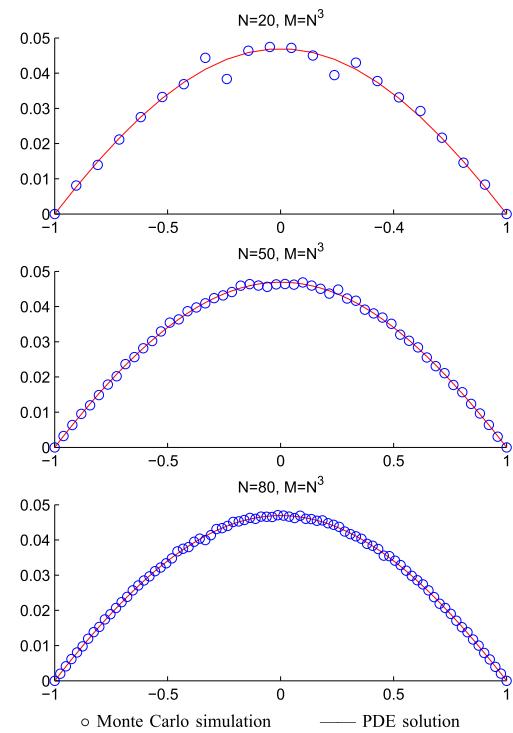


FIGURE 4. The Monte Carlo simulations (with different N and M) and the PDE solution of a one-dimensional network, with $b = 1/2$ and $c = 0$, at $t = 1s$.

$\sum_{n=1}^N X_{N,M}(k_o, n)/M$ of the network does go to ∞ ; however, the coverage ds_N of each node goes to 0. Hence the sum of the “data-coverage product” can be approximated by the finite area $C(t_o)$.

2) COMPARISONS OF THE PDE SOLUTIONS AND MONTE CARLO SIMULATIONS OF THE NETWORKS

In the remainder of this section, we compare the limiting PDE solutions with Monte Carlo simulations of the networks.¹

We first consider a one-dimensional network over the domain $\mathcal{D} = [-1, 1]$. We use the initial condition $z_0(s) = l_1 e^{-s^2}$, where $l_1 > 0$ is a constant, so that initially the nodes in the middle have messages to transmit, while those near the boundaries have very few. We set the message generation rate $g_p(s) = l_2 e^{-s^2}$, where $l_2 > 0$ is a parameter determining the total load of the system.

We use three sets of values of $N = 20, 50, 80$ and $M = N^3$, and show the PDE solution and the Monte Carlo simulation results with different N and M at $t = 1s$. The networks have diffusion $b = 1/2$ and convection $c = 0$ in Fig. 4 and $c = 1$ in Fig. 5, respectively, where the x-axis denotes the node location and y-axis denotes the normalized queue length.

For the three sets of the values of $N = 20, 50, 80$ and $M = N^3$, with $c = 0$, the maximum absolute errors of

¹This paper has supplementary downloadable material available at <http://ieeexplore.ieee.org>, provided by the authors. This includes eight multimedia AVI format movie clips, which show comparisons of network simulations and limiting PDE solutions. This material is 34.7 MB in size.

the PDE approximation are 5.6×10^{-3} , 1.3×10^{-3} , and 1.1×10^{-3} , respectively; and with $c = 1$, the errors are 4.4×10^{-3} , 1.5×10^{-3} , and 1.1×10^{-3} , respectively. As we can see, as N and M increase, the resemblance between the Monte Carlo simulations and the PDE solution becomes stronger. In the case of very large N and M , it is difficult to distinguish the results.

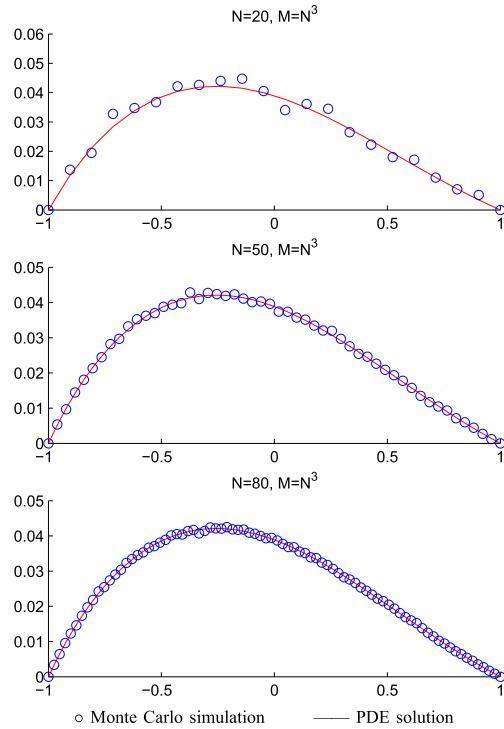


FIGURE 5. The Monte Carlo simulations (with different N and M) and the PDE solution of a one-dimensional network, with $b = 1/2$ and $c = 1$, at $t = 1s$.

We stress that the PDEs only took fractions of a second to solve on a computer, while the Monte Carlo simulations took on the order of tens of hours.

3) A TWO-DIMENSIONAL NETWORK

The generalization of the continuum model to higher dimensions is straightforward, except for more arduous algebraic manipulation. Likewise, the convergence analysis is similar to the one dimensional case.

We consider a two-dimensional network of $N = N_1 \times N_2$ sensor nodes uniformly placed over a domain $\mathcal{D} \subset \mathbb{R}^2$. Here we switch to a two-dimensional labeling scheme. We label the nodes by (n_1, n_2) , where $n_1 = 1, \dots, N_1$ and $n_2 = 1, \dots, N_2$, and denote the grid point in \mathcal{D} corresponding to node (n_1, n_2) by $v_N(n_1, n_2)$. This labeling scheme is more intuitive for this two-dimensional scenario, but is essentially equivalent to the single-label one. (e.g., if we set $n := (n_1 - 1)N_2 + n_2$ and $\hat{v}_N(n) := v_N(n_1, n_2)$, then $\hat{v}_N(n)$ form the same grid.)

Again let the distance between any two neighboring nodes be ds_N . Assume that node (n_1, n_2) randomly chooses to transmit to the east, west, north, or south immediate neighbor with

probabilities $P_e(n_1, n_2) = b_1(v_N(n_1, n_2)) + c_e(v_N(n_1, n_2))ds_N$, $P_w(n_1, n_2) = b_1(v_N(n_1, n_2)) + c_w(v_N(n_1, n_2))ds_N$, $P_n(n_1, n_2) = b_2(v_N(n_1, n_2)) + c_n(v_N(n_1, n_2))ds_N$, and $P_s(n_1, n_2) = b_2(v_N(n_1, n_2)) + c_s(v_N(n_1, n_2))ds_N$, respectively, where $P_e(n_1, n_2) + P_w(n_1, n_2) + P_n(n_1, n_2) + P_s(n_1, n_2) \leq 1$. Therefore it is necessary that $b_1(s) + b_2(s) \leq 1/2$. Define $c_1 = c_w - c_e$ and $c_2 = c_s - c_n$.

The derivation of the limiting PDE is similar to those of the one-dimensional case, except that we now have to consider transmission to and interference from four directions instead of two. We present the limiting PDE here without the detailed derivation:

$$\dot{z} = \sum_{j=1}^2 b_j \frac{\partial}{\partial s_j} \left((1+5z)(1-z)^3 \frac{\partial z}{\partial s_j} \right) + 2(1-z)^3 \frac{\partial z}{\partial s_j} \frac{db_j}{ds_j} \\ + z(1-z)^4 \frac{d^2 b_j}{ds_j^2} + \frac{\partial}{\partial s_j} (c_j z(1-z)^4) + g_p,$$

with boundary condition $z(t, s) = 0$ and initial condition $z(0, s) = z_0(s)$, where $t \in [0, T]$ and $s = (s_1, s_2) \in \mathcal{D}$.

We now compare the PDE approximation and the Monte Carlo simulations of a network over the domain $\mathcal{D} = [-1, 1] \times [-1, 1]$. We use the initial condition $z_0(s) = l_1 e^{-(s_1^2+s_2^2)}$, where $l_1 > 0$ is a constant. We set the message generation rate $g_p(s) = l_2 e^{-(s_1^2+s_2^2)}$, where $l_2 > 0$ is a constant.

We use three different sets of the values of $N_1 \times N_2$ and M , where $N_1 = N_2 = 20, 50, 80$ and $M = N_1^3$. We show the contours of the normalized queue length from the PDE solution and the Monte Carlo simulation results with different sets of values of N_1 , N_2 , and M , at $t = 0.1s$. The networks have diffusion $b_1 = b_2 = 1/4$ and convection $c_1 = c_2 = 0$ in Fig. 6 and $c_1 = -2, c_2 = -4$ in Fig. 7, respectively.

For the three sets of values of $N_1 = N_2 = 20, 50, 80$ and $M = N_1^3$, with $c_1 = c_2 = 0$, the maximum absolute errors are $3.2 \times 10^{-3}, 1.1 \times 10^{-3}$, and 6.8×10^{-4} , respectively; and with $c_1 = -2, c_2 = -4$, the errors are $4.1 \times 10^{-3}, 1.0 \times 10^{-3}$, and 6.6×10^{-4} , respectively. Again the accuracy of the continuum model increases with N_1, N_2 , and M .

It took 3 days to do the Monte Carlo simulation of the network at $t = 0.1s$ with 80×80 nodes and the maximum queue length $M = 80^3$, while the PDE solved on the same machine took less than a second. We could not do Monte Carlo simulations of any larger networks or greater values of t because of prohibitively long computation time.

III. PROOFS OF THE MAIN RESULTS

This section is devoted solely to the proofs of the results in Sec. II. As such, the material here is highly technical and might be tedious to follow in detail, though we have tried our best to make it as readable as possible. The reader can safely skip this section without doing violence to the main ideas of the paper, though much of our hard work is reflected here.

We first prove Theorem 1 (Main Theorem) by analyzing the convergence of the Markov chains $X_{N,M}(k)$ to the solution of the limiting PDE in a two-step procedure. In the first

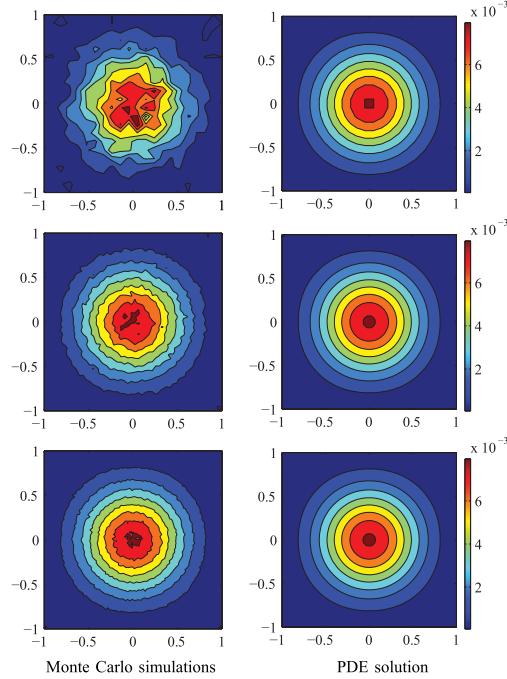


FIGURE 6. The Monte Carlo simulations (from top to bottom, with $N_1 = N_2 = 20, 50, 80$, respectively, and $M = N_1^3$) and the PDE solution of a two-dimensional network, with $b_1 = b_2 = 1/4$ and $c_1 = c_2 = 0$, at $t = 0.1s$.

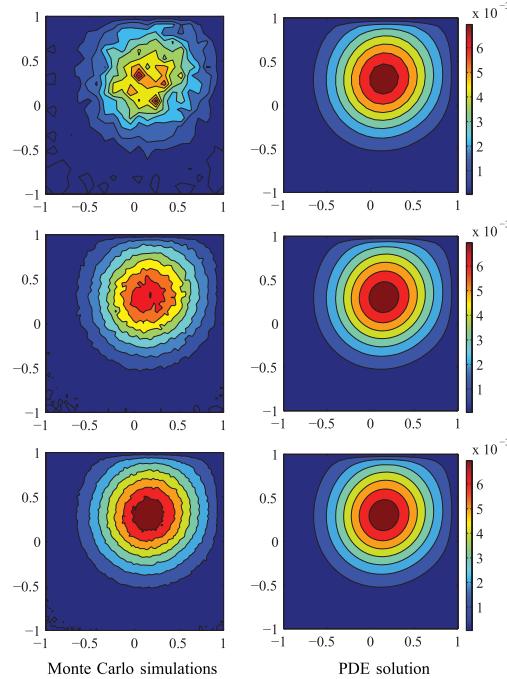


FIGURE 7. The Monte Carlo simulations (from top to bottom, with $N_1 = N_2 = 20, 50, 80$, respectively, and $M = N_1^3$) and the PDE solution of a two-dimensional network, with $b_1 = b_2 = 1/4$ and $c_1 = -2, c_2 = -4$, at $t = 0.1s$.

step, for each N , we show in Sec. III-A that as $M \rightarrow \infty$, $X_{N,M}(k)/M$ converges to $x_{N,M}(k)$. In the second step, we treat M as a function of N , written M_N , and for any sequence $\{M_N\}$, we show in Sec. III-B that as $N \rightarrow \infty$, $x_N(k)$ converges

to the PDE solution. Based on the two steps, we show in Sec. III-C that as N and M_N go to ∞ in a dependent way, $X_N^{(p)}$ converges to the PDE solution, proving Theorem 1. We then prove Theorem 2 (Sufficient condition for key assumption) in Sec. III-D. Finally, in Sec. III-E, we prove Theorem 3 (Convergence of network model) using Theorem 1 and 2.

A. CONVERGENCE OF $X_{N,M}(k)$ AND $x_{N,M}(k)$ TO THE SOLUTION OF THE SAME ODE

In this subsection, we show that for each N , $X_{N,M}(k)/M$ and $x_{N,M}(k)$ are close in a certain sense for large M under certain conditions, by proving that both their continuous-time extensions converge to the solution of the same ODE.

For fixed T and N , by (10), \tilde{T}_N is fixed. As defined by (11) and (12) respectively, both $X_{N,M}^{(o)}(\tilde{t})$ and $x_{N,M}^{(o)}(\tilde{t})$ with $\tilde{t} \in [0, \tilde{T}_N]$ are in the space $D^N[0, \tilde{T}_N]$ of \mathbb{R}^N -valued Càdlàg functions on $[0, \tilde{T}_N]$. Since they both depend on M , each one of them forms a sequence of functions in $D^N[0, \tilde{T}_N]$ indexed by $M = 1, 2, \dots$. Define the ∞ -norm $\|\cdot\|_\infty^{(o)}$ on $D^N[0, \tilde{T}_N]$; i.e., for $x \in D^N[0, \tilde{T}_N]$,

$$\|x\|_\infty^{(o)} = \max_{n=1,\dots,N} \sup_{t \in [0, \tilde{T}_N]} |x_n(t)|,$$

where x_n is the n th components of x .

Now we present a lemma stating that under some conditions, for each N , as $M \rightarrow \infty$, $X_{N,M}^{(o)}$ converges uniformly to the solution of the ODE $\dot{y} = f_N(y)$, and $x_{N,M}^{(o)}$ converges uniformly to the same solution, both on $[0, \tilde{T}_N]$.

Lemma 1: Assume, for each N , that:

- L1.1. there exists an identically distributed sequence $\{\lambda_N(k)\}$ of integrable random variables such that for each k and x , $|F_N(x, U_N(k))| \leq \lambda_N(k)$ a.s.;
- L1.2. the function $F_N(x, U_N(k))$ is continuous in x a.s.; and
- L1.3. the ODE $\dot{y} = f_N(y)$ has a unique solution on $[0, \tilde{T}_N]$ for any initial condition $y(0)$.

Suppose that as $M \rightarrow \infty$, $X_{N,M}^{(o)}(0) \xrightarrow{P} y(0)$ and $x_{N,M}^{(o)}(0) \rightarrow y(0)$, where “ \xrightarrow{P} ” represents convergence in probability. Then, for each N , as $M \rightarrow \infty$, $\|X_{N,M}^{(o)} - y\|_\infty^{(o)} \xrightarrow{P} 0$ and $\|x_{N,M}^{(o)} - y\|_\infty^{(o)} \rightarrow 0$ on $[0, \tilde{T}_N]$, where y is the unique solution of $\dot{y} = f_N(y)$ with initial condition $y(0)$.

To prove Lemma 1, we first present a lemma due to Kushner [19].

Lemma 2: Assume, for each N , that:

- L2.1. the set $\{|F_N(x, U_N(k))| : k \geq 0\}$ is uniformly integrable;
- L2.2. for each k and each bounded random variable X ,

$$\lim_{\delta \rightarrow 0} E \sup_{|Y| \leq \delta} |F_N(X, U_N(k)) - F_N(X + Y, U_N(k))| = 0;$$

and

- L2.3. there is a function $\hat{f}_N(\cdot)$ [continuous by $\hat{2}$] such that as $n \rightarrow \infty$,

$$\frac{1}{n} \sum_{k=0}^n F_N(x, U_N(k)) \xrightarrow{P} \hat{f}_N(x).$$

Suppose that, for each N , $\dot{y} = \hat{f}_N(y)$ has a unique solution on $[0, \tilde{T}_N]$ for any initial condition, and that $X_{N,M}^{(o)}(0) \Rightarrow y(0)$, where “ \Rightarrow ” represents weak convergence. Then for each N , as $M \rightarrow \infty$, $\|X_{N,M}^{(o)} - y\|_\infty^{(o)} \Rightarrow 0$ on $[0, \tilde{T}_N]$.

We note that in Kushner's original theorem, the convergence of $X_{N,M}^{(o)}$ to y is stated in terms of Skorokhod norm [19], but it is equivalent to the ∞ -norm in our case where the time interval $[0, \tilde{T}_N]$ is finite and the limit y is continuous [110].

We now prove Lemma 1 by showing that the Assumptions L2.1, L2.2, and L2.3 of Lemma 2 hold under the Assumptions L1.1, L1.2, and L1.3 of Lemma 1.

Proof of Lemma 1: Since $\lambda_N(k)$ is integrable, as $a \rightarrow \infty$, $E|\lambda_N(k)|1_{\{|\lambda_N(k)|>a\}} \rightarrow 0$, where 1_A is the indicator function of set A . By Assumption L1.1, for each k, x , and $a > 0$,

$$\begin{aligned} E|F_N(x, U_N(k))|1_{\{|F_N(x, U_N(k))|>a\}} \\ \leq E|\lambda_N(k)|1_{\{|F_N(x, U_N(k))|>a\}} \\ \leq E|\lambda_N(k)|1_{\{|\lambda_N(k)|>a\}}. \end{aligned}$$

Therefore as $a \rightarrow \infty$,

$$\sup_{k \geq 0} E|F_N(x, U_N(k))|1_{\{|F_N(x, U_N(k))|>a\}} \rightarrow 0;$$

i.e., the family $\{|F_N(x, U_N(k))| : k \geq 0\}$ is uniformly integrable, and hence Assumption L2.1 holds.

By Assumption L1.2, for each k and each bounded X , a.s.,

$$\lim_{\delta \rightarrow 0} \sup_{|Y| \leq \delta} |F_N(X, U_N(k)) - F_N(X + Y, U_N(k))| = 0.$$

By Assumption L1.1, for each k and each bounded X and Y , a.s.,

$$\begin{aligned} |F_N(X, U_N(k)) - F_N(X + Y, U_N(k))| \\ \leq |F_N(X, U_N(k))| + |F_N(X + Y, U_N(k))| \leq 2\lambda_N(k). \end{aligned}$$

Therefore, for each k , each bounded X , and each δ , a.s.,

$$\left| \sup_{|Y| \leq \delta} |F_N(X, U_N(k)) - F_N(X + Y, U_N(k))| \right| \leq 2\lambda_N(k),$$

an integrable random variable. By the dominant convergence theorem,

$$\begin{aligned} \lim_{\delta \rightarrow 0} E \sup_{|Y| \leq \delta} |F_N(X, U_N(k)) - F_N(X + Y, U_N(k))| \\ = E \lim_{\delta \rightarrow 0} \sup_{|Y| \leq \delta} |F_N(X, U_N(k)) - F_N(X + Y, U_N(k))| = 0. \end{aligned}$$

Hence Assumption L2.2 holds.

Since $U_N(k)$ are i.i.d., by the weak law of large numbers and the definition of f_N in (3), as $n \rightarrow \infty$,

$$\frac{1}{n} \sum_{k=0}^n F_N(x, U_N(k)) \xrightarrow{P} f_N(x).$$

Hence Assumption L2.3 holds.

Therefore, by Lemma 2, for each N , as $M \rightarrow \infty$, $\|X_{N,M}^{(o)} - y\|_\infty^{(o)} \Rightarrow 0$ on $[0, \tilde{T}_N]$. For any sequence of random processes $\{X_n\}$, if A is a constant, $X_n \Rightarrow A$ if and only if $X_n \xrightarrow{P} A$. Therefore, as $M \rightarrow \infty$, $\|X_{N,M}^{(o)} - y\|_\infty^{(o)} \xrightarrow{P} 0$ on $[0, \tilde{T}_N]$.

The same argument implies the deterministic convergence of $x_{N,M}^{(o)}$: as $M \rightarrow \infty$, $\|x_{N,M}^{(o)} - y\|_\infty^{(o)} \rightarrow 0$ on $[0, \tilde{T}_N]$. \square

Based on Lemma 1, we get the following lemma, which states that, for each N , $X_{N,M}^{(o)}$ and $x_{N,M}^{(o)}$ are close with high probability for large M .

Lemma 3: Let the assumptions of Lemma 1 hold. Then for any sequence $\{\zeta_N\}$, for each N and for M sufficiently large,

$$P\{\|X_{N,M}^{(o)} - x_{N,M}^{(o)}\|_\infty^{(o)} > \zeta_N\} \leq 1/N^2 \text{ on } [0, \tilde{T}_N].$$

Proof: By the triangle inequality,

$$\|X_{N,M}^{(o)} - x_{N,M}^{(o)}\|_\infty^{(o)} \leq \|X_{N,M}^{(o)} - y\|_\infty^{(o)} + \|x_{N,M}^{(o)} - y\|_\infty^{(o)}.$$

By Lemma 1, for each N , as $M \rightarrow \infty$, $\|X_{N,M}^{(o)} - x_{N,M}^{(o)}\|_\infty^{(o)} \xrightarrow{P} 0$ on $[0, \tilde{T}_N]$. This completes the proof. \blacksquare

Since $X_{N,M}^{(o)}$ and $x_{N,M}^{(o)}$ are the continuous-time extensions of $X_{N,M}(k)$ and $x_{N,M}(k)$ by piecewise-constant extensions, respectively, we have the following corollary stating that for each N , as $M \rightarrow \infty$, $X_{N,M}(k)/M$ converges uniformly to $x_{N,M}(k)$.

Corollary 1: Let the assumptions of Lemma 1 hold. Then for any sequence $\{\zeta_N\}$, for each N and for M sufficiently large, we have that

$$P \left\{ \max_{\substack{k=1, \dots, K_{N,M} \\ n=1, \dots, N}} \left| \frac{X_{N,M}(k, n)}{M} - x_{N,M}(k, n) \right| > \zeta_N \right\} \leq \frac{1}{N^2}.$$

B. CONVERGENCE OF $x_N(k)$ TO THE LIMITING PDE

For the remainder of this section, we treat M as a function of N , written M_N . We now state conditions under which ε_N converges to 0 for any sequence $\{M_N\}$ as $N \rightarrow \infty$.

Lemma 4: Assume that:

L4.1 there exist a sequence $\{\xi_N\}$ and $c_1 < \infty$ such that as $N \rightarrow \infty$, $\xi_N \rightarrow 0$, and for N sufficiently large, $\|u_N\|^{(N)} < c_1 \xi_N$;

L4.2 for each N , (17) holds; and

L4.3 the sequence $\{\mu_N\}$ is bounded.

Then there exists $c_0 < \infty$ such that for any sequence $\{M_N\}$ and N sufficiently large, $\|\varepsilon_N\|^{(N)} < c_0 \xi_N$.

Proof: By the definition of μ_N (20), for each N , there exists $\delta > 0$ such that for $\alpha < \delta$,

$$\sup_{\|u\|^{(N)} \leq \alpha} \frac{\|H_N(u)\|^{(N)}}{\|u\|^{(N)}} \leq \mu_N + 1.$$

By Assumption L4.1, $\|u_N\|^{(N)} \rightarrow 0$ as $N \rightarrow \infty$. Then there exists α_1 such that for N sufficiently large, $\|u_N\|^{(N)} \leq \alpha_1 < \delta$, and hence

$$\frac{\|H_N(u_N)\|^{(N)}}{\|u_N\|^{(N)}} \leq \sup_{\|u\|^{(N)} \leq \alpha_1} \frac{\|H_N(u)\|^{(N)}}{\|u\|^{(N)}} \leq \mu_N + 1.$$

Therefore, for N sufficiently large,

$$\|\varepsilon_N\|^{(N)} = \|H_N(u_N)\|^{(N)} \leq (\mu_N + 1) \|u_N\|^{(N)}.$$

By Assumption L4.3, and because the derivation above does not depend on the choice of the sequence $\{M_N\}$, the proof is completed. \blacksquare

C. PROOF OF THEOREM 1

We now prove the main theorem.

Proof of Theorem 1: By Lemma 4, there exist a sequence $\{\xi_N\}$ and $c_2 < \infty$ such that as $N \rightarrow \infty$, $\xi_N \rightarrow 0$, and for N sufficiently large, $\|\varepsilon_N\|^{(N)} \leq c_2 \xi_N$.

Let $X_N = [X_N(1)^\top, \dots, X_N(K_N)^\top]^\top / M_N$, $x_N = [x_N(1)^\top, \dots, x_N(K_N)^\top]^\top$, and $z_N = [z_N(1), \dots, z_N(K_N)]^\top$ denote vectors in $\mathbb{R}^{K_N N}$. Hence $\varepsilon_N = x_N - z_N$.

For $x \in \mathbb{R}^{K_N N}$, where $x = [x(1)^\top, \dots, x(K_N)^\top]^\top$ and $x(k) = [x(k, 1), \dots, x(k, N)]^\top \in \mathbb{R}^N$, we have that

$$\|x\|^{(N)} \leq \max_{\substack{k=1, \dots, K_N \\ n=1, \dots, N}} |x(k, n)|.$$

Therefore, by Corollary 1, there exists a sequence $\{\tilde{M}_N\}$ such that if for each N , $M_N \geq \tilde{M}_N$, then

$$\sum_{N=1}^{\infty} P \left\{ \|X_N - x_N\|^{(N)} > \xi_N \right\} \leq \sum_{N=1}^{\infty} 1/N^2 < \infty.$$

It follows from the first Borel-Cantelli Lemma that a.s., there exists N_1 such that for $N \geq N_1$ and $M_N \geq \tilde{M}_N$, $\|X_N - x_N\|^{(N)} \leq \xi_N$.

By the triangle inequality,

$$\|X_N - z_N\|^{(N)} \leq \|X_N - x_N\|^{(N)} + \|\varepsilon_N\|^{(N)}.$$

Therefore, a.s., there exists N_2 such that for $N \geq N_2$ and $M_N > \tilde{M}_N$,

$$\|X_N - z_N\|^{(N)} < (c_2 + 1)\xi_N. \quad (27)$$

Let $z_N^{(p)}(t, s)$, where $(t, s) \in [0, T] \times \mathcal{D}$, be the continuous-time-space extension of $z_N(k)$ defined in the same way as $X_N^{(p)}(t, s)$ is defined from $X_N(k)$. Then by its definition, we have that

$$\|X_N^{(p)} - z_N^{(p)}\|^{(p)} = \|X_N - z_N\|^{(N)}. \quad (28)$$

Let $\Omega_N(k, n) = \Omega_N^{(t)}(k) \times \Omega_N^{(s)}(n)$ be the subset of $[0, T] \times \mathcal{D}$ containing $(t_N(k), v_N(n))$ over which $z_N^{(p)}$ is piecewise constant; i.e., $t_N(k) \in \Omega_N^{(t)}(k)$ and $v_N(n) \in \Omega_N^{(s)}(n)$, and for all $(t, s) \in \Omega_N(k, n)$, $z_N^{(p)}(t, s) = z_N^{(p)}(t_N(k), v_N(n)) = z(t_N(k), v_N(n))$.

By (10), there exists a sequence $\{\bar{M}_N\}$ such that if for each N , $M_N \geq \bar{M}_N$, then for N sufficiently large, $dt_N \leq ds_N$. By Assumption T1.5, there exists $c_3 < \infty$ such that for N sufficiently large, for $M_N \geq \bar{M}_N$, and for $k = 1, \dots, K_N$ and $n = 1, \dots, N$,

$$|z(t_N(k), v_N(n)) - z(t, s)| \leq c_3 ds_N, \quad (t, s) \in \Omega_N(k, n).$$

Then we have that

$$\begin{aligned} \|z_N^{(p)} - z\|^{(p)} &= \sup_{t \in [0, T]} \int_{\mathcal{D}} |z_N^{(p)}(t, s) - z(t, s)| ds \\ &= \sup_{t \in [0, T]} \sum_n \int_{\Omega_N^{(s)}(n)} |z_N^{(p)}(t, s) - z(t, s)| ds \\ &= \max_k \sup_{t \in \Omega_N^{(t)}(k)} \sum_n \int_{\Omega_N^{(s)}(n)} |z_N^{(p)}(t, s) - z(t, s)| ds \end{aligned}$$

$$\begin{aligned} &\leq \max_k \sum_n \int_{\Omega_N^{(s)}(n)} \sup_{t \in \Omega_N^{(t)}(k)} |z_N^{(p)}(t, s) - z(t, s)| ds \\ &= \max_k \sum_n \int_{\Omega_N^{(s)}(n)} \sup_{t \in \Omega_N^{(t)}(k)} |z(t_N(k), v_N(n)) - z(t, s)| ds \\ &\leq \max_k \sum_n \int_{\Omega_N^{(s)}(n)} c_3 ds_N = c_3 ds_N |\mathcal{D}|, \end{aligned} \quad (29)$$

where $|\mathcal{D}|$ is the Lebesgue measure of \mathcal{D} .

By the triangle inequality,

$$\|X_N^{(p)} - z\|^{(p)} \leq \|X_N^{(p)} - z_N^{(p)}\|^{(p)} + \|z_N^{(p)} - z\|^{(p)}.$$

Set $\hat{M}_N = \max\{\tilde{M}_N, \bar{M}_N\}$. By (27), (28), and (29), a.s., there exist $c_0 < \infty$ and N_0 such that for $N \geq N_0$ and $M_N \geq \hat{M}_N$,

$$\|X_N^{(p)} - z\|^{(p)} < c_0 \max\{\xi_N, ds_N\}. \quad \square$$

D. PROOF OF THEOREM 2

To prove Theorem 2, we first prove Lemma 5 and 6 below.

First we provide in Lemma 5 a sequence bounding $\{\mu_N\}$ from above. By (18), for each N , for $k = 1, \dots, K_N$ and $n = 1, \dots, N$, we can write $\varepsilon_N(k, n) = H_N^{(k, n)}(u_N)$, where $H_N^{(k, n)}$ is from $\mathbb{R}^{K_N N}$ to \mathbb{R} . Suppose that H_N is differentiable at 0. Define

$$DH_N = \max_{k=1, \dots, K_N} \sum_{i=1}^{K_N} \max_{j=1, \dots, N} \sum_{n=1}^N \left| \frac{\partial H_N^{(k, n)}}{\partial u(i, j)}(0) \right|. \quad (30)$$

We have that

Lemma 5: Assume that:

L5.1 for each N , (17) holds; and

L5.2 for each N , $H_N \in C^1$ locally at 0.

Then we have that for each N , $\mu_N \leq DH_N$.

Proof: Let J_N be the jacobian matrix of H_N at 0. Note that $J_N \in \mathbb{R}^{K_N N \times K_N N}$. Let $J_N(l, m)$ be its (l, m) th component, where $l, m = 1, \dots, K_N N$. Then we have that for $k, i = 1, \dots, K_N$ and $n, j = 1, \dots, N$,

$$\frac{\partial H_N^{(k, n)}}{\partial u(i, j)}(0) = J_N((k-1)N+n, (i-1)N+j).$$

Let $C_N(k, i)$ be the matrix in $\mathbb{R}^{N \times N}$ such that for $n, j = 1, \dots, N$, the (n, j) th component of $C_N(k, i)$ is

$$\frac{\partial H_N^{(k, n)}}{\partial u(i, j)}(0);$$

i.e., $C_N(k, i)$ is the (k, i) th block in the partition of J_N into $N \times N$ blocks (there are $K_N \times K_N$ such blocks), where $k, i = 1, \dots, K_N$. Then by (30),

$$DH_N = \max_{k=1, \dots, K_N} \sum_{i=1}^{K_N} \|C_N(k, i)\|_1^{(N)}. \quad (31)$$

($\|\cdot\|_1^{(N)}$ is defined by (23).)

By (19) and (22), for $u = [u(1)^\top, \dots, u(K_N)^\top]^\top \in \mathbb{R}^{K_N N}$, where $u(k) = [u(k, 1), \dots, u(k, N)]^\top \in \mathbb{R}^N$,

$$\begin{aligned} \|J_N u\|^{(N)} &= ds_N \max_{k=1, \dots, K_N} \left\| \sum_{i=1}^{K_N} C_N(k, i) u(i) \right\|_1^{(N)} \\ &\leq ds_N \max_{k=1, \dots, K_N} \sum_{i=1}^{K_N} \|C_N(k, i) u(i)\|_1^{(N)} \\ &\leq ds_N \max_{k=1, \dots, K_N} \sum_{i=1}^{K_N} \|C_N(k, i)\|_1^{(N)} \|u(i)\|_1^{(N)} \\ &\leq \max_{k=1, \dots, K_N} \sum_{i=1}^{K_N} \|C_N(k, i)\|_1^{(N)} ds_N \max_{l=1, \dots, K_N} \|u(l)\|_1^{(N)} \\ &= DH_N \|u\|^{(N)}, \end{aligned}$$

where the last equation follows from (31), (19), and (22). Therefore, for $u \neq 0$,

$$DH_N \geq \frac{\|J_N u\|^{(N)}}{\|u\|^{(N)}}. \quad (32)$$

Note that if $u_N = 0$, then by (16) and (17), $\varepsilon_N = 0$. Therefore

$$H_N(0) = 0. \quad (33)$$

By Assumption L5.2 and Taylor's theorem, there exists a function \tilde{H}_N such that

$$H_N(u) = J_N u + \tilde{H}_N(u), \quad (34)$$

and

$$\lim_{\alpha \rightarrow 0} \sup_{\|u\|^{(N)} \leq \alpha} \frac{\|\tilde{H}_N(u)\|^{(N)}}{\|u\|^{(N)}} = 0. \quad (35)$$

By (34) and the triangle inequality, we have that

$$\|H_N(u)\|^{(N)} \leq \|J_N u\|^{(N)} + \|\tilde{H}_N(u)\|^{(N)}.$$

Therefore by (20),

$$\mu_N \leq \lim_{\alpha \rightarrow 0} \sup_{\|u\|^{(N)} \leq \alpha} \left(\frac{\|J_N u\|^{(N)}}{\|u\|^{(N)}} + \frac{\|\tilde{H}_N(u)\|^{(N)}}{\|u\|^{(N)}} \right).$$

Hence by (32) and (35), we complete the proof. ■

Next we present in Lemma 6 a relationship between f_N and DH_N . Define for each N and for $k, l = 1, \dots, K_N$,

$$B_N^{(k,l)} = \begin{cases} A_N(k-1)A_N(k-2)\dots A_N(l), & 1 \leq l < k; \\ I_N, & l = k; \\ 0, & l > k, \end{cases} \quad (36)$$

where $A_N(l)$ is as defined by (21). We have that

Lemma 6: Assume that:

L6.1 for each N , (17) holds; and

L6.2 for each N , $f_N \in \mathcal{C}^1$.

Then we have that for each N , for $k, i = 1, \dots, K_N$ and $n, j = 1, \dots, N$,

$$\frac{\partial H_N^{(k,n)}}{\partial u(i,j)}(0) = B_N^{(k,i)}(n, j) dt_N.$$

Proof: By Assumption L6.2 and Taylor's theorem, for fixed z , there exists a function \tilde{f}_N such that

$$\begin{aligned} f_N(x_N(k)) - f_N(z_N(k)) &= Df_N(z_N(k)) \varepsilon_N(k) \\ &\quad + t l d f_N(z_N(k) + \varepsilon_N(k), z_N(k)), \end{aligned}$$

and for each z ,

$$\tilde{f}_N(z, z) = 0, \quad (37)$$

and

$$\lim_{\|\varepsilon\|^{(N)} \rightarrow 0} \frac{\|\tilde{f}_N(z + \varepsilon, z)\|^{(N)}}{\|\varepsilon\|^{(N)}} = 0. \quad (38)$$

Then we have from (16) that for $k = 0, \dots, K_N - 1$,

$$\begin{aligned} \varepsilon_N(k+1) &= \varepsilon_N(k) + \frac{1}{M_N} Df_N(z_N(k)) \varepsilon_N(k) \\ &\quad + \frac{1}{M_N} \tilde{f}_N(z_N(k) + \varepsilon_N(k), z_N(k)) + dt_N u_N(k). \end{aligned}$$

Therefore

$$\begin{aligned} \varepsilon_N(k+1) &= A_N(k) \varepsilon_N(k) + dt_N u_N(k) \\ &\quad + \frac{\tilde{f}_N(z_N(k) + \varepsilon_N(k), z_N(k))}{M_N}. \end{aligned}$$

For $k = 0, \dots, K_N - 1$, define

$$\eta_N(k) = dt_N u_N(k) + \frac{\tilde{f}_N(z_N(k) + \varepsilon_N(k), z_N(k))}{M_N}. \quad (39)$$

Then $\varepsilon_N(k+1) = A_N(k) \varepsilon_N(k) + \eta_N(k)$. Therefore for $k = 1, \dots, K_N$,

$$\begin{aligned} \varepsilon_N(k) &= A_N(k-1) \dots A_N(1) \eta_N(0) \\ &\quad + A_N(k-1) \dots A_N(2) \eta_N(1) \\ &\quad + \dots + A_N(k-1) \eta_N(k-2) + \eta_N(k-1). \end{aligned}$$

Then it follows from (36) that for $k = 1, \dots, K_N$,

$$\varepsilon_N(k) = \sum_{l=1}^k B_N^{(k,l)} \eta_N(l-1). \quad (40)$$

Write $\varepsilon_N(k) = H_N^{(k)}(u_N)$. By (39),

$$\eta_N(k) = dt_N u_N(k) + \frac{\tilde{f}_N(z_N(k) + H_N^{(k)}(u_N), z_N(k))}{M_N}.$$

Hence by (40), for $k = 1, \dots, K_N$,

$$\begin{aligned} \varepsilon_N(k) &= \sum_{l=1}^k B_N^{(k,l)} dt_N u_N(l-1) \\ &\quad + \sum_{l=1}^k B_N^{(k,l)} \frac{\tilde{f}_N(z_N(l-1) + H_N^{(l-1)}(u_N), z_N(l-1))}{M_N}. \end{aligned}$$

Denote by $g_N^{(k,l,n)}(\cdot) : \mathbb{R}^{K_N N} \rightarrow \mathbb{R}^N$ the n th component of

$$B_N^{(k,l)} \tilde{f}_N(z_N(l-1) + H_N^{(l-1)}(\cdot), z_N(l-1)).$$

By (37) and (33), $g_N^{(k,l,n)}(0) = 0$.

Let $\{e(i, j) : i = 1, \dots, K_N, j = 1, \dots, N\}$ be the standard basis for $\mathbb{R}^{K_N N}$; i.e., $e(i, j)$ is the element of $\mathbb{R}^{K_N N}$ with the (i, j) th entry being 1 and all other entries being 0. Then

$$\begin{aligned} \frac{\partial H_N^{(k,n)}}{\partial u(i,j)}(0) &= B_N^{(k,i)}(n, j) dt_N \\ &+ \frac{1}{M_N} \sum_{l=1}^k \left(\lim_{h \rightarrow 0} \frac{g_N^{(k,l,n)}(h e(i,j))}{h} \right). \end{aligned}$$

It remains to show that

$$\lim_{h \rightarrow 0} \frac{g_N^{(k,l,n)}(h e(i,j))}{h} = 0.$$

Denote by $\theta_N^{(l,d)}(\cdot) : \mathbb{R}^{K_N N} \rightarrow \mathbb{R}$ the d th component of $\tilde{f}_N(z_N(l) + H_N^{(l)}(\cdot), z_N(l))$. Then

$$g_N^{(k,l,n)}(u) = \sum_{d=1}^N B_N^{(k,l)}(n, d) \theta_N^{(l-1,d)}(u).$$

Denote by $\tilde{f}_N^{(l,d)}(\cdot) : \mathbb{R}^N \rightarrow \mathbb{R}$ the d th component of $\tilde{f}_N(z_N(l) + (\cdot), z_N(l))$. Then

$$\theta_N^{(l,d)}(u) = \tilde{f}_N^{(l,d)}(H_N^{(l)}(u)). \quad (41)$$

Then it remains to show that

$$\lim_{\|u\|^{(N)} \rightarrow 0} \frac{\theta_N^{(l,d)}(u)}{\|u\|^{(N)}} = 0. \quad (42)$$

By Assumption L6.2 and by induction, it follows from (16) that for fixed z , ε_N is a C^1 function of u_N , because the composition of functions in C^1 is still in C^1 . Hence Assumption L6.2 here implies Assumption L5.2 of Lemma 5. By Assumption L5.2 and (33), there exists c such that $|c| < \infty$, and for each $\varepsilon_1 > 0$, there exists $\delta_1(\varepsilon_1)$ such that for $\|u\|^{(N)} < \delta_1(\varepsilon_1)$,

$$\left| \frac{\|H_N^{(l)}(u)\|^{(N)}}{\|u\|^{(N)}} - c \right| < \varepsilon_1. \text{ Hence for } \|u\|^{(N)} < \delta_1(\varepsilon_1),$$

$$\|H_N^{(l)}(u)\|^{(N)} < (|c| + \varepsilon_1) \|u\|^{(N)}. \quad (43)$$

By (38), $\lim_{\|x\|^{(N)} \rightarrow 0} \frac{\tilde{f}_N^{(l,d)}(x)}{\|x\|^{(N)}} = 0$. Hence for each $\varepsilon_2 > 0$, there exists $\delta_2(\varepsilon_2)$ such that for $\|x\|^{(N)} < \delta_2(\varepsilon_2)$,

$$\left| \tilde{f}_N^{(l,d)}(x) \right| < \frac{\varepsilon_2}{|c| + 1} \|x\|^{(N)}. \quad (44)$$

For each ε , let $\hat{\varepsilon}(\varepsilon)$ be sufficiently small such that

$$(|c| + \hat{\varepsilon}(\varepsilon)) \delta_1(\hat{\varepsilon}(\varepsilon)) < \delta_2(\varepsilon), \quad (45)$$

and

$$\hat{\varepsilon}(\varepsilon) < 1. \quad (46)$$

Then by (43) and (45), for $\|u\|^{(N)} < \delta_1(\hat{\varepsilon}(\varepsilon))$, $\|H_N^{(l)}(u)\|^{(N)} < \delta_2(\varepsilon)$. Therefore, in the case that $\|H_N^{(l)}(u)\|^{(N)} > 0$, by (41) and (44),

$$\left| \theta_N^{(l,d)}(u) \right| = \left| \tilde{f}_N^{(l,d)}(H_N^{(l)}(u)) \right| < \frac{\varepsilon}{|c| + 1} \|H_N^{(l)}(u)\|^{(N)}.$$

By (43) and (46),

$$\left\| H_N^{(l)}(u) \right\|^{(N)} < (|c| + \hat{\varepsilon}(\varepsilon)) \|u\|^{(N)} < (|c| + 1) \|u\|^{(N)}.$$

By the above two inequalities,

$$\frac{\left| \theta_N^{(l,d)}(u) \right|}{\|u\|^{(N)}} < \varepsilon. \quad (47)$$

By (37), $\tilde{f}_N^{(l,d)}(0) = 0$. Therefore, in the case that $\|H_N^{(l)}(u)\|^{(N)} = 0$, $\theta_N^{(l,d)}(u) = 0$, and thus (47) still holds. Therefore, (42) holds. ■

Now we prove Theorem 2 using the preceding lemmas.

Proof of Theorem 2: By (30), Lemma 5, and Lemma 6, we have that

$$\begin{aligned} \mu_N &\leq \max_{k=1,\dots,K_N} \sum_{i=1}^{K_N} \max_{j=1,\dots,N} \sum_{n=1}^N \left| B_N^{(k,i)}(n, j) \right| dt_N \\ &= \max_{k=1,\dots,K_N} \sum_{i=1}^{K_N} \left\| B_N^{(k,i)} \right\|_1^{(N)} dt_N \\ &\leq \max_{k=1,\dots,K_N} K_N \max_{i=1,\dots,K_N} \left\| B_N^{(k,i)} \right\|_1^{(N)} dt_N \\ &\leq T \max_{k=1,\dots,K_N} \left\| B_N^{(k,i)} \right\|_1^{(N)}. \end{aligned}$$

($\|\cdot\|_1^{(N)}$ is defined by (23).) Therefore, by (36) and by the sub-multiplicative property of induced norm, we have that

$$\begin{aligned} \mu_N &\leq T \max_{\substack{k=1,\dots,K_N \\ i=1,\dots,k-1}} \|A_N(k-1) A_N(k-2) \dots A_N(i)\|_1^{(N)} \\ &\leq T \max_{\substack{k=1,\dots,K_N \\ i=1,\dots,k-1}} \|A_N(k-1)\|_1^{(N)} \dots \|A_N(i)\|_1^{(N)}. \end{aligned}$$

Then by Assumption T2.3, there exists $c < \infty$ such that for N sufficiently large,

$$\mu_N \leq T(1 + c dt_N)^{K_N}.$$

As $N \rightarrow \infty$, $K_N \rightarrow \infty$, and

$$(1 + c dt_N)^{K_N} = \left(1 + \frac{c T}{K_N} \right)^{K_N} \rightarrow e^{cT}.$$

Therefore $\{\mu_N\}$ is bounded. □

E. PROOF OF THEOREM 3

We now prove Theorem 3 using Theorem 1 and 2.

Proof of Theorem 3: It follows from (5) that there exists $c_1, c_2 < \infty$ such that for N sufficient large and $k = 0, \dots, K_N - 1$,

$$\begin{cases} |u_N(k, n)| < c_1, & n = 1, N; \\ |u_N(k, n)| < c_2 ds_N, & n = 2, \dots, N-1. \end{cases} \quad (48)$$

Therefore, there exists $c_3 < \infty$ such that for N sufficient large,

$$\max_{k=0,\dots,K_N-1} \sum_{n=1}^N |u_N(k, n)| < c_3,$$

and hence by (19), we have that for N sufficient large,

$$\|u_N\|^{(N)} < c_3 ds_N.$$

Hence the Assumption T1.1 of Theorem 1 holds.

By (5), for each N , for $x = [x_1, \dots, x_N]^\top \in [0, 1]^N$, the (n, m) th component of $Df_N(x)$, where $n, m = 1, \dots, N$, is

$$\left\{ \begin{array}{ll} P_l(n)x_n(1 - x_{n-1}), & m = n - 2; \\ (1 - x_n)[P_r(n-1)(1 - x_{n+1}) \\ \quad - P_l(n+1)x_{n+1}] \\ \quad + P_l(n)x_n(1 - x_{n-2}), & m = n - 1; \\ -P_r(n-1)x_{n-1}(1 - x_{n+1}) \\ \quad - P_l(n+1)x_{n+1}(1 - x_{n-1}) \\ \quad - P_r(n)(1 - x_{n+1})(1 - x_{n+2}) \\ \quad - P_l(n)(1 - x_{n-1})(1 - x_{n-2}), & m = n; \\ (1 - x_n)[P_l(n+1)(1 - x_{n-1}) \\ \quad - P_r(n-1)x_{n-1}] \\ \quad + P_r(n)x_n(1 - x_{n+2}), & m = n + 1; \\ P_r(n)x_n(1 - x_{n+1}), & m = n + 2; \\ 0 & \text{other wise,} \end{array} \right.$$

where x_n with $n \leq 0$ or $n \geq N + 1$ are defined to be zero. It then follows that for each k ,

$$\|A_N(k)\|_1^{(N)} = 1. \quad (49)$$

Hence Assumption T2.3 of Theorem 2 holds. We note that obtaining (49) and (48) requires tedious, but elementary, algebraic manipulation. One can also verify that the other assumptions of Theorem 1 and 2 hold. By Theorem 1, this completes the proof. \square

IV. CONCLUSION

In this paper we analyze the convergence of a sequence of Markov chains to its continuum limit, the solution of a PDE, in a two-step procedure. We provide precise sufficient conditions for the convergence and the explicit rate of convergence. Based on such convergence we approximate the Markov chain modeling a large wireless sensor network by a nonlinear diffusion-convection PDE.

With the well-developed mathematical tools available for PDEs, this approach provides a framework to model and simulate networks with a very large number of components, which is practically infeasible for Monte Carlo simulation. Such a tool enables us to tackle problems such as performance analysis and prototyping, resource provisioning, network design, network parametric optimization, network control, network tomography, and inverse problems, for very large networks. For example, we can now use the PDE model to optimize certain performance metrics (e.g., throughput) of a large network by adjusting the placement of destination nodes or the routing parameters (e.g., coefficients in convection terms), with relatively negligible computation overhead compared with that of the same task done by Monte Carlo simulations.

For simplicity, we have treated sequences of grid points that are uniformly located. As with finite difference methods for differential equations, the convergence results can be extended to models that have nonuniform points spacing under assumptions that insure the points in the sequence should become dense in the underlying domain uniformly in

the limit. For example, we could consider a double sequence of minimum point spacing $\{h_i\}$ and maximum point spacing $\{H_i\}$ with $H_i/h_i = \text{constant}$, and for each i , we can consider a model with nonhomogeneous point spacing between h_i and H_i . We can also introduce a spatial change of variables that maps a nonuniform model to a uniform model. This changes the coefficients in the resulting PDE, by substitution and the chain rule. In this way we can extend our approach to nonuniform, even mobile, networks. We can further consider the control of nodes such that global characteristics of the network are invariant under node locations and mobility. (See our paper [106], [111] for details.)

The assumption made in (24) that the probabilities of transmission behave continuously insures that there is a limiting behavior in the limit of large numbers of nodes and relates the behavior of networks with different numbers of nodes. The convergence results can be extended to the situation in which the probabilities change discontinuously at a finite number of lower dimensional linear manifolds (e.g., points in one dimension, lines in two dimensions, planes in three dimensions) in space provided that all of the discrete networks under consideration have nodes on the manifolds of discontinuity.

There are other considerations regarding the network that can significantly affect the derivation of the continuum model. For example, transmissions could happen beyond immediate nodes, and the interference between nodes could behave differently in the presence of power control; we can consider more boundary conditions other than sinks, including walls, semi-permeating walls, and their composition; and we can seek to establish continuum models for other domains such as the Internet, cellular networks, traffic networks, and human crowds.

REFERENCES

- [1] M. Guizani, A. Rayes, B. Khan, and A. Al-Fuqaha, *Network Modeling and Simulation: A Practical Perspective*. New York, NY, USA: Wiley, 2010.
- [2] Y. Zhang, E. Chong, J. Hannig, and D. Estep, “On continuum limits of Markov chains and network modeling,” in *Proc. 49th IEEE Conf. CDC*, Dec. 2010, pp. 6779–6784.
- [3] E. K. P. Chong, D. Estep, and J. Hannig, “Continuum modeling of large networks,” *Int. J. Numer. Model., Electron. Netw., Devices Fields*, vol. 21, no. 3, pp. 169–186, 2008.
- [4] K. Wong, C. Gin, R. Rajeswari, W. Woon, and E. Chong, “Feasibility of a PDE-based teletraffic model for cellular networks,” in *Proc. ICIAS*, Nov. 2007, pp. 383–388.
- [5] N. Burch, E. Chong, D. Estep, and J. Hannig, “Analysis of routing protocols and interference-limited communication in large wireless networks,” *J. Eng. Math.*, vol. 79, no. 1, pp. 183–199, 2012.
- [6] G. R. Liu and S. S. Quek, *The Finite Element Method: A Practical Course*. Oxford, U.K.: Butterworth-Heinemann, 2003.
- [7] A. R. Mitchell and D. F. Griffiths, *The Finite Difference Method in Partial Differential Equations*. New York, NY, USA: Wiley, 1980.
- [8] M. Burger, P. Markowich, and J. Pietschmann, “Continuous limit of a crowd motion and herding model: Analysis and numerical simulations,” *Kinetic Rel. Methods*, vol. 4, no. 4, pp. 1025–1047, 2011.
- [9] S. Bhatt, R. Fujimoto, A. Ogielski, and K. Perumalla, “Parallel simulation techniques for large-scale networks,” *IEEE Commun. Mag.*, vol. 36, no. 8, pp. 42–47, Aug. 1998.
- [10] R. M. Fujimoto, *Parallel and Distributed Simulation Systems*. New York, NY, USA: Wiley, 2000.
- [11] R. M. Fujimoto, K. Perumalla, A. Park, H. Wu, M. H. Ammar, and G. F. Riley, “Large-scale network simulation—How big? How fast?” in *Proc. Symp. MASCOTS*, 2003, pp. 1–8.

- [12] B. Szymanski, Y. Liu, and R. Gupta, "Parallel network simulation under distributed genesis," in *Proc. 17th Workshop Parallel Distrib. Simul.*, Jun. 2003, pp. 61–68.
- [13] M. Villén-Altamirano and J. Villén-Altamirano, "On the efficiency of RESTART for multidimensional state systems," *ACM Trans. Model. Comput. Simul.*, vol. 16, pp. 251–279, Jul. 2006.
- [14] P. M. Papazoglou, D. A. Karas, and R. C. Papademetriou, "An efficient scheduling mechanism for simulating concurrent events in wireless communications based on an improved priority queue (PQ) TDM layered multi-threading approach," *WSEAS Trans. Commun.*, vol. 7, no. 3, pp. 207–215, Mar. 2008.
- [15] D. Anick, D. Mitra, and M. Sondhi, "Stochastic theory of a data handling system with multiple sources," *Bell Syst. Tech. J.*, vol. 61, no. 8, pp. 1871–1894, Oct. 1982.
- [16] G. Kesidis, A. Singh, D. Cheung, and W. Kwok, "Feasibility of fluid event-driven simulation for ATM networks," in *Proc. Commun., Key Global Prosperity, Global Telecommun. Conf.*, vol. 3, Nov. 1996, pp. 2013–2017.
- [17] K. Kumaran and D. Mitra, "Performance and fluid simulations of a novel shared buffer management system," in *Proc. IEEE 17th Annu. Joint Conf. Comput. Commun. Soc.*, vol. 3, Mar./Apr. 1998, pp. 1449–1461.
- [18] B. Liu, D. Figueiredo, Y. Guo, J. Kurose, and D. Towsley, "A study of networks simulation efficiency: Fluid simulation vs. packet-level simulation," in *Proc. 20th Annu. Joint Conf. IEEE Comput. Commun. Soc.*, vol. 3, Apr. 2001, pp. 1244–1253.
- [19] H. J. Kushner, *Approximation and Weak Convergence Methods for Random Processes, with Applications to Stochastic Systems Theory*. Cambridge, MA, USA: MIT Press, 1984.
- [20] H. Robbins and S. Monro, "A stochastic approximation method," *Ann. Math. Stat.*, vol. 22, no. 3, pp. 400–407, 1951.
- [21] J. Kiefer and J. Wolfowitz, "Stochastic estimation of the maximum of a regression function," *Ann. Math. Stat.*, vol. 23, pp. 462–466, 1952.
- [22] M. Benaim, "Dynamics of stochastic approximation algorithms," in *Séminaire de Probabilités XXXIII* (Lecture Notes in Mathematics), vol. 1709. Berlin/Heidelberg, Germany: Springer-Verlag, 1999, pp. 1–68.
- [23] T. L. Lai, "Stochastic approximation: Invited paper," *Ann. Stat.*, vol. 31, no. 2, pp. 391–406, 2003.
- [24] L. Ljung, "Analysis of recursive stochastic algorithms," *IEEE Trans. Autom. Control*, vol. 22, no. 4, pp. 551–575, Aug. 1977.
- [25] H. J. Kushner and D. S. Clark, *Stochastic Approximation for Constrained and Unconstrained Systems*. New York, NY, USA: Springer-Verlag, 1978.
- [26] A. Benveniste, M. MéTivier, and P. Priouret, *Stochastic Approximations and Adaptive Algorithms*. New York, NY, USA: Springer-Verlag, 1990.
- [27] M. Benaim and M. W. Hirsch, "Stochastic approximation algorithms with constant step size whose average is cooperative," *Ann. Appl. Probab.*, vol. 9, no. 1, pp. 216–241, 1999.
- [28] H. J. Kushner and G. G. Yin, *Stochastic Approximation and Recursive Algorithms and Applications*. New York, NY, USA: Springer-Verlag, 2003.
- [29] T. G. Kurtz, "Solutions of ordinary differential equations as limits of pure jump Markov processes," *J. Appl. Probab.*, vol. 7, no. 1, pp. 49–58, 1970.
- [30] R. W. R. Darling, (2002, Dec.). *Fluid Limits of Pure Jump Markov Processes: A Practical Guide* [Online]. Available: <http://arxiv.org/abs/math/0210109>
- [31] R. W. R. Darling and J. R. Norris, "Differential equation approximations for Markov chains," *Probab. Surveys*, vol. 5, pp. 37–79, 2008.
- [32] R. McVinish and P. Pollett, "The deterministic limit of a stochastic logistic model with individual variation," *Math. Biosci.*, vol. 241, no. 1, pp. 109–114, Jan. 2013.
- [33] S. N. Ethier and T. G. Kurtz, *Markov Processes: Characterization and Convergence* (Probability and Mathematical Statistics: Probability and Mathematical Statistics). New York, NY, USA: Wiley, 1986.
- [34] P. Billingsley, *Convergence of Probability Measures*. New York, NY, USA: Wiley, 1999.
- [35] H. F. Trotter, "Approximation of semi-groups of operators," *Pacific J. Math.*, vol. 8, no. 4, pp. 887–919, 1958.
- [36] D. W. Stroock and S. S. Varadhan, *Multidimensional Diffusion Processes*. Berlin, Germany: Springer-Verlag, 1979.
- [37] P. Gupta and P. R. Kumar, "The capacity of wireless networks," *IEEE Trans. Inf. Theory*, vol. 46, no. 2, pp. 388–404, Mar. 2000.
- [38] M. Grossglauser and D. N. C. Tse, "Mobility increases the capacity of ad hoc wireless networks," *IEEE/ACM Trans. Netw.*, vol. 10, no. 4, pp. 477–486, Aug. 2002.
- [39] L. L. Xie and P. Kumar, "A network information theory for wireless communication: Scaling laws and optimal operation," *IEEE Trans. Inf. Theory*, vol. 50, no. 5, pp. 748–767, May 2004.
- [40] J. Herdtner and E. Chong, "Throughput-storage tradeoff in ad hoc networks," in *Proc. 24th Annu. Joint Conf. IEEE Comput. Commun. Soc.*, vol. 4, Mar. 2005, pp. 2536–2542.
- [41] P. Jelenkovic, P. Momicilovic, and M. Squillante, "Scalability of wireless networks," *IEEE/ACM Trans. Netw.*, vol. 15, no. 2, pp. 295–308, Apr. 2007.
- [42] C. Graham, "Chaosity on path space for a queueing network with selection of the shortest queue among several," *Ann. Appl. Probab.*, vol. 37, no. 1, pp. 198–211, 2000.
- [43] M. Benaim and J.-Y. L. Boudec, "A class of mean field interaction models for computer and communication systems," *Perform. Evaluat.*, vol. 65, nos. 11–12, pp. 823–838, 2008.
- [44] R. Bakhshi, L. Cloth, W. Fokkink, and B. R. Haverkort, "Mean-field analysis for the evaluation of gossip protocols," *ACM SIGMETRICS Perform. Evaluat. Rev.*, vol. 36, pp. 31–39, Nov. 2008.
- [45] R. Bakhshi, J. Endrullis, S. Endrullis, W. Fokkink, and B. Haverkort, "Automating the mean-field method for large dynamic gossip networks," in *Proc. 7th Int. Conf. QEST*, Sep. 2010, pp. 241–250.
- [46] C. Bordenave, D. McDonald, and A. Proutière, "A particle system in interaction with a rapidly varying environment: Mean field limits and applications," *Netw. Heterogeneous Media*, vol. 5, no. 1, pp. 31–62, 2010.
- [47] H. Tembine, J.-Y. Le Boudec, R. El-Azouzi, and E. Altman, "Mean field asymptotics of Markov decision evolutionary games and teams," in *Proc. Int. Conf. Game Theory Netw.*, May 2009, pp. 140–150.
- [48] P. E. Caines, "Mean field stochastic control," in *Proc. Bode Lecture 48th IEEE Conf. Decision Control*, Shanghai, China, Dec. 2009.
- [49] M. Nourian, R. Malhamé, M. Huang, and P. Caines, "Optimality of adaption based mean field control laws in leader-follower stochastic collective dynamics," in *Proc. 49th IEEE Conf. CDC*, Dec. 2010, pp. 2270–2275.
- [50] M. Nourian, P. Caines, and R. Malhamé, "Synthesis of Cucker-Smale type flocking via mean field stochastic control theory: Nash equilibria," in *Proc. 48th Annu. Allerton Conf. Commun., Control, Comput.*, Sep./Oct. 2010, pp. 814–819.
- [51] N. Gast, B. Gaujal, and J.-Y. L. Boudec, "Mean field for Markov decision processes: From discrete to continuous optimization," *IEEE Trans. Autom. Control*, vol. 57, no. 9, pp. 2266–2280, Sep. 2012.
- [52] O. Zeitouni and M. Gutman, "On universal hypotheses testing via large deviations," *IEEE Trans. Inf. Theory*, vol. 37, no. 2, pp. 285–290, Mar. 1991.
- [53] Y. Amit and M. Miller, "Large deviations for coding Markov chains and Gibbs random fields," *IEEE Trans. Inf. Theory*, vol. 39, no. 1, pp. 109–118, Jan. 1993.
- [54] A. Weiss, "An introduction to large deviations for communication networks," *IEEE J. Sel. Areas Commun.*, vol. 13, no. 6, pp. 938–952, Aug. 1995.
- [55] I. Paschalidis and Y. Chen, "Anomaly detection in sensor networks based on large deviations of Markov chain models," in *Proc. 47th IEEE CDC*, Dec. 2008, pp. 2338–2343.
- [56] I. Paschalidis and G. Smaragdakis, "Spatio-temporal network anomaly detection by assessing deviations of empirical measures," *IEEE/ACM Trans. Netw.*, vol. 17, no. 3, pp. 685–697, Jun. 2009.
- [57] S. L. Sobolev, *Partial Differential Equations of Mathematical Physics*. New York, NY, USA: Dover, 1964.
- [58] D. Blount, "Law of large numbers in the supremum norm for a chemical reaction with diffusion," *Ann. Appl. Probab.*, vol. 2, no. 1, pp. 131–141, 1992.
- [59] Y. Tyutyunov, I. Senina, and R. Arditi, "Clustering due to acceleration in the response to population gradient: A simple self-organization model," *Amer. Naturalist*, vol. 164, no. 6, pp. 722–735, 2004.
- [60] C. F. Huang and H. Pages, "Optimal consumption and portfolio policies with an infinite horizon: Existence and convergence," *Ann. Appl. Probab.*, vol. 2, no. 1, pp. 36–64, 1992.
- [61] G. L. Xu and S. E. Shreve, "A duality method for optimal consumption and investment under short-selling prohibition. II. Constant market coefficients," *Ann. Appl. Probab.*, vol. 2, no. 2, pp. 314–328, 1992.
- [62] W. F. Phillips, *A New Continuum Model for Traffic Flow*. Washington, DC, USA: U.S. Dept. of Transportation, Research and Special Programs Administration, 1981.
- [63] P. Kachroo, *Pedestrian Dynamics Feedback Control of Crowd Evacuation*. New York, NY, USA: Springer-Verlag, 2008.
- [64] C. B. Morrey, "On the derivation of the equations of hydrodynamics from statistical mechanics," *Proc. Nat. Acad. Sci. USA*, vol. 40, no. 5, pp. 317–322, 1954.

- [65] R. L. Dobrushin and R. Siegmund-Schultze, "The hydrodynamic limit for systems of particles with independent evolution," *Math. Nachrichten*, vol. 105, no. 1, pp. 199–224, 1982.
- [66] M. Z. Guo, G. C. Papanicolaou, and S. R. S. Varadhan, "Nonlinear diffusion limit for a system with nearest neighbor interactions," *Commun. Math. Phys.*, vol. 118, no. 1, pp. 31–59, 1988.
- [67] C. Kipnis and C. Landim, *Scaling Limits of Interacting Particle Systems*. New York, NY, USA: Springer-Verlag, 1999.
- [68] A. Koukous, "Hydrodynamic behavior of symmetric zero-range processes with random rates," *Stochastic Process. Appl.*, vol. 84, no. 2, pp. 297–312, 1999.
- [69] C. Bahadoran, H. Guiol, K. Ravishankar, and E. Saada, "Euler hydrodynamics of one-dimensional attractive particle systems," *Ann. Probab.*, vol. 34, no. 4, pp. 1339–1369, 2006.
- [70] H. Chen and A. Mandelbaum, "Discrete flow networks: Bottleneck analysis and fluid approximations," *Math. Oper. Res.*, vol. 16, no. 2, pp. 408–446, 1991.
- [71] H. Chen, "Fluid approximations and stability of multiclass queueing networks: Work-conserving disciplines," *Ann. Appl. Probab.*, vol. 5, no. 3, pp. 637–665, 1995.
- [72] J. G. Dai, "On positive Harris recurrence of multiclass queueing networks: A unified approach via fluid limit models," *Ann. Appl. Probab.*, vol. 5, no. 1, pp. 49–77, 1995.
- [73] J. G. Dai, "A fluid limit model criterion for instability of multiclass queueing networks," *Ann. Appl. Probab.*, vol. 6, no. 3, pp. 751–757, 1996.
- [74] S. P. Meyn, "Transience of multiclass queueing networks via fluid limit models," *Ann. Appl. Probab.*, vol. 5, no. 4, pp. 946–957, 1995.
- [75] A. Eryilmaz and R. Srikant, "Fluid models of wireless networks," in *Proc. SCSC*, Montreal, QC, Canada, Aug. 2003, pp. 1–5.
- [76] F. P. Kelly and R. J. Williams, "Fluid model for a network operating under a fair bandwidth-sharing policy," *Ann. Appl. Probab.*, vol. 14, no. 3, pp. 1055–1083, 2004.
- [77] M. Lelarge, "Fluid limit of generalized Jackson queueing networks with stationary and ergodic arrivals and service times," *J. Appl. Probab.*, vol. 42, no. 2, pp. 491–512, 2005.
- [78] R. Egorova, S. Borst, and B. Zwart, "Bandwidth-sharing networks in overload," *Perform. Evaluat.*, vol. 64, pp. 978–993, Oct. 2007.
- [79] H. C. Gromoll and R. J. Williams, "Fluid limits for networks with bandwidth sharing and general document size distributions," *Ann. Appl. Probab.*, vol. 19, no. 1, pp. 243–280, 2009.
- [80] D. L. Iglehart, "Multiple channel queues in heavy traffic," *Probab. Theory Rel. Fields*, vol. 17, no. 2, pp. 168–180, 1971.
- [81] J. M. Harrison, "A limit theorem for priority queues in heavy traffic," *J. Appl. Probab.*, vol. 10, no. 4, pp. 907–912, 1973.
- [82] G. Foschini and J. Salz, "A basic dynamic routing problem and diffusion," *IEEE Trans. Commun.*, vol. 26, no. 3, pp. 320–327, Mar. 1978.
- [83] S. Halfin and W. Whitt, "Heavy-traffic limits for queues with many exponential servers," *Oper. Res.*, vol. 29, no. 3, pp. 567–588, 1981.
- [84] J. M. Harrison and M. I. Reiman, "Reflected Brownian motion on an orthant," *Ann. Probab.*, vol. 9, no. 2, pp. 302–308, 1981.
- [85] W. Whitt, "On the heavy-traffic limit theorem for $GI/G/\infty$ queues," *Adv. Appl. Probab.*, vol. 14, no. 1, pp. 171–190, 1982.
- [86] J. M. Harrison and R. J. Williams, "Brownian models of feedforward queueing networks: Quasireversibility and product form solutions," *Ann. Appl. Probab.*, vol. 2, no. 2, pp. 263–293, 1992.
- [87] J. M. Harrison and R. J. Williams, "A multiclass closed queueing network with unconventional heavy traffic behavior," *Ann. Appl. Probab.*, vol. 6, no. 1, pp. 1–47, 1996.
- [88] J. M. Harrison, "Heavy traffic analysis of a system with parallel servers: Asymptotic optimality of discrete-review policies," *Ann. Appl. Probab.*, vol. 8, no. 3, pp. 822–848, 1998.
- [89] J. M. Harrison, "Brownian models of open processing networks: Canonical representation of workload," *Ann. Appl. Probab.*, vol. 10, no. 1, pp. 75–103, 2000.
- [90] A. Budhiraja and A. P. Ghosh, "Diffusion approximations for controlled stochastic networks: An asymptotic bound for the value function," *Ann. Appl. Probab.*, vol. 16, no. 4, pp. 1962–2006, 2006.
- [91] M. Bramson, J. G. Dai, and J. M. Harrison, "Positive recurrence of reflecting Brownian motion in three dimensions," *Ann. Appl. Probab.*, vol. 20, no. 2, pp. 753–783, 2010.
- [92] H. Chen and J. Shanthikumar, "Fluid limits and diffusion approximations for networks of multi-server queues in heavy traffic," *Discrete Event Dyn. Syst.*, vol. 4, no. 3, pp. 269–291, 1994.
- [93] V. Nguyen, "Fluid and diffusion approximations of a two-station mixed queueing network," *Math. Oper. Res.*, vol. 20, no. 2, pp. 321–354, 1995.
- [94] W. Chen, D. Huang, A. Kulkarni, J. Unnikrishnan, Q. Zhu, P. Mehta, S. Meyn, and A. Wierman, "Approximate dynamic programming using fluid and diffusion approximations with applications to power management," in *Proc. 48th IEEE CDC/CCC*, Dec. 2009, pp. 3575–3580.
- [95] E. W. Grundke and A. N. Zincir-Heywood, "A uniform continuum model for scaling of ad hoc networks," in *Proc. 2nd Int. Conf. ADHOC-NOW*, 2003, pp. 96–103.
- [96] P. Jacquet, "Geometry of information propagation in massively dense ad hoc networks," in *Proc. 5th ACM Int. Symp. Mobile Ad Hoc Netw. Comput.*, 2004, pp. 157–162.
- [97] B. Sirkeci-Mergen, A. Scaglione, and G. Mergen, "Asymptotic analysis of multistage cooperative broadcast in wireless networks," *IEEE Trans. Inf. Theory*, vol. 52, no. 6, pp. 2531–2550, Jun. 2006.
- [98] S. Tournipas and L. Tassiulas, "Optimal deployment of large wireless sensor networks," *IEEE Trans. Inf. Theory*, vol. 52, no. 7, pp. 2935–2953, Jul. 2006.
- [99] E. Altman, A. Silva, P. Bernhard, and M. Debbah, "Continuum equilibria for routing in dense ad-hoc networks," in *Proc. 45th Annu. Allerton Conf. Commun., Control, Comput.*, 2007, pp. 1–31.
- [100] J.-M. Kelif, M. Coupechoux, and P. Godlewski, "Spatial outage probability for cellular networks," in *Proc. IEEE Global Telecommun. Conf.*, Nov. 2007, pp. 4445–4450.
- [101] C. F. Chiasserini, R. Gaeta, M. Garetto, M. Gribaudo, D. Manini, and M. Sereno, "Fluid models for large-scale wireless sensor networks," *Perform. Evaluat.*, vol. 64, nos. 7–8, pp. 715–736, 2007.
- [102] V. Blondel, J. Hendrickx, and J. Tsitsiklis, "On Krause's multi-agent consensus model with state-dependent connectivity," *IEEE Trans. Autom. Control*, vol. 54, no. 11, pp. 2586–2597, Nov. 2009.
- [103] V. Misra, W. B. Gong, and D. Towsley, "Fluid-based analysis of a network of AQM routers supporting TCP flows with an application to RED," in *Proc. Conf. Appl., Technol., Archit., Protocols Comput. Commun.*, 2000, pp. 151–160.
- [104] L. Liu, P. Parag, J. Tang, W. Y. Chen, and J. F. Chamberland, "Resource allocation and quality of service evaluation for wireless communication systems using fluid models," *IEEE Trans. Inf. Theory*, vol. 53, no. 5, pp. 1767–1777, May 2007.
- [105] J. Lee, S. Bohacek, J. Hespanha, and K. Obraczka, "Modeling communication networks with hybrid systems," *IEEE/ACM Trans. Netw.*, vol. 15, no. 3, pp. 630–643, Jun. 2007.
- [106] Y. Zhang, E. K. P. Chong, J. Hannig, and D. Estep, "Continuum modeling and control of large nonuniform networks," in *Proc. 49th Annu. Allerton Conf. Commun., Control, Comput.*, 2011, pp. 1670–1677.
- [107] M. Neely, E. Modiano, and C. Rohrs, "Dynamic power allocation and routing for time-varying wireless networks," *IEEE J. Sel. Areas Commun.*, vol. 23, no. 1, pp. 89–103, Jan. 2005.
- [108] M. K. Jain, *Numerical Solution of Differential Equations*. New York, NY, USA: Wiley, 1984.
- [109] E. Isaacson and H. B. Keller, *Analysis of Numerical Methods*. New York, NY, USA: Wiley, 1966.
- [110] P. Billingsley, *Probability and Measure*. New York, NY, USA: Wiley, 1995.
- [111] Y. Zhang, E. K. P. Chong, J. Hannig, and D. Estep, "Continuum modeling and control of large nonuniform wireless networks via nonlinear partial differential equations," *Abstract Appl. Anal.*, vol. 2013, pp. 262581-1–262581-16, Mar. 2013.



YANG ZHANG received the B.S. and M.S. degrees in electrical engineering from Xi'an Jiaotong University, Xi'an, China, in 2004 and 2008, respectively. He is currently pursuing the Ph.D. degree in electrical engineering from Colorado State University, Fort Collins, CO, USA. His current research interests include network modeling, stochastic analysis, simulation, optimization, and control.



EDWIN K. P. CHONG (F'04) received the B.E. (Hons.) degree from the University of Adelaide, Adelaide, South Australia, in 1987 and the M.A. and Ph.D. degrees from Princeton University, Princeton, NJ, USA, in 1989 and 1991, respectively, where he held an IBM Fellowship. He was with the School of Electrical and Computer Engineering, Purdue University, West Lafayette, IN, USA, in 1991, where he was named a University Faculty Scholar in 1999. Since August 2001, he has been a Professor of electrical and computer engineering and Professor of mathematics from Colorado State University, Fort Collins, CO, USA. His current research interests include stochastic modeling and control, optimization methods, and communication and sensor networks. He coauthored the best-selling book, *An Introduction to Optimization* (Fourth Edition, Wiley-Interscience, 2013). He received the NSF CAREER Award in 1995 and the ASEE Frederick Emmons Terman Award in 1998. He was a co-recipient of the 2004 Best Paper Award for a paper in the journal *Computer Networks*. In 2010, he received the IEEE Control Systems Society Distinguished Member Award.

Prof. Chong was the Founding Chairman of the IEEE Control Systems Society Technical Committee on Discrete Event Systems, and served as an IEEE Control Systems Society Distinguished Lecturer. He is currently a Senior Editor of the IEEE Transactions on Automatic Control, and also serves on the editorial boards of *Computer Networks* and the *Journal of Control Science and Engineering*. He served as a member of the IEEE Control Systems Society Board of Governors and is currently a Vice President for Financial Activities. He has also served on the organizing committees of several international conferences. He has been on the program committees for the IEEE Conference on Decision and Control, the American Control Conference, the IEEE International Symposium on Intelligent Control, IEEE Symposium on Computers and Communications, and the IEEE Global Telecommunications Conference. He has also served in the executive committees for the IEEE Conference on Decision and Control, the American Control Conference, the IEEE Annual Computer Communications Workshop, the International Conference on Industrial Electronics, Technology and Automation, and the IEEE International Conference on Communications. He was the Conference (General) Chair for the Conference on Modeling and Design of Wireless Networks, part of SPIE ITCOM 2001. He was the General Chair for the 2011 Joint 50th IEEE Conference on Decision and Control and European Control Conference.



DONALD ESTEP received the B.A. degree from Columbia College, Columbia University, New York, NY, USA, in 1981 and the M.S. and Ph.D. degrees in applied mathematics from the University of Michigan, Ann Arbor, Michigan, USA, in 1987. He joined the faculty in the School of Mathematics with the Georgia Institute of Technology, Atlanta, GA, USA, in 1987. In 2000, he joined the faculties in the Departments of Mathematics and of Statistics at Colorado State University, Fort Collins, CO, USA. He was appointed as the Colorado State University Interdisciplinary Research Scholar in 2009. At Colorado State University, he directed the NSF IGERT Program for Interdisciplinary Mathematics, Ecology, and Statistics from 2003 to 2009, and has directed the Center for Interdisciplinary Mathematics and Statistics, since 2006. He was an NSF International Research Fellow from 1993 to 1995, won the Computational and Mathematical Methods in Sciences and Engineering Prize in 2005, and won the Pennock Distinguished Service Award in 2009 and University Scholarship Impact Award with Colorado State University in 2011. From 2013 to 2014, he holds the Chalmers Jubilee Professor honor with the Chalmers University of Technology, Gothenburg, Sweden. He is the founding Editor in Chief of the SIAM/ASA Journal on Uncertainty Quantification, Editor in Chief of the SIAM Book Series on Computational Science and Engineering in 2009, and Editor for five other journals. He served as the founding Chair of the SIAM Activity Group on Uncertainty Quantification, and currently serves on the SIAM Book Committee and on the Governing Board of the Statistical and Applied Mathematical Sciences Institute.



JAN HANNIG is a Professor in the Department of Statistics and Operations Research with the University of North Carolina, Chapel Hill, NC, USA. His current research interests include applied probability, theoretical statistics, generalized fiducial inference, and applications to biology and engineering. He received the M.gr. degree in mathematics from the Charles University, Prague, Czech Republic, in 1996. He received the Ph.D. degree in statistics and probability from Michigan State University, Ann Arbor, MI, USA, in 2000 under the direction of Prof. A.V. Skorokhod. From 2000 to 2008, he was the Faculty with the Department of Statistics with Colorado State University, Fort Collins, CO, USA where he was promoted to an Associate Professor in 2006. He was with the Department of Statistics and Operations with the University of North Carolina, Chapel Hill, NC, USA, in 2008 and was promoted to Professor in 2013, he is an Associate Editor of Electronic Journal of Statistics, Stat-The ISI's Journal for the Rapid Dissemination of Statistics Research and Journal of Computational and Graphical Statistics. He is an elected member of International Statistical Institute. He was a PI and co-PI on several federally funded projects.