On Rates of Convergence for Markov Chains Under Random Time State Dependent Drift Criteria

Ramiro Zurkowski, Serdar Yüksel, *Member, IEEE*, and Tamás Linder, *Fellow, IEEE*

Abstract—**Many applications in networked control require intermittent access of a controller to a system, as in event-triggered systems or information constrained control applications. Motivated by such applications and extending previous work on Lyapunov-theoretic drift criteria, we establish both subgeometric and geometric rates of convergence for Markov chains under state dependent random time drift criteria. We quantify how the rate of ergodicity, nature of Lyapunov functions, their drift properties, and the distributions of stopping times are related. We finally study an application in networked control.**

Index Terms—**Foster-Lyapunov criteria, Markov Chain Monte-Carlo (MCMC), Markov processes, networked control systems, stochastic stability.**

I. INTRODUCTION AND LITERATURE REVIEW

STOCHASTIC stability of Markov chains has an almost complete theory, and forms a foundation for several other general techniques such as dynamic programming, linear programming approach to Markov Decision Processes [1], and Markov Chain Monte-Carlo (MCMC) [2]. One powerful approach to establish stochastic stability is through single-stage (Foster-Lyapunov) drift criteria [3]. The *state-dependent* criteria [4]–[6] relax the one-stage criteria to criteria involving time instances which are state-dependent but deterministic. Such criteria form the basis of the fluid-model (or ODE) approach to stability in stochastic networks and other general models [2], [7]–[10]. Building on [3] and [4], [11] considered stability criteria based on a state-dependent *random sampling* of the Markov chain of the following form: It was assumed that there is positive real-valued function V on the state space $\mathbb{X}$ of a discrete-time Markov chain $\{x_t\}_{t \geq 0}$, and an increasing sequence of stopping times $\{\mathcal{T}_i\}_{t \geq 0}$, with $\mathcal{T}_0 = 0$, such that for each i

$$E\left[V\left(x_{\mathcal{T}_{i+1}}\right) \mid \mathcal{F}_{\mathcal{T}_i}\right] \leq V\left(x_{\mathcal{T}_i}\right) - \delta\left(x_{\mathcal{T}_i}\right) \tag{1}$$

where the function $\delta : \mathbb{X} \to \mathbb{R}$ is positive (bounded away from zero) outside of a "small set," and $\mathcal{F}_{\mathcal{T}_i}$ denotes the filtration of "events up to time $\mathcal{T}_i$." We will make this more precise later in the paper. Further relevant work include [12] and [6].

Motivation for studying such problems comes from networked control systems and communication systems: For many networked control scenarios, access to information or application of a control action in a system is limited to random event times. As examples for such settings, there has been significant research on stochastic stabilization of networked control systems and information theory; as in stabilization of adaptive quantizers studied in source coding [13], [14] and control theory [15]–[18]. A specific example involving control over an erasure channel is given in [11], where non-zero stabilizing actions of a controller are applied to a system at certain event driven times and stochastic stability is shown using drift conditions and martingale techniques. For an extensive discussion, see [19]. The methods of random-time drift criteria can also be applied to models of networked control systems with delay-sensitive information transmission, for example, for studying the effects of randomness in the delay for transmission of sensor or controller signals (see, e.g., [20]–[23]).

One other, increasingly prominent, area is *event-triggered feedback control* systems (see, e.g., [24]–[28]) where the event instances constitute the stopping-times. The study of such systems is practically relevant since an event-based clock is usually more efficient than a time-triggered clock for control under information or actuation costs. The literature on such systems has primarily focused on the stabilization of such systems and we hope that the analysis in this paper will be useful for both stabilization and optimization of such systems: If the objective is to compute optimal solutions to an average cost optimization problem for an event triggered setup, a powerful approach is the *discounted limit approach* [29], [30]. This method typically requires geometric or sufficiently fast subgeometric convergence conditions to establish the existence of a solution to an average cost optimality equation or inequality [29]. The rate of convergence results in this paper will be useful in such contexts. Furthermore, rates of convergence to equilibrium in Markov chains are useful in bounding the distribution of transient events and the approximate computation of optimal costs under ergodicity properties. In addition, as documented extensively in the literature, Markov Chain Monte Carlo algorithms require a tedious analysis on rates of convergence bounds to obtain probabilistically guaranteed simulation times, see, e.g., [2], [31]. Furthermore, as has been discussed in [32] and [33], approximation methods for optimization of Markov Decision Processes benefit from the presence of sufficiently fast mixing/rates of convergence conditions.

In this paper, we extend recent works on random-time drift analysis [11] to obtain criteria for rates of convergence under subgeometric and geometric rate functions.

The rest of the paper is organized as follows. In Section II, we provide background information on Markov chains and rates of convergence to equilibrium. Section III contains the rate of convergence results under random-time state-dependent drift conditions. Section IV contains an example from networked control.

II. Markov Chains, Stochastic Stability, and Rates of Convergence

In this section, we review some definitions and background material relating to Markov chains and their convergence to equilibrium.

A. Preliminaries

We let $\{x_t\}_{t \geq 0}$ denote a discrete-time Markov chain with state space $\mathbb{X}$. The basic assumptions of [3] are adopted, see [34] for a more comprehensive introduction: It is assumed that $\mathbb{X}$ is a complete separable metric space; its Borel σ-field is denoted by $\mathcal{B}(\mathbb{X})$. The transition probability is denoted by P, so that for any $x \in \mathbb{X}$, $A \in \mathcal{B}(\mathbb{X})$, the probability of moving in one step from the state x to the set A is given by $P(x_{t+1} \in A | x_t = x) = P(x, A)$. The n-step transitions are obtained via composition in the usual way, $P(x_{t+n} \in A | x_t = x) = P^n(x, A)$, for any $n \geq 1$. The transition law acts on measurable functions $f : \mathbb{X} \to \mathbb{R}$ and measures μ on $\mathcal{B}(\mathbb{X})$ via $Pf(x) := \int_{\mathbb{X}} P(x, dy) f(y)$, $x \in \mathbb{X}$, and $\mu P(A) := \int_{\mathbb{X}} \mu(dx) P(x, A)$, $A \in \mathcal{B}(\mathbb{X})$. A probability measure π on $\mathcal{B}(\mathbb{X})$ is called invariant if $\pi P = \pi$, i.e.,

$$\int \pi(dx) P(x, A) = \pi(A), \qquad A \in \mathcal{B}(\mathbb{X}).$$

For any initial probability measure ν on $\mathcal{B}(\mathbb{X})$ we can construct a stochastic process $\{x_t\}$ with transition law P satisfying $x_0 \sim \nu$. We let P_ν denote the resulting probability measure on the sample space $(\mathbb{X}, \mathcal{B}(\mathbb{X}))^\infty$, with the usual convention for $\nu = \delta_x$ (where δ_x is the probability measure defined by $\delta_x(A) = 1_A(x)$ for all Borel A and $1_E(x)$ denotes the indicator function for the event $\{x \in E\}$) when the initial state is $x \in \mathbb{X}$, in which case we write P_x for the resulting probability measure. Likewise, E_x denotes the expectation operator when the initial condition is given by $x_0 = x$.

When $\nu = \pi$ (the invariant measure), the resulting process is stationary. For a set $A \in \mathcal{B}(\mathbb{X})$ we denote

$$\mathcal{T}_A := \min\{t \geq 1 : x_t \in A\}. \tag{2}$$

Definition II.1: Let φ denote a σ-finite measure on $\mathcal{B}(\mathbb{X})$. The Markov chain is called φ-irreducible if for any $x \in \mathbb{X}$, and any $B \in \mathcal{B}(\mathbb{X})$ satisfying $\varphi(B) > 0$, we have $P_x\{\mathcal{T}_B < \infty\} > 0$. A φ-irreducible Markov chain is aperiodic if for any $x \in \mathbb{X}$, and any $B \in \mathcal{B}(\mathbb{X})$ satisfying $\varphi(B) > 0$, there exists $n_0 = n_0(x, B)$ such that $P^n(x, B) > 0$ for all $n \geq n_0$. A φ-irreducible Markov chain is Harris recurrent if $P_x(\mathcal{T}_B < \infty) = 1$ for any $x \in \mathbb{X}$, and any $B \in \mathcal{B}(\mathbb{X})$ satisfying $\varphi(B) > 0$. It is *positive Harris recurrent* if in addition there is an invariant probability measure π.

A maximal irreducibility measure is one with respect to which all other irreducibility measures are absolutely continuous. Define $\mathcal{B}^+(\mathbb{X}) = \{A \in \mathcal{B}(\mathbb{X}) : \psi(A) > 0\}$, where ψ is a maximal irreducibility measure. We refer to sets in $\mathcal{B}^+(\mathbb{X})$ as *reachable*.

A set $A \in \mathcal{B}(\mathbb{X})$ is full if $\psi(A^c) = 0$ for a maximal irreducibility measure ψ. A set $A \in \mathcal{B}(\mathbb{X})$ is absorbing if $P(x, A) = 1$ for all $x \in A$. In an irreducible Markov chain every absorbing set is *full*.

Definition II.2: A set $\alpha \in \mathcal{B}^+(\mathbb{X})$ is an atom if for all $x, y \in \alpha$, $P(x, \cdot) = P(y, \cdot)$.

The concept of an atom is extremely important as it gives us a fundamental unit, where all the points of a reachable set act together. This allows, through the *cycle* equation, an invariant probability measure $\pi(A) = E_\alpha[\sum_{k=0}^{\mathcal{T}_\alpha - 1} 1_A(x_k)] / E_\alpha[\mathcal{T}_\alpha]$. When the state space is not countable, one typically needs to artificially construct such an atom, as we discuss further below.

Definition II.3: A set $C \in \mathcal{B}^+(\mathbb{X})$ is (n_0, ϵ, ν)-small if

$$P^{n_0}(x, B) \geq \epsilon \nu(B) \quad \forall B \in \mathcal{B}(\mathbb{X}), x \in C$$

where $n_0 \geq 1$, $\epsilon \in (0, 1)$, and ν is a positive measure on $(\mathbb{X}, \mathcal{B}(\mathbb{X}))$.

An important fact is that small sets exist, see [3, Theorem 5.2.1].

Fact II.1: For an irreducible Markov chain, every set $A \in \mathcal{B}^+(\mathbb{X})$ contains a small set in $\mathcal{B}^+(\mathbb{X})$.

Definition II.4: A set $C \in \mathcal{B}^+(\mathbb{X})$ is called κ-petite if there is a positive measure κ on $\mathcal{B}(\mathbb{X})$ and a probability distribution a on $\mathbb{Z}_+ = \{0, 1, 2 \ldots\}$ such that

$$\sum_{n=0}^\infty a(n) P^n(x, B) \geq \kappa(B) \quad \text{for all } B \in \mathcal{B}(\mathbb{X}), \ x \in C. \tag{3}$$

The convolution of two functions $f, g : \mathbb{Z}_+ \to \mathbb{R}$, denoted by $f * g$, is defined as usual by $f * g(n) = \sum_{k=0}^n f(k) g(n-k)$, for all $n \in \mathbb{Z}_+$. The next lemma follows from [3, Lemma 5.5.2] and allows us to assume without loss of generality that for an irreducible Markov chain, if a set is κ-petite, then κ can be replaced by maximal irreducibility measure (or equivalently κ can be assumed maximal).

Lemma II.1: If an irreducible Markov chain has some set $C \in \mathcal{B}^+(\mathbb{X})$ that is κ-petite for some distribution a, then C is ψ-petite for the distribution $a * f(n)$, where $f(n) = 2^{-n-1}$ and ψ is a maximal irreducibility measure.

An important result is the equivalence of small sets and petite sets.

Theorem II.2 ([3, Theorem 5.5.3]): For an aperiodic and irreducible Markov chain every petite set is small.

Small sets are analogous to compact sets in the stability theory for φ-irreducible Markov chains. In most applications of φ-irreducible Markov chains we find that any compact set is small—in this case, $\{x_t\}$ is called a *T-chain* [3]. The equivalence of small sets and petite sets can be used cleverly to show that all petite sets are petite for some distribution that has finite mean. The next theorem follows from [3, Propositions 5.5.5 and 5.5.6].

Theorem II.3: For an aperiodic and irreducible Markov chain every petite set is petite with a maximal irreducibility measure for a distribution with finite mean.

Invoking (3), we will use Theorem II.3 repeatedly with a set C that is κ-petite for some distribution $a(\cdot)$ to achieve bounds on hitting times for any $B \in \mathcal{B}^+(\mathbb{X})$

$$E_x \left[\sum_{k=0}^{\mathcal{T}_B - 1} 1_C(x_k) \right] \leq \frac{1}{\kappa(B)} E_x \left[\sum_{k=0}^{\mathcal{T}_B - 1} \sum_{n=1}^\infty 1_B(x_{k+n}) a(n) \right]$$

$$\leq \frac{1}{\kappa(B)} \sum_{n=0}^\infty n a(n) =: c(B) < \infty. \tag{4}$$

B. Regularity and Ergodicity

Regularity and ergodicity are concepts closely related through the work of Meyn and Tweedie [3], [4] and Tuominen and Tweedie [35]. The definitions below are in terms of functions $f : \mathbb{X} \to [1, \infty)$ and $r : \mathbb{Z}_+ \to (0, \infty)$.

Definition II.5: A set $A \in \mathcal{B}(\mathbb{X})$ is called (f, r)-regular if

$$\sup_{x \in A} E_x \left[\sum_{k=0}^{\mathcal{T}_B - 1} r(k) f(x_k) \right] < \infty$$

for all $B \in \mathcal{B}^+(\mathbb{X})$. A finite measure ν on $\mathcal{B}(\mathbb{X})$ is called (f, r)-regular if

$$E_\nu \left[\sum_{k=0}^{\mathcal{T}_B - 1} r(k) f(x_k) \right] < \infty$$

for all $B \in \mathcal{B}^+(\mathbb{X})$, and a point x is called (f, r)-regular if the measure δ_x is (f, r)-regular.

To make sense of ergodicity we first need to define the f-norm, denoted $\|.\|_f$.

Definition II.6: For a function $f : \mathbb{X} \to [1, \infty)$ the f-norm of a measure μ defined on $(\mathbb{X}, \mathcal{B}(\mathbb{X}))$ is given by

$$\|\mu\|_f = \sup_{g \leq f} \left| \int \mu(dx) g(x) \right|$$

where the supremum is taken over all measurable g such that $g(x) \leq f(x)$ for all x.

The commonly used total variation norm, or TV-norm, is the f-norm when $f = 1$, and is denoted by $\| \cdot \|_{TV}$.

Definition II.7: A Markov chain $\{x_t\}$ with invariant distribution π is (f, r)-ergodic if

$$r(n) \|P^n(x, \cdot) - \pi(\cdot)\|_f \to 0 \quad \text{as } n \to \infty \text{ for all } x \in \mathbb{X}. \quad (5)$$

If (5) is satisfied for a geometric r (so that $r(n) = M\zeta^n$ for some $\zeta > 1$, $M < \infty$) and $f = 1$ then the Markov chain $\{x_t\}$ is called geometrically ergodic.

C. The Splitting Technique and the Coupling Inequality

Nummelin's splitting technique [36] (see also [37]) is a widely used method in the study of Markov chains; see, e.g., [3, Chapter 5], [35, Proposition 3.7 and Theorem 4.1], [31, Section 4.2]. With an irreducible, aperiodic Markov chain $\{x_t\}$ on state space $\mathbb{X}$ with transition probability P and a (m, δ, ν)-small set C with finite return time, we construct an atom in order to construct an invariant distribution for the chain.

We first review the splitting technique for the case $m = 1$ (i.e., C is a $(1, \delta, \nu$-small set). Construct a new Markov chain $\{z_t\}$ on $\mathbb{X} \times \{0, 1\}$ by letting $z_t = (x_t, a_t)$, where $\{a_t\}$ is a sequence of random variables on $\{0,1\}$, independent of $\{x_t\}$, except when $x_t \in C$.

1) If $x_t \notin C$ then $x_{t+1} \sim P(x_t, \cdot)$
2) If $x_t \in C$, then
   with probability δ : $a_t = 1$ and $x_{t+1} \sim \nu(\cdot)$
   with probability $(1-\delta)$: $a_t = 0$ and $x_{t+1} \sim ((P(x_t, \cdot) - \delta\nu(\cdot))/(1 - \delta))$.

Thus the distribution of x_{t+1} given z_t is

$$P(x_{t+1} \in B \mid z_t = (x_t, a_t) \in C \times \{1\}) = \nu(B)$$
$$P(x_{t+1} \in B \mid z_t = (x_t, a_t) \in C \times \{0\}) = \frac{P(x_t, B) - \delta\nu(B)}{1 - \delta}.$$

Note that $(P(x_t, \cdot) - \delta\nu(\cdot))/(1 - \delta) \geq 0$ is a valid probability measure since C is $(1, \delta, \nu)$-small. If $x_t \in C$, then

$$x_{t+1} \sim \delta\nu(\cdot) + (1 - \delta)\frac{P(x_t, \cdot) - \delta\nu(\cdot)}{1 - \delta} = P(x_t, \cdot)$$

so the one-step transition probabilities are unchanged for $\{x_t\}$.

This construction allows one to define $S = C \times \{1\}$ as an atom for $\{z_t\}$, and to construct an invariant distribution for $\{x_t\}$ using $\{z_t\}$.

We specified the technique for the one step transition probability, but the same construction applies for (m, ϵ, ν)-small sets where $m > 1$ with the only change that the $m - 1$ steps after hitting C at x_t are distributed conditionally on x_t and x_{t+m} (see [31, Section 4.2]). When $m > 1$, the Markov chain $\{z_t\}$ does not have an atom; instead it has an "m-step atom" in the sense that $P^m((x, 1), \cdot) = P^m((y, 1), \cdot)$ for all $x, y \in C$.

A useful method to obtain bounds of convergence is through the *coupling inequality*. The coupling inequality bounds the total variation distance between the distributions of two random variables by the probability they are different. Let X, Y be two jointly distributed random variables. The following is the well known coupling inequality:

$$\|P(X \in \cdot) - P(Y \in \cdot)\|_{TV} \leq P(X \neq Y).$$

This inequality is useful in discussions of ergodicity when used in conjunction with parallel Markov chains, as in [31, Theorem 4.1], and [38, Theorem 4.2]: One tries to create two Markov chains, x_t and x'_t, having the same one-step transition probability distribution but driven independently until they are coupled on a small set with some fixed probability whenever they visit the small set. Here, x'_t is a stationary Markov chain. By the Coupling Inequality and the previous discussion with Nummelin's splitting technique we have $\|P^n(x, \cdot) - \pi(\cdot)\|_{TV} \leq P(x_n \neq x'_n)$, where $x'_n \sim \pi P^n = \pi$.

D. Drift Criteria for Positivity

We now consider specific formulations of the random-time drift criterion (1). Throughout the paper the sequence of stopping times $\{\mathcal{T}_i\}_{i \geq 0}$ is assumed to be non-decreasing, with $\mathcal{T}_0 = 0$.

Theorem II.4 is the general result of [11], providing a single criterion for positive Harris recurrence, as well as finite "moments" (the steady-state mean of the function f appearing in the drift condition (6)). The drift condition (6) is a refinement of (1).

Theorem II.4 [11]: Suppose that $\{x_t\}$ is a φ-irreducible and aperiodic Markov chain. Suppose moreover that there are functions $V : \mathbb{X} \to (0, \infty)$, $\delta : \mathbb{X} \to [1, \infty)$, $f : \mathbb{X} \to [1, \infty)$, a small set C on which V is bounded, and a constant $b \in \mathbb{R}$, such that

$$E\left[V\left(x_{\mathcal{T}_{i+1}}\right) \mid \mathcal{F}_{\mathcal{T}_i}\right] \leq V\left(x_{\mathcal{T}_i}\right) - \delta\left(x_{\mathcal{T}_i}\right) + b 1_C\left(x_{\mathcal{T}_i}\right)$$
$$E\left[\sum_{k=\mathcal{T}_i}^{\mathcal{T}_{i+1}-1} f(x_k) \middle| \mathcal{F}_{\mathcal{T}_i}\right] \leq \delta\left(x_{\mathcal{T}_i}\right). \qquad i \geq 0. \quad (6)$$

Then the following hold:

(i) $\{x_t\}$ is positive Harris recurrent, with unique invariant distribution π

(ii) $\pi(f) := \int f(x)\pi(dx) < \infty$.

(iii) For any function g that is bounded by f, in the sense that $\sup_x |g(x)|/f(x) < \infty$, we have convergence of moments in the mean, and the strong law of large numbers holds

$$\lim_{t\to\infty} E_x[g(x_t)] = \pi(g)$$

$$\lim_{N\to\infty} \frac{1}{N} \sum_{t=0}^{N-1} g(x_t) = \pi(g) \qquad a.s., \ x \in \mathbb{X}.$$

E. Rates of Convergence: Geometric Ergodicity

In this section, following [3] and [31], we review results stating that a strong type of ergodicity, geometric ergodicity, follows from a simple drift condition. An irreducible Markov chain is said to satisfy the *univariate drift condition* if there are constants $\lambda \in (0,1)$ and $b < \infty$, along with a function $V : \mathbb{X} \to [1,\infty)$, and a small set C such that

$$PV \leq \lambda V + b1_C. \tag{7}$$

Using the coupling inequality, Roberts and Rosenthal [31] prove that geometric ergodicity follows from the univariate drift condition. We also note that the univariate drift condition allows us to assume that V is bounded on C without any loss (see [31, Lemma 14]).

Theorem II.5 ([31, Theorem 9]): Suppose $\{x_t\}$ is an aperiodic, irreducible Markov chain with invariant distribution π. Suppose C is a $(1, \epsilon, \nu)$-small set and $V : \mathbb{X} \to [1,\infty)$ satisfies the univariate drift condition with constants $\lambda \in (0,1)$ and $b < \infty$. Then $\{x_t\}$ is geometrically ergodic.

That geometric ergodicity follows from the univariate drift condition with a small set C is proven by Roberts and Rosenthal by using the coupling inequality to bound the TV-norm, but an alternate proof is given by Meyn and Tweedie [3] resulting in the following theorem.

Theorem II.6 ([3, Theorem 15.0.1]): Suppose $\{x_t\}$ is an aperiodic and irreducible Markov chain. Then the following are equivalent:

(i) $E_x[\mathcal{T}_B] < \infty$ for all $x \in \mathbb{X}$, $B \in \mathcal{B}^+(\mathbb{X})$, the invariant distribution π of $\{x_t\}$ exists and there exists a petite set C, constants $\gamma < 1$, $M > 0$ such that for all $x \in C$

$$|P(x,C) - \pi(C)| < M\gamma^n.$$

(ii) For a petite set C and for some $\kappa > 1$

$$\sup_{x \in C} E_x[\kappa^{\mathcal{T}_C}] < \infty.$$

(iii) For a petite set C, constants $b > 0$ $\lambda \in (0,1)$, and a function $V : \mathbb{X} \to [1,\infty]$ (finite for some x) such that

$$PV \leq \lambda V + b1_C.$$

Any of the conditions imply that there exists $r > 1$, $R < \infty$ such that for any x

$$\sum_{n=0}^{\infty} r^n \|P^n(x,\cdot) - \pi(\cdot)\|_V \leq RV(x).$$

We note for future reference that if (iii) above holds, (ii) holds for for all $\kappa \in (1, \lambda^{-1})$.

F. Rates of Convergence: Subgeometric Ergodicity

Here, we review the class of subgeometric rate functions (see [38, Section 4], [6, Section 5], and [3], [4], [35], [39]). Let Λ_0 be the family of functions $r : \mathbb{Z}_+ \to [0,\infty)$ satisfying

$$r \text{ is non-decreasing,} \quad r(1) \geq 2$$

and

$$\frac{\log r(n)}{n} \downarrow 0 \quad \text{as } n \to \infty.$$

The second condition implies that for all $r \in \Lambda_0$

$$r(m+n) \leq r(m)r(n) \quad \text{for all } m, n \in \mathbb{Z}_+. \tag{8}$$

The class of subgeometric rate functions Λ defined in [35] is the class of sequences r for which there exists a function $r_0 \in \Lambda_0$ such that

$$0 < \liminf_{n\to\infty} \frac{r(n)}{r_0(n)} \leq \limsup_{n\to\infty} \frac{r(n)}{r_0(n)} < \infty.$$

The main theorems we use to construct conditions on subgeometric rates of convergence are due to Tuominen and Tweedie [35].

Theorem II.7 ([35], Theorem 2.1): Suppose $\{x_t\}$ is an irreducible and aperiodic Markov chain with state space $\mathbb{X}$ and transition probability P. Let $f : \mathbb{X} \to [1,\infty)$ and $r \in \Lambda$ be given. The following are equivalent:

(i) There exists a petite set $C \in \mathcal{B}(\mathbb{X})$ such that

$$\sup_{x \in C} E_x \left[\sum_{k=0}^{\mathcal{T}_C - 1} r(k)f(x_k) \right] < \infty.$$

(ii) There exist a sequence $\{V_n\}$ of functions $V_n : \mathbb{X} \to [0,\infty]$, a petite set $C \in \mathcal{B}(\mathbb{X})$, and $b > 0$ such that V_0 is bounded on C

$$V_0(x) = \infty \text{ implies } V_1(x) = \infty$$

and

$$PV_{n+1} \leq V_n - r(n)f + br(n)1_C, \quad n \in \mathbb{Z}_+. \tag{9}$$

(iii) There exists an (f,r)-regular set $A \in \mathcal{B}^+(\mathbb{X})$.

(iv) There exists a full absorbing set S which can be covered by a countable number of (f,r)-regular sets.

Theorem II.8 ([35, Theorem 4.1]): Suppose an aperiodic and irreducible Markov chain $\{x_t\}$ satisfies the equivalent conditions (i)–(iv) of Theorem II.7 with $f : \mathbb{X} \to [1,\infty)$ and $r \in \Lambda$. Then the Markov chain is (f,r)-ergodic, i.e.,

$$\lim_{n\to\infty} r(n) \|P^n(x,\cdot) - \pi\|_f = 0.$$

The proof of this result relies on a first-entrance last-exit decomposition [3] of the transition probabilities; see [3, Section 13.2.3].

The conditions of Theorem II.7 may be hard to check, especially (ii), comparing a sequence of Lyapunov functions $\{V_k\}$ at each time step. We briefly discuss the methods of

Douc *et al.* [39] (see also Hairer [38]) that extend the subgeometric ergodicity results and show how to construct subgeometric rates of ergodicity from a simpler drift condition. [39] assumes that there exists a function $V : \mathbb{X} \to [1, \infty]$, a concave monotone non-decreasing differentiable function $\phi : [1, \infty] \to (0, \infty]$, a set $C \in \mathcal{B}(\mathbb{X})$ and a constant $b \in \mathbb{R}$ such that

$$PV + \phi \circ V \leq V + b 1_C. \tag{10}$$

If an aperiodic and irreducible Markov chain $\{x_t\}$ satisfies the above with a petite set C, and if $V(x_0) < \infty$, then it can be shown that $\{x_t\}$ satisfies Theorem II.7(ii). Therefore $\{x_t\}$ has invariant distribution π and is $(\phi \circ V, 1)$-ergodic so that $\lim_{n \to \infty} \|P^n(x, \cdot) - \pi(\cdot)\|_{\phi \circ V} = 0$ for all x in the set $\{x : V(x) < \infty\}$ of π-measure 1. The results by Douc *et al.* build then on trading off $(\phi \circ V, 1)$ ergodicity for $(1, r_\phi)$-ergodicity for some rate function r_ϕ, by carefully constructing the function utilizing concavity; see [39, Propositions 2.1 and 2.5] and [38, Theorem 4.1(3)].

To achieve ergodicity with a nontrivial rate and norm one can invoke a result involving the class of *pairs of ultimately non-decreasing functions*, defined in [39]. The class $\mathcal{Y}$ of pairs of ultimately non-decreasing functions consists of pairs $\Psi_1, \Psi_2 : \mathbb{X} \to [1, \infty)$ such that $\Psi_1(x)\Psi_2(y) \leq x + y$ and $\Psi_i(x) \to \infty$ for one of $i = 1, 2$.

Proposition II.9: Suppose $\{x_t\}$ is an aperiodic and irreducible Markov chain that is both $(1, r)$-ergodic and $(f, 1)$-ergodic for some $r \in \Lambda$ and $f : \mathbb{X} \to [1, \infty)$. Suppose $\Psi_1, \Psi_2 : \mathbb{X} \to [1, \infty)$ are a pair of ultimately non-decreasing functions. Then $\{x_t\}$ is $(\Psi_1 \circ f, \Psi_2 \circ r)$-ergodic.

Therefore we can show that if $(\Psi_1, \Psi_2) \in \mathcal{Y}$ and a Markov chain satisfies the condition (10), then it is $(\Psi_1 \circ \phi \circ V, \Psi_2 \circ r_\phi)$-ergodic.

III. RATES OF CONVERGENCE UNDER RANDOM-TIME STATE-DEPENDENT DRIFT

The second condition of Theorem II.7 assumes that a deterministic sequence of functions $\{V_n\}$ exists and satisfies the drift condition (9).

We apply Theorem II.7 to the case where the Foster-Lyapunov drift condition holds not for every n but for a sequence of stopping times $\{\mathcal{T}_n\}$. Our goal is to reveal a relation between the stopping times $\{\mathcal{T}_n\}$ where a drift condition holds and the rate function r, so that we obtain (f, r)-ergodicity.

A. A General Result on Ergodicity

The following result builds on and generalizes [11, Theorem 2.1].

Theorem III.1: Let $\{x_t\}$ be an aperiodic and irreducible Markov chain with a small set C. Suppose there are functions $V : \mathbb{X} \to (0, \infty)$ with V bounded on C, $f : \mathbb{X} \to [1, \infty)$, $\delta : \mathbb{X} \to [1, \infty)$, a constant $b \in \mathbb{R}$, and $r \in \Lambda$ such that for a sequence of stopping times $\{\mathcal{T}_n\}$

$$E\left[V\left(x_{\mathcal{T}_{n+1}}\right) \mid x_{\mathcal{T}_n}\right] \leq V\left(x_{\mathcal{T}_n}\right) - \delta\left(x_{\mathcal{T}_n}\right) + b 1_C\left(x_{\mathcal{T}_n}\right)$$
$$E\left[\sum_{k=\mathcal{T}_n}^{\mathcal{T}_{n+1}-1} f(x_k)r(k) \,\middle|\, \mathcal{F}_{\mathcal{T}_n}\right] \leq \delta\left(x_{\mathcal{T}_n}\right). \tag{11}$$

Then $\{x_t\}$ satisfies Theorem II.7 and is (f, r)-ergodic.

Proof: The proof is similar to the proof of the Comparison Theorem of [3] as well as [11, Theorem 2.1(i)]. We may assume $r \in \Lambda_0$. We define sampled hitting times $\gamma_B = \min\{n > 0 : x_{\mathcal{T}_n} \in B\}$ for all $B \in \mathcal{B}^+(\mathbb{X})$ and $\gamma_B^N = \min(N, \gamma_B)$. Since $\{x_{\mathcal{T}_n}\}$ satisfies the drift condition, it follows that for $x \in C$:

$$E_x\left[\sum_{n=0}^{\gamma_C^N - 1} \delta(x_{\mathcal{T}_n})\right] \leq V(x) + b E_x\left[\sum_{n=0}^{\gamma_C^N - 1} 1_C\left(x_{\mathcal{T}_n}\right)\right] \leq V(x) + b$$

which is finite since V is bounded on C by assumption. An application of the monotone convergence theorem then gives

$$E_x\left[\sum_{n=0}^{\gamma_C - 1} \delta(x_{\mathcal{T}_n})\right] \leq V(x) + b E_x\left[\sum_{n=0}^{\gamma_C - 1} 1_C(x_{\mathcal{T}_n})\right] \leq V(x) + b$$

Since $\mathcal{T}_B \leq \mathcal{T}_{\gamma_B}$ for all $B \in \mathcal{B}^+(\mathbb{X})$ by definition, we have

$$E_x\left[\sum_{n=0}^{\mathcal{T}_C - 1} f(x_n)r(n)\right] \leq E_x\left[\sum_{n=0}^{\gamma_C - 1} \delta\left(x_{\mathcal{T}_n}\right)\right] \leq V(x) + b$$

so C is a petite set which satisfies

$$\sup_{x \in C} E_x\left[\sum_{n=0}^{\mathcal{T}_C - 1} r(n)f(x_n)\right] \leq \sup_{x \in C} V(x) + b < \infty.$$

This means that the Markov chain $\{x_n\}$ satisfies Theorem II.7(i) and is (f, r)-ergodic. ∎

B. On Petite Sets and Sampling

Unfortunately the techniques we reviewed earlier that rely on petite sets (specifically Theorem II.3) become unavailable in the random time drift setting as a petite set C for $\{x_n\}$ is not necessarily petite for $\{x_{\mathcal{T}_n}\}$. To be able to relax conditions on the behavior of V on C, we can place one of the following two conditions on the stopping times or require that V is bounded on C.

For an analogous application of Theorem II.3 in the random time setting we define sampled hitting times for any $B \in \mathcal{B}^+(\mathbb{X})$ as $\gamma_B = \min\{n > 0 : x_{\mathcal{T}_n} \in B\}$.

Lemma III.2: Suppose $\{x_t\}$ is an aperiodic and irreducible Markov chain. If there exists sequence of stopping times $\{\mathcal{T}_n\}$ independent of $\{x_t\}$, then any C that is small for $\{x_t\}$ is petite for $\{x_{\mathcal{T}_n}\}$.

Proof: Since C is petite, it is small by Theorem II.2 for some m. Let C be (m, δ, ν)-small for $\{x_t\}$

$$P^{\mathcal{T}_1}(x, \cdot) = \sum_{k=1}^{\infty} P(\mathcal{T}_1 = k) P^k(x, \cdot)$$

$$\geq \sum_{k=m}^{\infty} P(\mathcal{T}_1 = k) \int P^m(x, dy) P^{k-m}(y, \cdot)$$

$$\geq \sum_{k=m}^{\infty} P(\mathcal{T}_1 = k) \int 1_C(x) \delta \nu(dy) P^{k-m}(y, \cdot) \tag{12}$$

which is a well defined measure. Therefore defining $\kappa(\cdot) = \int \nu(dy) \sum_{k=m}^{\infty} P(\mathcal{T}_1 = k) P^{k-m}(y, \cdot)$, we have that C is $(1, \delta, \kappa)$-small for $\{x_{\mathcal{T}_n}\}$. ∎

The above allows us to uniformly bound $E_x[\sum_{n=0}^{\gamma_B - 1} 1_C(x_{\mathcal{T}_n})]$ when the stopping times are independent of the Markov chain, by an application of Theorem II.3 and (4).

The independence of stopping times $\{\mathcal{T}_n\}$ of $\{x_t\}$ is a restrictive condition that *event triggered systems* cannot satisfy since in such systems the stopping times depend explicitly on the state process hitting certain sets. One useful example where independence of stopping times can be enforced is given in [21] where a system controlled over an unreliable network is affected by variable transmission delays between the controller and the plant.

For the event-triggered case we will derive a useful result which will be used to show that in the drift equations of the form (11), the Lyapunov function V may not need to be assumed bounded on C.

The proof of the next result follows directly from the definition of $\mathcal{T}_n$.

Lemma III.3: Suppose $\{\mathcal{T}_n\}$ are the subsequent hitting times of a sequence of sets $\{E_n\}$ in $\mathcal{B}^+(\mathbb{X})$, so that $\mathcal{T}_{n+1} = \min\{t > \mathcal{T}_n : x_t \in E_{n+1}\}$. If $\bigcap_{n=0}^{\infty} E_n \in \mathcal{B}^+(\mathbb{X})$ then for any reachable $B \subset \bigcap_n E_n$, we have $\mathcal{T}_{\gamma_B} = \mathcal{T}_B$.

Assumption III.1: The stopping times are as in Lemma III.3 and $C \subset \bigcap_{n=0}^{\infty} E_n$.

Recall that by Theorem II.3 a petite set C is petite with a maximal irreducibility measure κ for a distribution a with finite mean, so for any $B \in \mathcal{B}^+(\mathbb{X})$ we have that $\sum_{n=0}^{\infty} a(n) P^n(x, B) \geq \kappa(B) 1_C(x)$. Assumption III.1 then implies that if any C is petite for $\{x_t\}$, then for some $\tilde{C} \subset C \subset \bigcap_{n=0}^{\infty} E_n$, we have that

$$
\begin{aligned}
E_x \left[\sum_{k=0}^{\gamma_{\tilde{C}} - 1} 1_C(x_{\mathcal{T}_k}) \right] &\leq E_x \left[\sum_{k=0}^{\mathcal{T}_{\tilde{C}} - 1} 1_C(x_k) \right] \\
&\leq \frac{1}{\kappa(\tilde{C})} E_x \left[\sum_{k=0}^{\mathcal{T}_{\tilde{C}} - 1} \sum_n 1_{\tilde{C}}(x_{k+n}) a(n) \right] \\
&= \frac{1}{\kappa(\tilde{C})} \sum_n a(n) E_x \left[\sum_{k=0}^{\mathcal{T}_{\tilde{C}} - 1} 1_{\tilde{C}}(x_{k+n}) \right] \\
&\leq \frac{1}{\kappa(\tilde{C})} \sum_n a(n) n = c(\tilde{C}) < \infty. \quad (13)
\end{aligned}
$$

Hence if the stopping times satisfy the conditions in Lemma III.2 or Lemma III.3, we can drop the condition that V is bounded on C, by applying [3, Chap. 11] and [3, Proposition 5.5.6] to $\{x_{\mathcal{T}_n}\}$ and noting that by (11), $\{x_t\}$ satisfies Theorem II.7(i). This follows since the drift condition implies for any $B \in \mathcal{B}^+(\mathbb{X})$, $E_x[\sum_{n=0}^{\gamma_B - 1} \delta(x_{\mathcal{T}_n})] \leq V(x) + b E_x[\sum_{n=0}^{\gamma_B - 1} 1_C(x_{\mathcal{T}_n})]$, where the last term is bounded if the conditions of either Lemma III.2 or Lemma III.3 and Assumption III.1 are satisfied.

It is interesting to note that the two extreme cases of the stopping times, either independent of or completely dependent on the Markov chain, both give similarly useful relaxations.

C. Subgeometric Ergodicity

The second inequality (11) may be hard to check as it does not provide means for checking the relation between the stopping times $\{\mathcal{T}_n\}$ and the rate function r since the function depends on k in a non-explicit fashion. In the following, the relationship of the criteria with the rate function r is relative to the stopping time.

We assume that $r \in \Lambda_0$ and thus r satisfies $r(m + n) \leq r(m) r(n)$.

Theorem III.4: Let $\{x_t\}$ be an aperiodic and irreducible Markov chain with a small set C. Suppose there exist $V : \mathbb{X} \to [1, \infty)$ which is bounded on C and for some $\epsilon > 0$, $\lambda \in (0, 1)$, $\lambda V(x) \leq V(x) - \epsilon$ for all $x \notin C$, and $b \in \mathbb{R}$ such that for an increasing sequence of stopping times $\{\mathcal{T}_n\}$

$$
E\left[V\left(x_{\mathcal{T}_{n+1}} \right) \mid \mathcal{F}_{\mathcal{T}_n} \right] \leq \lambda V\left(x_{\mathcal{T}_n} \right) + b 1_C\left(x_{\mathcal{T}_n} \right). \quad (14)
$$

If

$$
\sup_k E\left[\sum_{n=\mathcal{T}_k}^{\mathcal{T}_{k+1} - 1} r(n - \mathcal{T}_k) \mid \mathcal{F}_{\mathcal{T}_k} \right] =: M < \infty \quad (15)
$$

and

$$
\sup_k E\left[r(\mathcal{T}_{k+1} - \mathcal{T}_k) \mid \mathcal{F}_{\mathcal{T}_k} \right] \leq \lambda^{-1} \quad (16)
$$

then $\{x_t\}$ satisfies Theorem II.7 with $f = 1$ and is $(1, r)$-ergodic.

Proof: Suppose that instead of (14), we have that

$$
E\left[V(x_{n+1}) \mid \mathcal{F}_n \right] \leq \lambda V(x_n) + b 1_C(x_n). \quad (17)
$$

It follows then that the sequence $\{M_n\}$ defined by:

$$
M_n = \lambda^{-n} V(x_n) - \sum_{k=0}^{n-1} b 1_C(x_k) \lambda^{-(k+1)}
$$

with $M_0 = V(x_0)$, is a supermartingale. Then, with (14), defining $\gamma_B^N = \min\{N, \gamma_B\}$ for $B \in \mathcal{B}^+(\mathbb{X})$ gives, by Doob's optional sampling theorem

$$
E_x\left[\lambda^{-\gamma_B^N} V\left(x_{\mathcal{T}_{\gamma_B^N}} \right) \right] \leq V(x)
$$

$$
+ E_x\left[\sum_{n=0}^{\gamma_B^N - 1} b 1_C\left(x_{\mathcal{T}_n} \right) \lambda^{-(n+1)} \right] \quad (18)
$$

for any $B \in \mathcal{B}^+(\mathbb{X})$, and $N \in \mathbb{Z}_+$.

Since V is bounded above on C, we have that $C \subset \{V \leq L_1\}$ for some L_1 and thus

$$
\sup_{x \in C} E_x\left[\lambda^{-\gamma_C^N} V\left(x_{\mathcal{T}_{\gamma_C^N}} \right) \right] \leq L_1 + \lambda^{-1} b
$$

and by the monotone convergence theorem, and the fact that V is bounded from below by 1 everywhere and bounded from above on C

$$
\sup_{x \in C} E_x\left[\lambda^{-\gamma_C} V\left(x_{\mathcal{T}_{\gamma_C}} \right) \right] \leq L_1(L_1 + \lambda^{-1} b).
$$

Now, for any $r \in \Lambda$, we have

$$\sup_{x \in C} E_x \left[\sum_{n=0}^{\mathcal{T}_C - 1} r(n) \right] \leq \sup_{x \in C} E_x \left[\sum_{n=0}^{\mathcal{T}_{\gamma_C} - 1} r(n) \right]$$

and since $r(m + n) \leq r(m)r(n)$ by (8), we obtain through iterated expectations that

$$\sup_{x \in C} E_x \left[\sum_{n=0}^{\mathcal{T}_C - 1} r(n) \right]$$

$$\leq \sup_{x \in C} E_x \left[\sum_{k=0}^{\gamma_C - 1} E \left[E \left[\sum_{n=\mathcal{T}_k}^{\mathcal{T}_{k+1} - 1} r(n - \mathcal{T}_k) \mid \mathcal{F}_{\mathcal{T}_k} \right] \right. \right.$$

$$\left. \left. \times \prod_{m=1}^{k} r(\mathcal{T}_m - \mathcal{T}_{m-1}) \right] \right].$$

Now, with (15), (16), and by the fact that V is bounded from below by 1, it follows that:

$$\sup_{x \in C} E_x \left[\sum_{n=0}^{\mathcal{T}_{\gamma_C} - 1} r(n) \right]$$

$$\leq \sup_{x \in C} E_x \left[\sum_{n=1}^{\gamma_C - 1} E_x \left[M V(x_{\mathcal{T}_n}) \lambda^{-n} \right] \right]$$

$$\leq L_1 \sup_{x \in C} E_x \left[\sum_{n=0}^{\gamma_C - 1} M \left(V(x) + \lambda^{-1} b \right) \right] \qquad (19)$$

so that

$$\sup_{x \in C} E_x \left[\sum_{n=0}^{\mathcal{T}_{\gamma_C} - 1} r(n) \right] \leq M L_1 (L_1 + \lambda^{-1} b) \sup_{x \in C} E_x[\gamma_C].$$

From (14) and the condition $\lambda V(x) \leq V(x) - \epsilon$ for $x \notin C$, we get that for all $x \in C$

$$E_x \left[V \left(x_{\mathcal{T}_{\gamma_C}} \right) \right]$$

$$\leq \left(V(x) - E_x \left[\sum_{n=0}^{\gamma_C - 1} \epsilon \right] + E_x \left[\sum_{n=0}^{\gamma_C - 1} b 1_C (x_{\mathcal{T}_n}) \right] \right)$$

and thus

$$\sup_{x \in C} E_x[\gamma_C] \leq \frac{L_1 + b}{\epsilon}.$$

Therefore $C \in \mathcal{B}^+(\mathbb{X})$ is a petite set such that $\sup_{x \in C} E_x[\sum_{n=0}^{\mathcal{T}_C - 1} r(n)]$ is finite and so $\{x_t\}$ satisfies Theorem II.7(i) with $f = 1$ and is $(1, r)$-ergodic. ∎

Remark III.1: We note that just as in the previous theorem, if the stopping times satisfy Lemma III.2, we can focus on return times for a petite set $A \subseteq \{V \leq L\}$ with $(V(x)/W(x)) \leq V(x) - \epsilon$ for all $x \notin A$ instead of C. Similarly, if the stopping times satisfy Lemma III.3 and Assumption III.1 we

can focus on return times for a petite set A in $\bigcap_n E_n$ with $(V(x)/W(x)) \leq V(x) - \epsilon$ for all $x \notin A \cap (\bigcap_n E_n)$ instead of C. This allows us to relax the conditions for V.

As an example, with $r(n) = 2n^\alpha$, let for all k, $E[\sum_{m=\mathcal{T}_k}^{\mathcal{T}_{k+1} - 1} (m - \mathcal{T}_k)^\alpha | \mathcal{F}_{\mathcal{T}_k}] < \infty$ and $E[(\mathcal{T}_{k+1} - \mathcal{T}_k)^\alpha | \mathcal{F}_{\mathcal{T}_k}] \leq \lambda^{-1}$. Then, the chain is *polynomially ergodic*. Note that one can obtain explicit expressions for a large class of sums of powers of the form $\sum_{k=0}^{\mathcal{T}_1} k^\alpha$ with $\alpha \in \mathbb{Z}_+$.

We also note that if r satisfies $\sup_k E_x[r(\mathcal{T}_{k+1} - \mathcal{T}_k)|x_{\mathcal{T}_k}] \leq M$ for some finite M, then by Jensen's inequality $r^{1/s}$ satisfies the bound $\sup_k E[r^{1/s}(\mathcal{T}_{k+1} - \mathcal{T}_k)|\mathcal{F}_{\mathcal{T}_k}] \leq \lambda^{-1}$ if $s > 1$ is large enough so that $M^{1/s} \leq \lambda^{-1}$.

Suppose now that the sequence of stopping times are state-dependent but deterministic, that is $\mathcal{T}_{k+1} = \mathcal{T}_k + n(x_{\mathcal{T}_k})$, $\mathcal{T}_0 = 0$.

Corollary III.5: Let $\{x_t\}$ be an aperiodic and irreducible Markov chain with a small set C. Suppose there exist $V : \mathbb{X} \to [1, \infty)$ which is bounded on C and with for some $\epsilon > 0$, $\lambda \in (0, 1)$, $\lambda V(x) \leq V(x) - \epsilon$ for all $x \notin C$, and $b \in \mathbb{R}$ such that for an increasing sequence of stopping times $\{\mathcal{T}_n\}$

$$E \left[V \left(x_{\mathcal{T}_{n+1}} \right) \mid \mathcal{F}_{\mathcal{T}_n} \right] \leq \lambda V \left(x_{\mathcal{T}_n} \right) + b 1_C \left(x_{\mathcal{T}_n} \right). \qquad (20)$$

Then for any $r \in \Lambda$ and $M > 0$ that satisfy

$$\sum_{k=0}^{n(x_0)} r(k) \leq M, \quad r(n(x_0)) \leq \frac{1}{\lambda}, \quad x_0 \in \mathbb{X}$$

then $\{x_t\}$ satisfies Theorem II.7 with $f = 1$ and it is $(1, r)$-ergodic.

We note that Theorem III.4 above is useful for proving $(1, r)$-ergodicity and Theorem III.1 is really only useful for proving $(f, 1)$-ergodicity, where r, f satisfy the respective hypotheses. In order to be able to prove more rate results, we may use results by Douc *et al.* [39] on the class $\mathcal{Y}$ of pairs of ultimately non-decreasing functions defined in Section II-F. If a Markov chain $\{x_t\}$ satisfies Theorem III.4 with $(1, r)$ and Theorem III.1 with $(f, 1)$, then $\{x_t\}$ is $(\Psi_1 \circ f, \Psi_2 \circ r)$-ergodic for $(\Psi_1, \Psi_2) \in \mathcal{Y}$ by Proposition II.9.

Before ending this section, we revisit a criterion by Connor and Fort [6] who studied rates of convergence under drift criteria which are based on state-dependent but deterministic sampling times so that

$$P^{n(x)} V(x) \leq \lambda V(x) + b 1_C(x)$$

where $n : \mathbb{X} \to \mathbb{Z}_+$ is the state dependent time where the drift condition is enforced. Now, consider the case where n is random and we have a sequence of stopping times $\{\mathcal{T}_k\}$ defined as $\mathcal{T}_{k+1} = \mathcal{T}_k + n(x_{\mathcal{T}_k})$ with $\mathcal{T}_0 = 0$. Theorem 3.2(i) of [6] can be partly generalized to the random-time case as follows.

Theorem III.6: Let $\{x_t\}$ be an aperiodic and irreducible Markov chain with a small set C. Suppose that the stopping times $\{\mathcal{T}_n\}$ satisfy the conditions of Lemma III.2 and that there exist a function $V : \mathbb{X} \to [1, \infty)$, V bounded on C, and constants $b \in \mathbb{R}$ and $\lambda \in (0, 1)$ such that for an increasing sequence of stopping times $\{\mathcal{T}_n\}$ with $\mathcal{T}_0 = 0$

$$E_x \left[V \left(x_{\mathcal{T}_{k+1}} \right) \mid \mathcal{F}_{\mathcal{T}_k} \right] \leq \lambda V \left(x_{\mathcal{T}_k} \right) + b 1_C \left(x_{\mathcal{T}_k} \right).$$

If there exists a strictly increasing function $R : (0, \infty) \to (0, \infty)$ such that $R(t)/t$ is non-increasing and $E[R(\mathcal{T}_{k+1} - \mathcal{T}_k)|x_{\mathcal{T}_k}] \leq V(x_{\mathcal{T}_k})$, then there exists a constant D such that $E_x[R(\mathcal{T}_C)] \leq DV(x)$. If in addition the invariant distribution π of $\{x_t\}$ exists, $\pi(V) < \infty$, then the Markov chain is $(1, R)$-ergodic.

Proof: Since $R(t)/t$ is non increasing, it follows that $\log R(t)/t \to 0$ and $R \in \Lambda$. It also follows that $R(a + b) \leq R(a) + R(b)$ for any $a, b > 0$. With R also increasing we have that

$$E_x\left[R(\mathcal{T}_C)\right] \leq E_x\left[R\left(\mathcal{T}_{\gamma_C}\right)\right] = E_x\left[R\left(\sum_{k=0}^{\gamma_C - 1} \mathcal{T}_{k+1} - \mathcal{T}_k\right)\right]$$

$$\leq E_x\left[\sum_{k=0}^{\gamma_C - 1} R(\mathcal{T}_{k+1} - \mathcal{T}_k)\right]$$

$$= E_x\left[\sum_{k=0}^{\gamma_C - 1} E_x\left[R(\mathcal{T}_{k+1} - \mathcal{T}_k) \mid \mathcal{F}_{\mathcal{T}_k}\right]\right]$$

$$\leq E_x\left[\sum_{k=0}^{\gamma_C - 1} V(x_{\mathcal{T}_k})\right].$$

With the drift condition $P^{\mathcal{T}_{k+1} - \mathcal{T}_k} V \leq V - (1 - \lambda)V + b1_C$ and V bounded on C, we have

$$(1 - \lambda)E_x\left[\sum_{k=0}^{\gamma_C - 1} V(x_{\mathcal{T}_k})\right] \leq V(x) + b$$

where $\mathcal{T}_0 = 0$ and so we obtain $E_x[R(\mathcal{T}_C)] < DV(x)$ for some $D > 0$.

If the invariant distribution π of $\{x_t\}$ exists and $\pi(V) < \infty$, then by [3, Theorem 14.2.11] there exists a small set A and $M \in \mathbb{R}$ such that $\sup_{x \in A} E_x[\sum_{k=0}^{\mathcal{T}_A - 1} V(x_k)] < M$. Defining the hitting time $\sigma_A = \min\{t \geq 0 : x_t \in A\}$, the function $W(x) = E_x[\sum_{n=0}^{\sigma_A} V(x_n)]$ satisfies the drift condition $PW \leq W - V + M1_A$ with A petite, and by Theorem II.3

$$E_x\left[\sum_{k=0}^{\mathcal{T}_B - 1} V(x_k)\right] \leq W(x) + ME_x\left[\sum_{k=0}^{\mathcal{T}_B - 1} 1_A(x_k)\right]$$

$$\leq W(x) + Mc(B)$$

for any $B \in \mathcal{B}^+(\mathbb{X})$. Therefore, since R is increasing

$$E_x\left[\sum_{k=0}^{\mathcal{T}_C - 1} R(k)\right] \leq E_x\left[\sum_{k=0}^{\mathcal{T}_C - 1} E[R(\mathcal{T}_C) \mid \mathcal{F}_k]\right]$$

$$\leq E_x\left[\sum_{k=0}^{\mathcal{T}_C - 1} DV(x_k)\right] \leq D(W(x) + Mc(C)). \quad (21)$$

To complete the proof, we show that W is bounded on C. If the stopping times are independent and thus satisfy the conditions of Lemma III.2, then C is petite for the randomly sampled chain $\{x_{\mathcal{T}_n}\}$ and the drift condition in the hypothesis gives $E_x[R(\mathcal{T}_B)] \leq (c(B) + 1)V(x)$ for any $B \in \mathcal{B}^+(\mathbb{X})$. Since W

satisfies a drift condition, $\{W < \infty\}$ is full and absorbing and we can find a petite set in $\{W < \infty\}$.

Combining the above with (21) gives

$$\sup_{x \in B} E_x\left[\sum_{k=0}^{\mathcal{T}_B - 1} R(k)\right] \leq \sup_{x \in B} (c(B) + 1)(W(x) + b) < \infty$$

for an appropriate petite set B when the conditions of Lemma III.2 are satisfied. Thus $\{x_t\}$ satisfies Theorem II.7(i) with $(f, r) = (1, R)$ and it is $(1, R)$ ergodic. ∎

D. Geometric Ergodicity

We use the same reasoning as before to obtain geometric ergodicity from a random time univariate drift condition.

Theorem III.7: Let $\{x_t\}$ be an aperiodic and irreducible Markov chain with a small set C. If there exists a function $V : \mathbb{X} \to [1, \infty)$, V bounded on C, constants $b \in \mathbb{R}$, $B > 0$, and $\lambda, \beta \in (0, 1)$ such that for a sequence of stopping times $\{\mathcal{T}_n\}$

$$E\left[V(x_{\mathcal{T}_{n+1}}) \mid \mathcal{F}_{\mathcal{T}_n}\right] \leq \lambda V(x_{\mathcal{T}_n}) + b1_C(x_{\mathcal{T}_n})$$

and

$$P(\mathcal{T}_{n+1} - \mathcal{T}_n = k \mid x_{\mathcal{T}_n}) \leq B\beta^k, \text{ for all } n, k, \text{ and } x_{\mathcal{T}_n} \notin C$$

with

$$\frac{1 - B\lambda}{\beta} > 1$$

and

$$\sup_{x \in C} E_x[a^{\mathcal{T}_1}] < \infty \quad (22)$$

for some $a > 1$, then x_t is geometrically ergodic.

Proof: By Theorem II.6 for $r \in (1, \lambda^{-1})$

$$\sup_{x \in C} E_x[r^{\gamma_C}] < \infty.$$

Let $\rho \in (1, (1 - B\lambda)/\beta)$. Then

$$E_x\left[\rho^{\mathcal{T}_{n+1} - \mathcal{T}_n} \mid \mathcal{F}_{\mathcal{T}_n}\right] \leq \frac{B}{1 - \rho\beta} < \lambda^{-1} \quad (23)$$

for $x \notin C$. By a use of iterated expectations

$$E_x\left[\rho^{\mathcal{T}_{\gamma_C}}\right] = E_x\left[\prod_{n=0}^{\gamma_C - 1} \rho^{\mathcal{T}_{n+1} - \mathcal{T}_n}\right] < E_x\left[\lambda^{-(\gamma_C - 1)}\rho^{\mathcal{T}_1}\right]. \quad (24)$$

By letting $1 < \rho < \min(a, (1 - B\lambda)/\beta)$, we obtain that $C \in \mathcal{B}^+(\mathbb{X})$ is a small set with a uniformly bounded $E_x[\rho^{\mathcal{T}_{\gamma_C}}]$ for $x \in C$. Therefore by Theorem II.6 the chain $\{x_t\}$ is geometrically ergodic. ∎

We also note that the rate of ergodicity relies on the constants m and δ for some (m, δ, ν)-small set C, so the ergodicity rate cannot be made explicit using only the information in the drift condition.
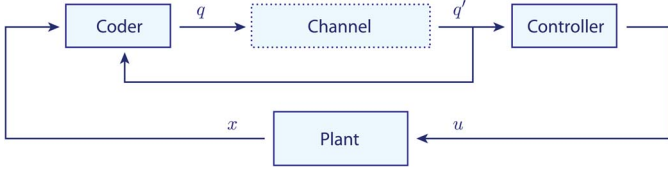
Fig. 1. Control of a system over a noisy channel.

IV. AN EXAMPLE IN NETWORKED CONTROL

We revisit the motivating example in [11], concerning the stabilization problem over erasure channels. In particular, we apply the results of the previous section to establish a rate of convergence to equilibrium provided that the information transmission rate satisfies a certain inequality. We consider a scalar LTI discrete-time system described by

$$x_{t+1} = ax_t + bu_t + w_t, \qquad t \geq 0 \qquad (25)$$

where x_t is the state at time t, u_t is the control input, the initial state x_0 is a random variable with a finite second moment, and $\{w_t\}$ is a sequence of zero-mean i.i.d. Gaussian random variables, also independent of x_0. We assume that the system is open-loop unstable and controllable, that is, $|a| \geq 1$ and $b \neq 0$. This system is connected over a noisy channel to a controller, as shown in Fig. 1. The channel is assumed to have finite input alphabet $\mathcal{M}$ and finite output alphabet $\mathcal{M}'$. A source coder maps the source symbols (state values) to corresponding channel inputs. The quantizer outputs are transmitted through the channel, after passing through a channel encoder. The receiver has access to noisy versions of the quantizer/coder outputs for each time instant t, which we denote by $q'_t \in \mathcal{M}'$.

The problem is to identify conditions on the channel so that there exist coding and control schemes leading to the stochastic stability of the controlled process. For a thorough review of such problems with necessity and sufficiency conditions, see [19].

The source output is quantized as follows:

$$Q_K^\Delta(x) = \begin{cases} \left(k - \frac{1}{2}(K+1)\right)\Delta, & \text{if } x \in \left[\left(k-1-\frac{1}{2}K\right)\Delta, \\ & \qquad \left(k-\frac{1}{2}K\right)\Delta\right) \\ \left(\frac{1}{2}(K-1)\right)\Delta, & \text{if } x = \frac{1}{2}K\Delta \\ 0, & \text{if } x \notin \left[-\frac{1}{2}K\Delta, \frac{1}{2}K\Delta\right] \end{cases}$$

where K is a positive integer. The quantizer outputs are transmitted through a memoryless erasure channel, after being subjected to a bijective mapping, which is performed by the channel encoder. At time t, the channel encoder $\mathcal{E}_t$ maps the quantizer output symbols to corresponding channel inputs $q_t \in \mathcal{M} := \{1, 2 \ldots, K+1\}$ so that $\mathcal{E}_t(Q_t(x_t)) = q_t$. The controller/decoder has access to noisy versions of the encoder outputs $q'_t \in \mathcal{M}' := \mathcal{M} \cup \{e\}$, with e denoting the erasure symbol, generated according to a probability distribution for every fixed $q \in \mathcal{M}$. The channel transition probabilities are given by

$$P(q' = i | q = i) = p, \qquad P(q' = e | q = i) = 1 - p, \quad i \in \mathcal{M}.$$

At each time t, the controller/decoder applies a mapping $\mathcal{D}_t : \mathcal{M} \cup \{e\} \to \mathbb{R}$, given by

$$\mathcal{D}_t(q'_t) = \mathcal{E}_t^{-1}(q'_t) \times 1_{\{q'_t \neq e\}} + 0 \times 1_{\{q'_t = e\}}.$$

Let $\{\Upsilon_t\}$ denote the sequence of i.i.d. binary random variables, representing the erasure process in the channel, where the event $\Upsilon_t = 1$ indicates that the signal is transmitted with no error through the channel at time t. Let $p = E[\Upsilon_t]$ denote the probability of success in transmission. The following key assumptions are imposed: Given $K \geq 2$ introduced in the definition of the quantizer, define the *rate variables*

$$R = \log_2(K+1) \qquad R' = \log_2(K). \qquad (26)$$

We fix positive scalars δ and α satisfying $|a|2^{-R'} < \alpha < 1$ and $\alpha(|a| + \delta)^{p^{-1}-1} < 1$. With $L > 0$ a constant, let $\bar{Q} : \mathbb{R} \times \mathbb{R} \times \{0, 1\} \to \mathbb{R}$ be defined as

$$\bar{Q}(\Delta, h, p) = \begin{cases} |a| + \delta, & \text{if } |h| > 1, \text{ or } p = 0 \\ \alpha, & \text{if } 0 \leq |h| \leq 1, p = 1, \Delta > L \\ 1, & \text{if } 0 \leq |h| \leq 1, p = 1, \Delta \leq L. \end{cases}$$

For each $t \geq 0$ and with $\Delta_0 \in \mathbb{R}$ selected arbitrarily, let

$$u_t = -\frac{a}{b}\hat{x}_t$$

$$\hat{x}_t = \mathcal{D}_t(q'_t) = \Upsilon_t Q_K^{\Delta_t}(x_t)$$

$$\Delta_{t+1} = \Delta_t \bar{Q}\left(\Delta_t, \left|\frac{x_t}{\Delta_t 2^{R'-1}}\right|, \Upsilon_t\right). \qquad (27)$$

Given the channel output $q'_t \neq e$, the controller can simultaneously deduce the realization of Υ_t and the event $\{|h_t| > 1\}$, where $h_t = x_t/(\Delta_t 2^{R'-1})$. This is due to the fact that if the channel output is not the erasure symbol, the controller knows that the signal is received with no error. If $q'_t = e$, however, then the controller applies 0 as its control input and enlarges the bin size of the quantizer.

By [11, Lemma 3.1], (x_t, Δ_t) is a Markov chain.

Consider now a sequence of stopping times which denote the times when there is a successful transmission of a source symbol in the *granular region* of the quantizer

$$\mathcal{T}_0 = 0, \quad \mathcal{T}_{z+1} = \inf\{k > \mathcal{T}_z : |h_k| \leq 1, p_k = 1\}, \quad z \in \mathbb{Z}_+.$$

By [11, Proposition 3.1], the stopping time distribution is bounded uniformly by a geometric measure.

Lemma IV.1 ([11, Proposition 3.1]): The discrete probability measure $P(\mathcal{T}_{i+1} - \mathcal{T}_i = k | x_{\mathcal{T}_i}, \Delta_{\mathcal{T}_i})$ satisfies

$$(1-p)^{k-1} \leq P(\mathcal{T}_{i+1} - \mathcal{T}_i \geq k \mid x_{\mathcal{T}_i}, \Delta_{\mathcal{T}_i}) \leq (1-p)^{k-1} + o(1)$$

where $o(1) \to 0$ as $\Delta_{\mathcal{T}_i} \to \infty$ uniformly in $x_{\mathcal{T}_i}$.

As a consequence, the probability $P(\mathcal{T}_{i+1} - \mathcal{T}_i \geq k | x_{\mathcal{T}_i}, \Delta_{\mathcal{T}_i})$ tends to $(1-p)^{k-1}p$ as $\Delta_{\mathcal{T}_i} \to \infty$.

Theorem IV.2: Suppose that

$$a^2 \left(1 - p + \frac{p}{(2^R - 1)^2}\right) < 1. \qquad (28)$$

Then, under the coding and control policy considered, the chain (x_t, Δ_t) is geometrically ergodic.

Proof: By the proof of [11, Theorem 3.2], with $V(x, \Delta) = \Delta^2$, and with

$$0 < \epsilon < 1 - \frac{p\alpha^2}{1 - (1-p)(|a|+\delta)^2}$$

$b < \infty$ and C a small set, it follows that:

$$E\left[V\left(x_{\mathcal{T}_{z+1}}, \Delta_{\mathcal{T}_{z+1}}\right) \mid x_{\mathcal{T}_z}\right]$$
$$\leq (1-\epsilon)V\left(x_{\mathcal{T}_z}, \Delta_{\mathcal{T}_z}\right) + b1_{\{(x_{\mathcal{T}_z}, \Delta_{\mathcal{T}_z}) \in C\}}. \quad (29)$$

Together with Lemma IV.1, and (34) in [11] leading to (22), these imply that for some $\lambda = 1 - \epsilon \in (0, 1)$, $B \in (p, p/\lambda)$ for some sufficiently large C, and $\beta = 1 - p$, Theorem III.7 holds. ∎

Remark IV.1: We recall from [11] that under (28), the system is quadratically stable in the sense that for each initial condition (x_0, Δ_0), $\lim_{t \to \infty} E[x_t^2] = E_\pi[x_0^2] < \infty$. We also note that by [11] if the goal is to only have the existence of an invariant probability measure, the requirements on the channel reduce to the conditions that $|a|2^{-R'} < \alpha < 1$ and $\alpha(|a|+\delta)^{p^{-1}-1} < 1$.

V. CONCLUSION

In this paper, we established random-time state-dependent drift criteria for Markov chains using Lyapunov-theoretic methods. We established drift criteria both for sub-geometric and geometric rates of convergence, where the conditions revealed the relationship between the distributions of the stopping times, the drift of the Lyapunov functions at random times, and the ergodicity rates. Future work includes the application of these results in event triggered control systems, as well as information theory problems for variable-length decoding.

REFERENCES

[1] V. S. Borkar, "Convex analytic methods in Markov decision processes," in *Handbook of Markov Decision Processes*, E. A. Feinberg and A. Shwartz, Eds. Boston, MA, USA: Kluwer, 2001, pp. 347–375.

[2] S. P. Meyn, *Control Techniques for Complex Networks*. Cambridge, U.K.: Cambridge Univ. Press, 2007.

[3] S. P. Meyn and R. Tweedie, *Markov Chains and Stochastic Stability*. London, U.K.: Springer-Verlag, 1993.

[4] S. P. Meyn and R. Tweedie, "State-dependent criteria for convergence of Markov chains," *Ann. Appl. Prob*, vol. 4, pp. 149–168, 1994.

[5] V. A. Malyšev and M. V. Menšikov, "Ergodicity, continuity and analyticity of countable Markov chains," *Trans. Moscow Math. Soc*, vol. 39, pp. 1–48, 1981.

[6] S. B. Connor and G. Fort, "State-dependent Foster-Lyapunov criteria for subgeometric convergence of Markov chains," *Stoch. Process Appl.*, vol. 119, pp. 176–4193, 2009.

[7] V. S. Borkar and S. P. Meyn, "The ODE method for convergence of stochastic approximation and reinforcement learning," *SIAM J. Control Optim.*, pp. 447–469, Dec. 2000.

[8] G. Fort, S. P. Meyn, E. Moulines, and P. Priouret, "ODE methods for skip-free Markov chain stability with applications to MCMC," *Annals Appl. Prob.*, pp. 664–707, 2008.

[9] J. G. Dai, "On positive harris recurrence of multiclass queueing networks: A unified approach via fluid limit models," *Annals Appl. Prob.*, vol. 5, pp. 49–77, 1995.

[10] J. G. Dai and S. P. Meyn, "Stability and convergence of moments for multiclass queueing networks via fluid limit models," *IEEE Trans. Autom. Control*, vol. 40, pp. 1889–1904, Nov. 1995.

[11] S. Yüksel and S. P. Meyn, "Random-time, state-dependent stochastic drift for Markov chains and application to stochastic stabilization over erasure channels," *IEEE Trans. Autom. Control*, vol. 58, pp. 47–59, Jan. 2013.

[12] B. H. Fralix, "Foster-type criteria for Markov chains on general spaces," *J. Appl. Prob.*, vol. 43, pp. 1194–1200, Dec. 2006.

[13] J. C. Kieffer, "Stochastic stability for feedback quantization schemes," *IEEE Trans. Inform. Theory*, vol. 28, pp. 248–254, Mar. 1982.

[14] J. C. Kieffer and J. G. Dunham, "On a type of stochastic stability for a class of encoding schemes," *IEEE Trans. Inform. Theory*, vol. 29, pp. 793–797, Nov. 1983.

[15] D. Liberzon and D. Nešič, "Input-to-state stabilization of linear systems with quantized state measurements," *IEEE Trans. Autom. Control*, vol. 52, pp. 767–781, May 2007.

[16] R. Brockett and D. Liberzon, "Quantized feedback stabilization of linear systems," *IEEE Trans. Autom. Control*, vol. 45, pp. 1279–1289, Jul. 2000.

[17] G. N. Nair and R. J. Evans, "Stabilizability of stochastic linear systems with finite feedback data rates," *SIAM J. Control Optim.*, vol. 43, pp. 413–436, Jul. 2004.

[18] S. Yüksel, "Stochastic stabilization of noisy linear systems with fixed-rate limited feedback," *IEEE Trans. Autom. Control*, vol. 55, pp. 1014–1029, Dec. 2010.

[19] S. Yüksel and T. Başar, *Stochastic Networked Control Systems: Stabilization and Optimization Under Information Constraints*. New York, NY, USA: Springer-Birkhäuser, 2013.

[20] M. Cloosterman, N. van de Wouw, W. Heemels, and H. Nijmeijer, "Stability of networked control systems with uncertain time-varying delays," *IEEE Trans. Autom. Control*, pp. 1575–1580, Jul. 2009.

[21] D. Quevedo, E. I. Silva, and G. C. Goodwin, "Control over unreliable networks affected by packet erasures and variable transmission delays," *IEEE J. Selected Areas Commun.*, vol. 26, pp. 672–685, 2008.

[22] G. M. Lipsa and N. C. Martins, "Remote state estimation with communication costs for first-order LTI systems," *IEEE Trans. Autom. Control*, vol. 56, pp. 2013–2025, Sep. 2011.

[23] W. Wu and A. Arapostathis, "Optimal sensor querying: General Markovian and LQG models with controlled observationss," *IEEE Trans. Autom. Control*, vol. 53, pp. 1392–1405, Jul. 2008.

[24] K. J. Aström and B. Bernhardsson, "Comparison of Riemann and Lebesgue sampling for first order stochastic systems," in *Proc. IEEE Conf. Decision Control*, Dec. 2002, pp. 2011–2016.

[25] D. Quevedo, V. Gupta, J. Ma, and S. Yüksel, "Stochastic stability of event-triggered anytime control," *IEEE Trans. Autom. Control*, vol. 59, pp. 3373–3379, Dec. 2014.

[26] K. J. M. Rabi and M. Johansson, "Optimal stopping for event-triggered sensing and actuation," in *Proc. IEEE Conf. Decision Control*, Dec. 2008, pp. 3607–3612.

[27] M. Lemmon, "Event-triggered feedback in control, estimation, and optimization," in *Networked Control Systems, ser. Lecture Notes in Control and Information Sciences*, A. Bemporad, M. Heemels, and M. Johansson, Eds. New York, NY, USA: Springer-Verlag, 2010, pp. 293–358.

[28] P. Tabuada, "Event-triggered real-time scheduling of stabilizing control tasks," *IEEE Trans. Autom. Control*, vol. 52, pp. 1680–1685, Sep. 2007.

[29] O. Hernandez-Lerma and J. Lasserre, *Discrete-Time Markov Control Processes*. New York, NY, USA: Springer-Verlag, 1996.

[30] A. Arapostathis, V. S. Borkar, E. Fernandez-Gaucherand, M. K. Ghosh, and S. I. Marcus, "Discrete-time controlled Markov processes with average cost criterion: A survey," *SIAM J. Control Optim.*, vol. 31, pp. 282–344, 1993.

[31] G. Roberts and J. Rosenthal, "General state space Markov chains and MCMC algorithms," *Probability Survery*, vol. 1, pp. 20–71, 2004.

[32] C. S. A. Antos and R. Munos, "Learning near-optimal policies with Bellman-residual minimization based fitted policy iteration and a single sample path," *Machine Learning*, vol. 71, no. 1, pp. 89–129, 2008.

[33] N. Saldi, T. Linder, and S. Yüksel, "Asymptotic optimality and rates of convergence of quantized stationary policies in stochastic control," *IEEE Trans. Autom. Control*, vol. 60, pp. 553–558, 2015.

[34] R. A. Zurkowski, "Lyapunov Analysis for Rates of Convergence in Markov Chains and Random-Time State-Dependent Drift," M.Sc. thesis, Queen's Univ., Kingston, ON, Canada, 2013.

[35] P. Tuominen and R. Tweedie, "Subgeometric rates of convergence of f-ergodic Markov chains," *Adv. Annals Appl. Prob.*, vol. 26, pp. 775–798, September 1994.

[36] E. Nummelin, "A splitting technique for Harris recurrent Markov chains," *Z. Wahrscheinlichkeitstheoric verw. Gebiete*, vol. 43, pp. 309–318, 1978.

[37] K. B. Athreya and P. Ney, "A new approach to the limit theory of recurrent Markov chains," *Trans. Amer. Math. Soc.*, vol. 245, pp. 493–501, 1978.

[38] University of Warwick, lecture notes M. Hairer, Convergence of Markov Processes, Aug. 2010. [Online]. Available: http://www.hairer.org/notes/Convergence.pdf lecture notes

[39] R. Douc, G. Fort, E. Moulines, and P. Soulier, "Practical drift conditions for subgeometric rates of convergence," *Annu. Appl. Probab*, vol. 14, pp. 1353–1377, 2004.

Ramiro Zurkowski received the B.Sc. degree in mathematics and engineering and the M.Sc. degree in mathematics from Queen's University, Kingston, ON, Canada, in 2012 and 2014, respectively.

Serdar Yüksel (S'02–M'11) received the B.Sc. degree in electrical and electronics engineering from Bilkent University, Ankara, Turkey, in 2001 and the M.S. and Ph.D. degrees in electrical and computer engineering from the University of Illinois at Urbana- Champaign, in 2003 and 2006, respectively.

He was a Post-Doctoral Researcher at Yale University, New Haven, CT, for a year before joining Queen's University, Kingston, ON, Canada, as an Assistant Professor of Mathematics and Engineering in the Department of Mathematics and Statistics, where he is now an Associate Professor. His research interests are on stochastic and decentralized control, information theory and applied probability.

Dr. Yüksel received the CAIMS/PIMS Early Career Award in Applied Mathematics in 2013. He is an Associate Editor for the IEEE TRANSACTIONS ON AUTOMATIC CONTROL.

Tamás Linder (S'92–M'93–SM'00–F'13) received the M.S. degree in electrical engineering from the Technical University of Budapest, Hungary, in 1988, and the Ph.D degree in electrical engineering from the Hungarian Academy of Sciences, Budapest, Hungary, in 1992.

He was a Post-Doctoral Researcher at the University of Hawaii, Manoa, in 1992 and a Visiting Fulbright Scholar at the Coordinated Science Laboratory, University of Illinois at Urbana-Champaign from 1993 to 1994. From 1994 to 1998, he was a faculty member in the Department of Computer Science and Information Theory, Technical University of Budapest. From 1996 to 1998 he was also a Visiting Research Scholar in the Department of Electrical and Computer Engineering, University of California, San Diego. In 1998, he joined Queen's University, Kingston, ON, Canada, where he is now a Professor of Mathematics and Engineering in the Department of Mathematics and Statistics. His research interests include communications and information theory, source coding and vector quantization, machine learning, and statistical pattern recognition.

Dr. Linder received the Premier's Research Excellence Award of the Province of Ontario in 2002 and the Chancellors Research Award of Queen's University in 2003. He was an Associate Editor for Source Coding of the IEEE TRANSACTIONS ON INFORMATION THEORY from 2003 to 2004.