Characterization of Information Channels for Asymptotic Mean Stationarity and Stochastic Stability of Nonstationary/Unstable Linear Systems

Serdar Yüksel

***Abstract*—Stabilization of nonstationary linear systems over noisy communication channels is considered. Stochastically stable sources, and unstable but noise-free or bounded-noise systems have been extensively studied in the information theory and control theory literature since the 1970s, with a renewed interest in the past decade. There have also been studies on noncausal and causal coding of unstable/nonstationary linear Gaussian sources. In this paper, tight necessary and sufficient conditions for stochastic stabilizability of unstable (nonstationary) possibly multidimensional linear systems driven by Gaussian noise over discrete channels (possibly with memory and feedback) are presented. Stochastic stability notions include recurrence, asymptotic mean stationarity and sample path ergodicity, and the existence of finite second moments. Our constructive proof uses random-time state-dependent stochastic drift criteria for stabilization of Markov chains. For asymptotic mean stationarity (and thus sample path ergodicity), it is sufficient that the capacity of a channel is (strictly) greater than the sum of the logarithms of the unstable pole magnitudes for memoryless channels and a class of channels with memory. This condition is also necessary under a mild technical condition. Sufficient conditions for the existence of finite average second moments for such systems driven by unbounded noise are provided.**

***Index Terms*—Asymptotic mean stationarity, feedback, Markov chains, nonasymptotic information theory, stochastic control, stochastic stability.**

I. Problem Formulation

THIS paper considers stochastic stabilization of linear systems controlled or estimated over discrete noisy channels with feedback. We consider first a scalar LTI discrete-time system (we consider multidimensional systems in Section IV) described by

$$x_{t+1} = ax_t + bu_t + d_t, \quad t \geq 0. \tag{1}$$

Here, x_t is the state at time t, u_t is the control input, the initial condition x_0 is a second-order random variable, and $\{d_t\}$ is a sequence of zero-mean independent, identically distributed (i.i.d.) Gaussian random variables. It is assumed that $|a| \geq 1$ and $b \neq 0$: the system is open-loop unstable, but it is stabilizable. This system is connected over a discrete noisy channel with a finite capacity to a controller, as shown in Fig. 1.

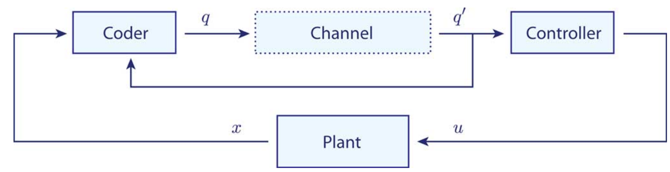

Fig. 1. Control over a discrete noisy channel with feedback.

The controller has access to the information it has received through the channel. The controller in our model estimates the state and then applies its control.

Remark 1.1: We note that the existence of the control can also be regarded as an estimation correction, and all results regarding stability may equivalently be viewed as the stability of the estimation error. Thus, the two problems are identical for such a controllable system and the reader unfamiliar with control theory can simply replace the stability of the state, with the stability of the estimation error.

Recall the following definitions.

Definition 1.1: A finite-alphabet channel with memory is characterized by a sequence of finite input alphabets $\mathcal{M}^{n+1}$, finite output alphabets $\mathcal{M}'^{n+1}$, and a sequence of conditional probability measures $P_n(q'_{[0,n]}|q_{[0,n]})$, from $\mathcal{M}^{n+1} \times \mathcal{M}'^{n+1}$ to $\mathbb{R}$, with

$$q'_{[0,n]} =: \{q'_0, q'_1, \ldots, q'_n\} \quad q_{[0,n]} := \{q_0, q_1, \ldots, q_n\}.$$

Definition 1.2: A discrete memoryless channel (DMC) is characterized by a finite input alphabet $\mathcal{M}$, a finite output alphabet $\mathcal{M}'$, and a conditional probability mass function $P(q'|q)$, from $\mathcal{M} \times \mathcal{M}'$ to $\mathbb{R}$. Let $q_{[0,n]} \in \mathcal{M}^{n+1}$ be a sequence of input symbols, and let $q'_{[0,n]} \in \mathcal{M}'^{n+1}$ be a sequence of output symbols, where $q_k \in \mathcal{M}$ and $q'_k \in \mathcal{M}'$ for all k. Let P^{n+1}_{DMC} denote the joint mass function on the $n+1$-tuple input and output spaces. A DMC from $\mathcal{M}^{n+1}$ to $\mathcal{M}'^{n+1}$ satisfies the following: $P^{n+1}_{\mathrm{DMC}}(q'_{[0,n]}, q_{[0,n]}) = \prod_{k=0}^{n} P_{\mathrm{DMC}}(q'_k, q_k)$, $\forall q_{[0,n]} \in \mathcal{M}^{n+1}$, $q'_{[0,n]} \in \mathcal{M}'^{n+1}$, where q_k, q'_k denote the kth component of the vectors $q_{[0,n]}, q'_{[0,n]}$, respectively. □

In the problem considered, a source coder maps the information at the encoder to corresponding channel inputs. This is done through quantization and a channel encoder. The quantizer outputs are transmitted through a channel, after being subjected to

Manuscript received April 02, 2011; revised January 24, 2012; accepted March 26, 2012. Date of publication June 08, 2012; date of current version September 11, 2012. S. Yüksel was supported by the Natural Sciences and Engineering Research Council of Canada. This paper was presented in part at the 2011 Annual Workshop on Information Theory and Applications, San Diego, CA, and in part at the 2011 IEEE Conference on Decision and Control.

The author is with the Department of Mathematics and Statistics, Queen's University, Kingston, ON K7L 3N6 Canada (e-mail: yuksel@mast.queensu.ca).

Communicated by T. Weissman, Associate Editor for Shannon Theory.

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIT.2012.2204033

a channel encoder. The receiver has access to noisy versions of the quantizer/coder outputs for each time, which we denote by $q'_t \in \mathcal{M}'$. The quantizer and the source coder policy are causal such that the channel input at time $t \geq 0$, q_t, is generated using the information vector I^s_t available at the encoder for $t > 0$

$$I^s_t = \{I^s_{t-1}, x_t, q_{t-1}, q'_{t-1}\}$$

and $I^s_0 = \{\nu_0, x_0\}$, where ν_0 is the probability measure for the initial state. The control policy at time t, also causal, is measurable on the sigma algebra generated by I^c_t, for $t \geq 1$

$$I^c_t = \{I^c_{t-1}, q'_t\}$$

and $I^c_0 = \{\nu_0\}$, and is a mapping to $\mathbb{R}$.

We will call such coding and control policies *admissible* policies.

The goal of this paper is to identify conditions on the channel under which the controlled process $\{x_t\}$ is stochastically stable in sense that $\{x_t\}$ is recurrent, $\{x_t\}$ is asymptotically mean stationary (AMS) and satisfies Birkhoff's sample path ergodic theorem, and that $\lim_{T\to\infty} \frac{1}{T}\sum_{t=0}^{T-1} \|x_t\|^2$ is finite almost surely, under admissible coding and control policies. We will make these notions and the contributions of this paper more precise after we discuss a literature review in the next section. The Appendix contains a review of relevant definitions and results on stochastic stability of Markov chains and ergodic theory.

Here is a brief summary of this paper. In the following, we will first provide a comprehensive literature review. In Section II, we state the main results of this paper. In Section III, we consider extensions to channels with memory, and in Section IV, we consider multidimensional settings. Section V contains the proofs of necessity and sufficiency results for stochastic stabilization. Section VI contains concluding remarks and discusses a number of extensions. This paper ends with Appendix which contains a review of stochastic stability of Markov chains and a discussion on ergodic processes.

A. Literature Review

There is a large literature on stochastic stabilization of sources via coding, both in the information theory and control theory communities.

In the information theory literature, stochastic stability results are established mostly for stationary sources, which are already in some appropriate sense stable sources. In this literature, the stability of the estimation errors as well as the encoder state processes is studied. These systems mainly involve causal and noncausal coding (block coding, as well as sliding-block coding) of stationary sources [29], [48], [53], and AMS sources [35]. Real-time settings such as sigma–delta quantization schemes have also been considered in the literature; see, for example, [36] among others.

There also have been important contributions on noncausal coding of nonstationary/unstable sources: consider the following Gaussian AR process:

$$x_t = -\sum_{k=1}^{m} a_k x_{t-k} + w_k \tag{2}$$

where $\{w_k\}$ is an independent and identical, zero-mean, Gaussian random sequence with variance $E[w_1^2] = \sigma^2$. If the roots of the polynomial $\mathcal{P}(z) := 1 + \sum_{k=1}^{m} a_k z^{-k}$ are all in the interior of the unit circle, then the process is stationary and its rate distortion function (with the distortion being the expected, normalized Euclidean error) is given parametrically (in terms of parameter θ) by the following Kolmogorov's formula [30], [54], obtained by considering the asymptotic distribution of the eigenvalues of the correlation matrix

$$D_\theta = \frac{1}{2\pi}\int_{-\pi}^{\pi} \min(\theta, \frac{1}{g(w)})dw$$
$$R(D_\theta) = \frac{1}{2\pi}\int_{-\pi}^{\pi} \max(\frac{1}{2}(\log\frac{1}{\theta g(w)}), 0)dw$$

with $g(w) = \frac{1}{\sigma^2}|1 + \sum_{k=1} a_k e^{-ikw}|^2$. If at least one root, however, is on or outside the unit circle, the analysis is more involved as the asymptotic eigenvalue distribution contains unbounded components. The authors of [30], [40], and [34] showed that, using the properties of the eigenvalues as well as Jensen's formula for integrations along the unit circle, $R(D_\theta)$ above should be replaced with

$$R(D_\theta) = \frac{1}{2\pi}\int_{-\pi}^{\pi} \max\left(\frac{1}{2}\log(\frac{1}{\theta g(w)}), 0\right)dw + \sum_{k=1}^{m}\frac{1}{2}\max\left(0, \log(|\rho_k|^2)\right) \tag{3}$$

where $\{\rho_k\}$ are the roots of the polynomial $\mathcal{P}$. We refer the reader to a review in [34] regarding rate-distortion results for such nonstationary processes and on the methods used in [30] and [40].

The author of [6] obtained the rate-distortion function for Wiener processes, and in addition, developed a two-part coding scheme, which was later generalized for more general processes in [75] and [78], which we will discuss later further, to unstable Markov processes. The scheme in [6] exploits the independent increment property of Wiener processes.

Thus, an important finding in the aforementioned literature is that the logarithms of the unstable poles in such linear systems appear in the rate-distortion formulations, an issue which has also been observed in the networked control literature, which we will discuss further later. We also wish to emphasize that these coding schemes are noncausal, that is the encoder has access to the entire ensemble before the encoding begins.

In contrast with information theory, due to the practical motivation of sensitivity to delay, the control theory literature has mainly considered causal/zero-delay coding for unstable (or nonstationary) sources, in the context of networked control systems. In the following, we will provide a discussion on the contributions in the literature which are contextually close to our paper.

The authors of [12] studied the tradeoff between delay and reliability, and posed questions leading to an accelerated pace of research efforts on what we today know as networked control problems. The authors of [45], [85], and [66] obtained the minimum lower bound needed for stabilization over noisy channels under a class of assumptions on the system noise and channels.

This result states that for stabilizability under information constraints, in the mean-square sense, a minimum rate needed for stabilizability has to be at least the sum of the logarithms of the unstable poles/eigenvalues in the system; that is

$$\sum_{k=1}^{m} \frac{1}{2} \max\left(0, \log(|\rho_k|^2)\right). \tag{4}$$

Comparing this result with (3), we observe that the rate requirement is not due to causality but due to the (differential) entropy rate of the unstable system.

For coding and information transmission for unstable linear systems, there is an important difference between continuous alphabet and finite-alphabet (discrete) channels as discussed in [95]: when the space is continuous alphabet, we do not necessarily need to consider adaptation in the encoders. On the other hand, when the channel is finite alphabet, and the system is driven by unbounded noise, a static quantizer leads to almost sure instability (see [66, Proposition 5.1] and [95, Th. 4.2]). With this observation, [66] considered a class of variable rate quantizer policies for such unstable linear systems driven by noise, with unbounded support set for its probability measure, controlled over noiseless channels, and obtained necessary and sufficient conditions for the boundedness of the following expression:

$$\limsup_{t\to\infty} E[\|x_t\|^2] < \infty.$$

With fixed rate, [97] obtained a somewhat stronger expression and established a limit

$$\lim_{t\to\infty} E[\|x_t\|^2] < \infty$$

and obtained a scheme which made the state process and the encoder process stochastically stable in the sense that the joint process is a positive Harris recurrent Markov chain and the sample path ergodic theorem is applicable.

The authors of [56] established that when a channel is present in a controlled linear system, under stationarity assumptions, the rate requirement in (4) is necessary for having finite second moments for the state variable. A related argument was made in [95] under the assumption of invariance conditions for the controlled state process under memoryless policies and finite second moments. In this paper, in Theorem 4.1, we will present a very general result along this direction for a general class of channels and a weaker stability notion. Such settings were further considered in the literature. The problem of control over noisy channels has been considered in many publications including [2], [55], [59], [60], [65], [78], [84], [85] among others. Many of the constructive results involve Gaussian channels or erasure channels (some modeled as infinite capacity erasure channels as in [79] and [42]). Other works have considered cases where there is either no disturbance or the disturbance is bounded, with regard to noisy sources and noisy channels. We discuss some of these in the following.

It is to be stressed that the notion of stochastic stability is very important in characterizing the conditions on the channel. The authors of [59] and [60] considered stabilization in the following sense, when the system noise is bounded:

$$\limsup_{t\to\infty} |x_t| < \infty \quad \text{a.s.}$$

and observed that one needs the zero-error capacity (with feedback) to be greater than a particular lower bound. A similar observation was made in [78], which we will discuss further in the following. When the system is driven by noise which admits a probability measure with unbounded support, the aforementioned stability requirement is impossible for an infinite horizon problem, even when the system is open-loop stable, since for any bound there exists almost surely a realization of a noise variable which will be larger.

The authors of [77] and [78] considered systems driven by bounded noise and considered a number of stability criteria: almost sure stability for noise-free systems, moment stability for systems with bounded noise ($\limsup_{t\to\infty} E[|x_t|^p] < \infty$) as well as *stability in probability* (defined in [59]) for systems with bounded noise. Stability in probability is defined as follows: for every $p > 0$, there exists ζ such that $P(|x_t| > \zeta) < p$ for all $t \in \mathbb{N}$. The authors of [77] and [78] also offered a novel and insightful characterization for reliability for controlling unstable processes, named, any-time capacity, as the characterization of channels for which the following criterion can be satisfied:

$$\limsup_{t\to\infty} E[|x_t|^p] < \infty$$

for positive moments p. A channel is α-any-time reliable for a sequential coding scheme if $P(\hat{m}^{t-d}(t) \neq m^{t-d}(t)) \leq K2^{-\alpha d}$ for all t, d. Here, m^{t-d} is the message transmitted at time $t-d$, estimated at time t. One interesting aspect of an any-time decoder is the independence from the delay, with a fixed encoder policy. The authors of [78] state that for a system driven by bounded noise, stabilization is possible if the maximum rate for which an any-time reliability of $2\log_2(|\rho_1|)$ is satisfied is greater than $\log_2(|\rho_1|)$, where ρ_1 is the unstable pole of a linear system.

In a related context, [55], [59], [78], and [58] considered the relevance to Shannon capacity. The authors of [55] observed that when the moment coefficient goes to zero, Shannon capacity provides the right characterization on whether a channel is sufficient or insufficient, when noise is bounded. A parallel argument is provided in [78, Section III.C.1], observing that in the limit when $p \to 0$, capacity should be the right measure for the objective of satisfying *stability in probability*. Their discussion was for bounded-noise signals. The authors of [59] also observed a parallel discussion, again for bounded-noise signals.

With a departure from the bounded-noise assumption, [58] extended the discussion in [78] and studied a more general model of multidimensional systems driven by an unbounded-noise process considering again *stability in probability*. The author of [58] also showed that when the discrete noisy channel has capacity less than $\log_2(|a|)$, where a is defined in (1), there exists no stabilizing scheme, and if the capacity is strictly greater than this number, there exists a stabilizing scheme in the sense of stability in probability.

Many network applications and networked control applications require the access of control and sensor information to be observed intermittently. Toward generating a solution for such problems, [94] and [96] developed random-time state-dependent drift conditions leading to the existence of an invariant distribution possibly with moment constraints, extending the earlier deterministic state-dependent results in [63]. Using drift arguments, [95] considered noisy (both discrete and continuous alphabet) channels, [97] considered noiseless channels, and [94] considered erasure channels for the following stability criteria: the existence of an invariant distribution and the existence of an invariant distribution with finite moments.

The authors of [24], [56], and [57] considered general channels (possibly with memory), and with a connection with Jensen's formula and Bode's sensitivity integral, developed necessary and sufficient rates for stabilization under various networked control settings. The authors of [65] considered erasure channels and obtained necessary and sufficient time-varying rate conditions for control over such channels. The authors of [17] considered second moment stability over a class of Markov channels with feedback and developed necessary and sufficient conditions, for systems driven by an unbounded noise. The authors of [38] considered the stochastic stability of quantizer parameters, parallel to the results of [97].

On the other hand, for more traditional information theoretic settings where the source is revealed at the beginning of transmission, and for cases where causality and delay are not important, the separation principle for source and channel coding results is applicable for ergodic sources and information stable channels. The separation principle for more general setups has been considered in [89], among others. The authors of [92] and [91] studied the optimal causal coding problem over, respectively, a noiseless channel and a noisy channel with noiseless feedback. Unknown sources have been considered in [15]. We also note that when noise is bounded, binning-based strategies, inspired from Wyner–Ziv and Slepian–Wolf coding schemes, are applicable. This type of consideration has been applied in [78], [95], and [37]. Finally, quantizer design for noiseless or bounded-noise systems includes [25], [26], and [43]. Channel coding algorithms for control systems have been presented recently in [70], [83], and [81].

There also has been progress on coding for noisy channels for the transmission of sources with memory. Due to practical relevance, for communication of sources with memory over channels, particular emphasis has been placed on Gaussian channels with feedback. For such channels, the fact that real-time linear schemes are rate-distortion achieving has been observed in [5], [28]; and [2] in a control theoretic context. Aside from such results (which involve matching between rate-distortion achieving test channels and capacity achieving source distributions [28]), capacity is known not to be a good measure of information reliability for channels for real-time (zero-delay or delay sensitive) control and estimation problems; see [91] and [78]. Such general aspects of *comparison of channels* for cost minimization have been investigated in [8] among others.

Also in the information theory literature, performance of information transmission schemes for channels with feedback has been a recurring avenue of research in information theory, for both variable-length and fixed-length coding schemes [14], [22], [41], [76]. In such setups, the source comes from a fixed alphabet, except the sequential setup in [76] and [22].

B. Contributions of This Paper

In view of the discussion above, this paper makes the following contributions. The question: *When does a linear system driven by unbounded noise, controlled over a channel (possibly with memory), satisfy Birkhoff's sample path ergodic theorem (or is asymptotically mean stationary)?* has not been answered to our knowledge. Also, the finite moment conditions for an arbitrary DMC for a system driven by unbounded noise have not been investigated to our knowledge, except for the bounded-noise analysis in [78]. The contributions of this paper are on these two problems. In this paper, we will show that the results in the literature can be strengthened to asymptotic mean stationarity and ergodicity. As a consequence of Kac's lemma [19], *stability in probability* can also be established. We will also consider conditions for finite second moment stability. We will use the random-time state-dependent drift approach [94] to prove our achievability results. Hence, we will find classes of channels under which the controlled process $\{x_t\}$ is stochastically stable in each of the following senses.

1) $\{x_t\}$ is recurrent: There exists a compact set A such that $\{x_t \in A\}$ infinitely often almost surely.
2) $\{x_t\}$ is AMS and satisfies Birkhoff's sample path ergodic theorem. We will establish that Shannon capacity provides a boundary point in the space of channels on whether this objective can be realized or not, provided a mild technical condition holds.
3) $\lim_{T\to\infty} \frac{1}{T}\sum_{t=0}^{T-1} \|x_t\|^2$ exists and is finite almost surely.

II. Stochastic Stabilization Over a DMC

A. Asymptotic Mean Stationarity and n-Ergodicity

Theorem 2.1: For a controlled linear source given in (1) over a DMC under any admissible coding and controller policy, to satisfy the AMS property under the following condition:

$$\liminf_{T\to\infty} \frac{1}{T} h(x_T) \le 0$$

the channel capacity C must satisfy

$$C \ge \log_2(|a|).$$

□

Proof: See the proof of Theorem 4.1 in Section V-A

Remark 2.1: The condition $\liminf_{T\to\infty} \frac{1}{T} h(x_T) \le 0$ is a very weak condition. For example, a stochastic process whose second moment grows subexponentially in time such that

$$\liminf_{T\to\infty} \frac{\log(E[x_T^2])}{T} = 0$$

satisfies this condition.

The aforementioned condition is almost sufficient as well, as we state in the following.

Theorem 2.2: For the existence of a compact coordinate recurrent set (see Definition 7.4), the following is sufficient: the channel capacity C satisfies $C > \log_2(|a|)$. □

Proof: See Section V-C. □

For the proof, we consider the following update algorithm. The algorithm and its variations have been used in source coding and networked control literature: see, for example, the earlier papers [29], [48], and more recent ones [13], [58]–[60], [66], [97], [98]. Our contribution is primarily on the stability analysis.

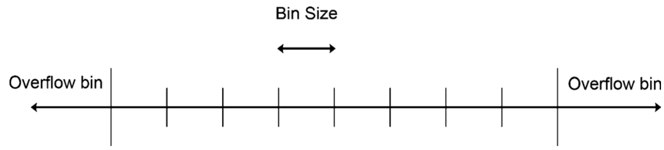

Fig. 2. Uniform quantizer with a single overflow bin.

Let n be a given block length. Consider the following setup. We will consider a class of uniform quantizers, defined by two parameters, with bin size $\Delta > 0$, and an even number $K(n) \geq 2$ (see Fig. 2). The uniform quantizer map is defined as follows: for $k = 1, 2 \ldots, K(n)$

$$Q^{\Delta}_{K(n)}(x) = \begin{cases} (k - \frac{1}{2}(K(n)+1))\Delta \\ \text{if} \quad x \in [(k-1-\frac{1}{2}K(n))\Delta, (k-\frac{1}{2}K(n))\Delta) \\ (\frac{1}{2}(K(n)-1))\Delta, \quad \text{if} \quad x = \frac{1}{2}K(n)\Delta \\ \mathcal{Z}, \qquad \text{if} \quad |x| > \frac{1}{2}K(n)\Delta \end{cases}$$

where $\mathcal{Z}$ denotes the overflow symbol in the quantizer. We define $\{x : Q^{\Delta}_{K(n)}(x) \neq \mathcal{Z}\}$ to be the *granular region* of the quantizer.

At every sampling instant $t = kn, k = 0, 1, 2, \ldots$, the source coder $\mathcal{E}^s_t$ quantizes output symbols in $\mathbb{R} \cup \{\mathcal{Z}\}$ to a set $\mathcal{M}(n) = \{1, 2, \ldots, K(n)+1\}$. A channel encoder $\mathcal{E}^c_t$ maps the elements in $\mathcal{M}(n)$ to corresponding channel inputs $q_{[kn,(k+1)n-1]} \in \mathcal{M}^n$.

For each time $t = kn-1$, $k = 1, 2, 3, \ldots$, the channel decoder applies a mapping $\mathcal{D}_{tn} : \mathcal{M}'^n \to \mathcal{M}(n)$, such that

$$c'_{(k+1)n-1} = \mathcal{D}_{kn}(q'_{[kn,(k+1)n-1]}).$$

Finally, the controller runs an estimator

$$\hat{x}_{kn} = (\mathcal{E}^s_{kn})^{-1}(c'_{(k+1)n-1}) \times 1_{\{c'_{(k+1)n-1} \neq \mathcal{Z}\}} + 0 \times 1_{\{c'_{(k+1)n-1} = \mathcal{Z}\}}$$

where 1_E denotes the indicator function for event E. Hence, when the decoder output is the overflow symbol, the estimation output is 0.

We consider quantizers that are adaptive. In our setup, the bin size of the uniform quantizer acts as the state of the quantizer. At time kn the bin size Δ_{kn} is assumed to be a function of the previous state $\Delta_{(k-1)n}$ and the past n channel outputs. We assume that the encoder has access to the previous channel outputs. Thus, such a quantizer is implementable at both the encoder and the decoder.

With $K(n) > \lceil |a|^n \rceil$, $R = \log_2(K(n)+1)$, let us define $R'(n) = \log_2(K(n))$ and let

$$R'(n) > n \log_2(\frac{|a|}{\alpha})$$

for some $\alpha, 0 < \alpha < 1$ and $\delta > 0$. When clear from the context, we will drop the index n in $R'(n)$.

We will consider the following update rules in the controller actions and the quantizers. For $t \geq 0$ and with $\Delta_0 > L$ for some $L \in \mathbb{R}_+$, and $\hat{x}_0 \in \mathbb{R}$, consider: for $t = kn$, $k \in \mathbb{N}$

$$u_t = -1_{\{t=(k+1)n-1\}} \frac{a^n}{b} \hat{x}_{kn}$$
$$\Delta_{(k+1)n} = \Delta_{kn} \bar{Q}(\Delta_{kn}, c'_{(k+1)n-1}) \tag{5}$$

where c' denotes the decoder output variable. If we use $\delta > 0$ and $L > 0$ such that

$$\bar{Q}(\Delta, c') = (|a|+\delta)^n \quad \text{if} \quad c' = \mathcal{Z}$$
$$\bar{Q}(\Delta, c') = \alpha^n \quad \text{if} \quad c' \neq \mathcal{Z}, \Delta \geq L$$
$$\bar{Q}(\Delta, c') = 1 \quad \text{if} \quad c' \neq \mathcal{Z}, \Delta < L \tag{6}$$

we will show that a recurrent set exists. The above imply that $\Delta_t \geq L\alpha^n =: L'$ for all $t \geq 0$.

Thus, we have three main events: when the decoder output is the overflow symbol, the quantizer is zoomed out (with a coefficient of $(|a|+\delta)^n$). When the decoder output is not the overflow symbol $\mathcal{Z}$, the quantizer is zoomed in (with a coefficient of α^n) if the current bin size is greater than or equal to L, and otherwise the bin size does not change.

We will establish our stability result through random-time stochastic drift criterion of Theorem 7.2, developed in [94] and [96]. This is because of the fact that the quantizer helps reduce the uncertainty on the system state only when the state is in the *granular* region of the quantizer. The times when the state is in this region are random. The reader is referred to Section B in the Appendix for a detailed discussion on the drift criteria.

In the following, we make the quantizer bin size process space countable and as a result establish the irreducibility of the sampled process (x_{tn}, Δ_{tn}).

Theorem 2.3: For an adaptive quantizer satisfying Theorem 2.2, suppose that the quantizer bin sizes are such that their logarithms are integer multiples of some scalar s, and $\log_2(\bar{Q}(\cdot))$ takes values in integer multiples of s. Suppose the integers taken are relatively prime (that is they share no common divisors except for 1). Then the sampled process (x_{tn}, Δ_{tn}) forms a positive Harris recurrent Markov chain at sampling times on the space of admissible quantizer bins and state values. □

Proof: See Section V-D. □

Theorem 2.4: Under the conditions of Theorems 2.2 and 2.3, the process $\{x_t, \Delta_t\}$ is n-stationary, n-ergodic, and hence AMS. □

Proof: See Section V-E. □

The proof follows from the observation that a positive Harris recurrent Markov chain is ergodic. It uses the property that if a sampled process is a positive Harris recurrent Markov chain, and if the intersampling time is fixed, with a time-homogenous update in the intersampling times, then the process is mixing, n-ergodic and n-stationary.

B. Quadratic Stability and Finite Second Moment

In this section, we discuss quadratic and finite second moment stability. Such an objective is important in applications. In control theory, quadratic cost functions are the most popular ones for linear and Gaussian systems. Furthermore, the two-part coding scheme of Berger in [5] can be generalized for more general unstable systems if one can prove finite moment boundedness of sampled end points.

For a given coding scheme with block length n and a message set $\mathcal{M}(n) = \{1, 2, \ldots, K(n)+1\}$, and a decoding function $\gamma : \mathcal{M}'^n \to \{1, 2, \ldots, K(n)+1\}$ define three types of errors.

1) Type I-A: Error from a granular symbol to another granular symbol. We define a bound for such errors. Define $P^e_{g|g}(n)$ to be
$$\max_{c \in \mathcal{M}(n) \setminus \mathcal{Z}} P(\gamma(q'_{[0,n-1]}) \neq c, \gamma(q'_{[0,n-1]}) \neq \mathcal{Z} | c)$$
where conditioning on c means that the symbol c is transmitted.

2) Type I-B: Error from a granular symbol to $\mathcal{Z}$. We define the following:
$$P^e_{g|g}(n) := \max_{c \in \mathcal{M}(n) \setminus \mathcal{Z}} P(\gamma(q'_{[0,n-1]}) = \mathcal{Z} | c).$$

3) Type II: Error from $\mathcal{Z}$ to a granular symbol
$$P^e_{g|\mathcal{Z}}(n) := P(\gamma(q'_{[0,n-1]}) \neq \mathcal{Z} | \mathcal{Z}).$$

Type II error will be shown to be crucial in the analysis of the error exponent. Type I-A and I-B will be important for establishing the drift properties. We summarize our results in the following.

Theorem 2.5: A sufficient condition for quadratic stability (for the joint (x_t, Δ_t) process) over a DMC is that
$$\lim_{n\to\infty} (\frac{1}{n} \log(P^e_{\mathcal{Z}|g}(n)) + 2\log(|a| + \delta) < 0$$
$$\lim_{n\to\infty} (\kappa \frac{1}{n} \log(P^e_{g|\mathcal{Z}}(n)) + 2\log(|a| + \delta) < 0$$
$$\lim_{n\to\infty} (\kappa \frac{1}{n} \log(P^e_{g|g}(n)) + 2\log(|a| + \delta) + 2\kappa \log(\alpha) < 0$$
$$R'(n) > n \log_2(|a|/\alpha)$$
with
$$\kappa < \frac{1}{\log_{\frac{|a|+\delta}{|a|}}(\frac{|a|+\delta}{\alpha})}.$$
□

Proof: See Section V-F. □

Let us define
$$\bar{P}_e(n) := \max_{c \in \mathcal{M}(n)} P(\gamma(q'_{[0,n-1]}) \neq c | c \quad \text{is transmitted}).$$

When the block length is clear from the context, we drop the index n. We have the following corollary to Theorem 2.5.

Corollary 2.1: A sufficient condition for quadratic stability (for the joint (x_t, Δ_t) process) over a DMC is that
$$\lim_{n\to\infty} (\kappa \frac{1}{n} \log(\bar{P}_e(n)) + 2\log(|a| + \delta) < 0$$
with rate $R'(n) > n \log_2(\frac{|a|}{\alpha})$. □

Remark 2.2: For a DMC with block length n, Shannon's random coding [27] satisfies
$$P_e(n) \leq e^{-nE(R) + o(n)}$$
uniformly for all codewords $c \in \{1, 2, \ldots, \mathcal{M}(n)\}$ with c' being the decoder output (thus, the random exponent also applies uniformly over the set). Here, $\frac{o(n)}{n} \to 0$ as $n \to \infty$ and $E(R) > 0$ for $0 < R < C$. Thus, under the aforementioned conditions, the exponent under random coding should satisfy $E(R) > \frac{2\log_2(|a|+\delta)}{\kappa}$.

Remark 2.3: The error exponent with feedback is typically improved with feedback, unlike capacity of DMCs. However, a precise solution to the error exponent problem of fixed-length block coding with noiseless feedback is not known. Some partial results have been reported in [21] (in particular, the sphere packing bound is optimal for a class of symmetric channels for rates above a critical number even with feedback), [4, Ch. 10], [10], [18], [23], [39], [99], and [67]. Particularly related to this section, [10] has considered the exponent maximization for a special message symbol, at rates close to capacity. At the end of this paper, a discussion for variable-length coding, in the context of Burnashev's [14] setup, will be discussed along with some other contributions in the literature. In case feedback is not used, Gilbert exponent [69] for low-rate regions may provide better bounds than the random coding exponent. □

Zero-Error Transmission for $\mathcal{Z}$: An important practical setup would be the case when $\mathcal{Z}$ is transmitted with no error and is not confused with messages from the granular region. We state this as follows.

Assumption A0: We have that $P^e_{\mathcal{Z}|g}(n) = P^e_{g|\mathcal{Z}}(n) = 0$ for $n \geq n_0$ for some $n_0 \in \mathbb{N}$. □

Theorem 2.6: Under Assumption **A0**, a sufficient condition for quadratic stability is
$$\lim_{n\to\infty} (\bar{P}_e(n))(|a| + \delta)^{2n} < 1$$
with rate $R'(n) > n \log_2(\frac{|a|}{\alpha})$ and $\kappa > 1/2$. □

Proof: See Section V-G. □

Remark 2.4: The result under (**A0**) is related to the notion of any-time capacity proposed by Sahai and Mitter [78]. We note that Sahai and Mitter considered also a block-coding setup, for the case when the noise is bounded, and were able to obtain a similar rate/reliability criterion as above. It is worth emphasizing that the reliability for sending one symbol $\mathcal{Z}$ for the under-zoom phase allows an improvement in the reliability requirements drastically. □

III. Channels With Memory

Definition 3.1: Let **Class A** be the set of channels which satisfy the following two properties.

a) The following Markov chain condition holds:
$$q'_t \leftrightarrow q_t, q_{[0,t-1]}, q'_{[0,t-1]} \leftrightarrow x_{[0,t]}$$
for all $t \geq 0$.

b) The channel capacity with feedback is given by
$$C = \lim_{T\to\infty} \max_{\{P(q_t | q_{[0,t-1]}, q'_{[0,t-1]})\}} \frac{1}{T} I(q_{[0,T-1]} \to q'_{[0,T-1]}) \quad (7)$$

where $0 \leq t \leq T-1$ and the directed mutual information is defined by

$$I(q_{[0,T-1]} \to q'_{[0,T-1]}) := \sum_{t=1}^{T-1} I(q_{[0,t]}; q'_t | q'_{[0,t-1]}) + I(q_0; q'_0).$$

DMCs naturally belong to this class of channels. For such channels, it is known that feedback does not increase the capacity. Such a class also includes finite state stationary Markov channels which are indecomposable [72], and non-Markov channels which satisfy certain symmetry properties [82]. Further examples are studied in [87] and [20].

Theorem 3.1: For a linear system controlled over a noisy channel with memory with feedback in **Class A**, if the channel capacity is less than $\log_2(|a|)$, then the AMS property under the following condition:

$$\liminf_{T\to\infty} \frac{1}{T} h(x_T) \leq 0$$

cannot be satisfied under any policy. □

Proof: See the proof of Theorem 4.1 in Section V-A. □

The proof of the above is presented in Section V-A.[1] If the channel is not information stable, then information spectrum methods lead to pessimistic realizations of capacity (known as the $\liminf$ in probability of the normalized information density; see [87] and [90]). We do not consider such channels in this paper, although the proof is generalizable to some cases when the channel state is Markov and the worst case initial input state is considered as in [72].

IV. Higher Order Sources

The proposed technique is also applicable to a class of settings for the multidimensional setup. Observe that a higher order ARMA model of the form (2) can be written in the following form:

$$x_{t+1} = Ax_t + Bu_t + Gd_t \tag{8}$$

where $x_t \in \mathbb{R}^N$ is the state at time t, u_t is the control input, and $\{d_t\}$ is a sequence of zero-mean i.i.d. zero-mean Gaussian random vectors of appropriate dimensions. Here, A is the system matrix with at least one eigenvalue greater than 1 in magnitude, that is, the system is open-loop unstable. Furthermore, (A, B) and (A, G) are controllable pairs, that is, the state process can be traced in finite time from any point in $\mathbb{R}^N$ to any other point in at most N time stages, by either the controller or the Gaussian noise process.

In the following, we assume that all modes with eigenvalues $\{\lambda_i, 1 \leq i \leq N\}$ of A are unstable, that is have magnitudes greater than or equal to 1. There is no loss here since if some eigenvalues are stable, by a similarity transformation, the unstable modes can be decoupled from stable modes; stable modes are already recurrent.

Theorem 4.1: Consider a multidimensional linear system with unstable eigenvalues, that is $|\lambda_i| \geq 1$ for $i = 1, \ldots, N$. For such a system controlled over a noisy channel with memory with feedback in **Class A**, if the channel capacity satisfies

$$C < \sum_i \log_2(|\lambda_i|)$$

there does not exist a stabilizing coding and control scheme with the property $\liminf_{T\to\infty} \frac{1}{T} h(x_T) \leq 0$. □

Proof: See Section V-A. □

For sufficiency, we will assume that A is a diagonalizable matrix with real eigenvalues (a sufficient condition being that the poles of the system are distinct real). In this case, the analysis follows from the discussion for scalar systems; as the identical recurrence analysis for the scalar case is applicable for each of the subsystems along each of the eigenvectors. The possibly correlated noise components will lead to the recurrence analysis discussed earlier. We thus have the following result.

Theorem 4.2: Consider a multidimensional system with a diagonalizable matrix A. If the Shannon capacity of the (DMC) channel used in the controlled system satisfies

$$C > \sum_{|\lambda_i|>1} \log_2(|\lambda_i|)$$

there exists a stabilizing (in the AMS sense) scheme. □

Proof: See Section V-H. □

The result can be extended to a case where the matrix A is in a Jordan form. Such an extension entails considerable details in the analysis for the stopping time computations and has not been included in the paper. A discussion for the special case of discrete noiseless channels is contained in [46] in the context of decentralized linear systems.

V. Proofs

A. Proof of Theorem 4.1

We present the proof for a multidimensional system since this case is more general. For channels under **Class A** (which includes the DMCs as a special case), the capacity is given by (7).

Let us define

$$R_T := \max_{\{P(q_t|q_{[0,t-1]}, q'_{[0,t-1]}), 0\leq t\leq T-1\}} \frac{1}{T} \sum_{t=0}^{T-1} I(q'_t; q_{[0,t]} | q'_{[0,t-1]}).$$

Observe that for $t > 0$

$$\begin{aligned}
& I(q'_t; q_{[0,t]} | q'_{[0,t-1]}) \\
&= H(q'_t | q'_{[0,t-1]}) - H(q'_t | q_{[0,t]}, q'_{[0,t-1]}) \\
&= H(q'_t | q'_{[0,t-1]}) - H(q'_t | q_{[0,t]}, x_t, q'_{[0,t-1]}) \qquad (9) \\
&\geq H(q'_t | q'_{[0,t-1]}) - H(q'_t | x_t, q'_{[0,t-1]}) \\
&= I(x_t; q'_t | q'_{[0,t-1]}). \qquad (10)
\end{aligned}$$

[1] One can also obtain a positive result: if the channel capacity is greater than $\log_2(|a|)$, then there exists a coding scheme leading to an AMS state process provided that the channel restarts itself with the sending of a new block. If this assumption does not hold, then using the proofs in this paper we can prove coordinate-recurrence under this condition. For the AMS property, however, new tools will be required. Our proof would have to be modified to account for the non-Markovian nature of the sampled state and quantizer process.

Here, (9) follows from the assumption that the channel is of **Class A**. It follows that since for two sequences such that $a_n \geq b_n$: $\limsup_n a_n \geq \limsup_n b_n$ and R_T is assumed to have a limit

$$\begin{aligned}
&\lim_{T\to\infty} R_T \\
&\geq \limsup_{T\to\infty} \frac{1}{T}\Bigg(\sum_{t=1}^{T-1} I(x_t; q'_t | q'_{[0,t-1]})) + I(x_0; q'_0)\Bigg) \\
&= \limsup_{T\to\infty} \frac{1}{T}\Bigg(\sum_{t=1}^{T-1}\Big(h(x_t|q'_{[0,t-1]}) - h(x_t|q'_{[0,t]})\Big) + I(x_0; q'_0)\Bigg) \\
&= \limsup_{T\to\infty} \frac{1}{T}\Bigg(\sum_{t=1}^{T-1}\Big(h(Ax_{t-1}+Gd_{t-1}+Bu_{t-1}|q'_{[0,t-1]}) - h(x_t|q'_{[0,t]})\Big) + I(x_0; q'_0)\Bigg) \\
&= \limsup_{T\to\infty} \frac{1}{T}\Bigg(\sum_{t=1}^{T-1}\Big(h(Ax_{t-1}+Gd_{t-1}|q'_{[0,t-1]}) - h(x_t|q'_{[0,t]})\Big) + I(x_0; q'_0)\Bigg) \quad (11) \\
&\geq \limsup_{T\to\infty} \frac{1}{T}\Bigg(\sum_{t=1}^{T-1}\Big(h(Ax_{t-1}+Gd_{t-1}|q'_{[0,t-1]}, d_{t-1}) - h(x_t|q'_{[0,t]})\Big) + I(x_0; q'_0)\Bigg) \quad (12) \\
&= \limsup_{T\to\infty} \frac{1}{T}\Bigg(\sum_{t=1}^{T-1}\Big(h(Ax_{t-1}|q'_{[0,t-1]}, d_{t-1}) - h(x_t|q'_{[0,t]})\Big) + I(x_0; q'_0)\Bigg) \\
&= \limsup_{T\to\infty} \frac{1}{T}\Bigg(\sum_{t=1}^{T-1}\Big(h(Ax_{t-1}|q'_{[0,t-1]}) - h(x_t|q'_{[0,t]})\Big) + I(x_0; q'_0)\Bigg) \quad (13) \\
&= \limsup_{T\to\infty} \frac{1}{T}\Bigg(\sum_{t=1}^{T-1}\Big(\log_2(|A|) + h(x_{t-1}|q'_{[0,t-1]}) - h(x_t|q'_{[0,t]})\Big) + I(x_0; q'_0)\Bigg) \\
&= \limsup_{T\to\infty} \frac{1}{T}\Bigg(\Big(\sum_{t=1}^{T-1}\log_2(|A|)\Big) + h(x_0|q'_0) - h(x_{T-1}|q'_{[0,T-1]}) + I(x_0; q'_0)\Bigg) \\
&= \log_2(|A|) - \liminf_{T\to\infty}\Big(\frac{1}{T}h(x_{T-1}|q'_{[0,T-1]})\Big) \\
&\geq \log_2(|A|) - \liminf_{T\to\infty}\Big(\frac{1}{T}h(x_{T-1})\Big) \quad (14) \\
&\geq \log_2(|A|).
\end{aligned}$$

Equation (11) follows from the fact that the control action is a function of the past channel outputs, (12) follows from the fact that conditioning does not increase entropy, and (13) from the observation that $\{d_t\}$ is an independent process. Equation (14) follows from conditioning. The other equations follow from the properties of mutual information. By the hypothesis, $\liminf_{t\to\infty} \frac{1}{t}h(x_t) \leq 0$, it must be that $\lim_{T\to\infty} R_T \geq \log_2(|A|)$. Thus, the capacity also needs to satisfy this bound. □

B. Stopping Time Analysis

This section presents an important supporting result on stopping time distributions, which is key in the application of Theorem 7.2 for the stochastic stability results. We begin with the following.

Lemma 5.1: Let $\mathcal{B}(\mathbb{R}\times\mathbb{R}_+)$ denote the Borel σ-field on $\mathbb{R}\times\mathbb{R}_+$. It follows that

$$\begin{aligned}
&P\Big((x_{kn}, \Delta_{kn}) \in (C\times D) | \{(x_{sn}, \Delta_{sn}), s<k\}\Big) \\
&\quad = P\Big((x_{kn}, \Delta_{kn}) \in (C\times D) | (x_{(k-1)n}, \Delta_{(k-1)n})\Big)
\end{aligned}$$

$\forall (C\times D) \in \mathcal{B}(\mathbb{R}\times\mathbb{R}_+)$, i.e., (x_{tn}, Δ_{tn}) is a Markov chain. □

The above follows from the observations that the channel is memoryless, the encoding only depends on the most recent samples of the state and the quantizer, and the control policies use the channel outputs received in the last block, which stochastically only depend on the parameters in the previous block.

Let us define $h_t := \frac{x_t}{\Delta_t 2^{R'-1}}$. We will say that the quantizer is perfectly zoomed when $|h_t| \leq 1$, and under-zoomed otherwise.

Define a sequence of stopping times for the perfect-zoom case with (where the initial state is perfectly zoomed at τ_0)

$$\tau_0 = 0, \quad \tau_{z+1} = \inf\{kn > \tau_z : |h_{kn}| \leq 1\}, \quad z,\ k \in \mathbb{Z}_+. \tag{15}$$

As discussed in Section II-B, there will be three types of errors.

1) Type I-A: Error from a granular symbol to another granular symbol. In this case, the quantizer will zoom in, yet an incorrect control action will be applied. As in Section II-B, $P^e_{g|g}(n)$ is an upper bound for such an error.
2) Type I-B: Error from a granular symbol to $\mathcal{Z}$. In this case, no control action will be applied and the quantizer will be zoomed out. As in Section II-B, $P^e_{\mathcal{Z}|g}(n)$ is an upper bound for such an error.
3) Type II: Error from $\mathcal{Z}$ to a granular symbol. At consecutive time stages, until the next stopping time, the quantizer should ideally zoom out. Hence, this error may take place in subsequent time stages (since at time 0 the quantizer is zoomed, this does not take place). The consequence of such an error is that the quantizer will be zoomed in and an incorrect control action will be applied. Let

$$\begin{aligned}
P_e(n) &:= P^e_{g|\mathcal{Z}}(n) \\
&= P(\gamma(q'_{[0,n-1]}) \neq \mathcal{Z} | \mathcal{Z} \quad \text{is transmitted}).
\end{aligned}$$

We will drop the dependence on n, when the block length is clear from the context.

Lemma 5.2: The discrete probability distribution $P(\tau_{z+1} - \tau_z | x_{\tau_z}, \Delta_{\tau_z})$ is asymptotically, in the limit of large Δ_{τ_z}, dominated (majorized) by a geometrically distributed measure. That is, for $k, \geq \lceil 1/\kappa \rceil + 1$

$$P(\tau_{z+1} - \tau_z \geq kn | x_{\tau_z}, \Delta_{\tau_z}) \leq \Xi(\Delta_{\tau_z}) \Bigg((1 - P^e_{g|g} - P^e_{\mathcal{Z}|g})(e P_e^{(\kappa)})^{k-2} + P^e_{g|g}(e P_e^{(\kappa - \frac{1-\kappa}{k-2})})^{k-2} + (P^e_{\mathcal{Z}|g})(e P_e^{(\kappa + \frac{\kappa}{k-2})})^{k-2} \Bigg) \quad (16)$$

where $\Xi(\Delta_{\tau_z}) < \infty$ and $\Xi(\Delta_{\tau_z}) \to 1$ as $\Delta_{\tau_z} \to \infty$ for every fixed n, uniformly in $|h_0| \leq 1$ and

$$\kappa < \frac{1}{\log_{\frac{|a|+\delta}{|a|}}(\frac{|a|+\delta}{\alpha})}. \quad (17)$$

□

Proof: Denote for $k \in \mathbb{N}$

$$\Theta_k := P(\tau_{z+1} - \tau_z \geq kn | x_{\tau_z}, \Delta_{\tau_z}). \quad (18)$$

Without any loss, let $z = 0$ and $\tau_0 = 0$, so that $\Theta_k = P(\tau_1 \geq kn | x_0, \Delta_0)$.

Before proceeding with the proof, we highlight the technical difficulties that will arise when the quantizer is in the under-zoom phase. As elaborated on above, the errors at time 0 are crucial for obtaining the error bounds: at time 0, at most with probability $P^e_{g|g}(n)$, an error will take place so that the quantizer will be zoomed in, yet an incorrect control signal will be applied. With probability at most $P^e_{\mathcal{Z}|g}(n)$, an error will take place so that no control action is applied and the quantizer is zoomed out. At consecutive time stages, until the next stopping time, the quantizer should ideally zoom out but an error takes place with probability $P^e_{g|\mathcal{Z}}(n)$ and leads the quantizer to be zoomed in, and a control action to be applied. Our analysis below will address all of these issues.

In the following, we will assume that the probabilities are conditioned on particular x_0, Δ_0 values, to ease the notational presentation.

We first consider the case where there is an intra-granular, Type I-A, error at time 0, which takes place at most with probability $P^e_{g|g}$ (this happens to be the worst case error for the stopping time distribution). Now

$$\begin{aligned} &P(\tau_1 \geq kn | \text{Type I-A error at time 0}) \\ &= P\Bigg(\bigcap_{m=1}^{k-1} (|h_{mn}| > 1) | \text{Type I-A error at time 0} \Bigg) \\ &= P\Bigg(\bigcap_{m=1}^{k-1} (|x_{mn}| \geq 2^{R'-1}(|a| + \delta)^{(m-s_m-1)n} \times \alpha^{(1+s_m)n} \Delta_0) \Bigg) \\ &= P\Bigg(\bigcap_{m=1}^{k-1} (|a^{mn}(x_0 + \sum_{i=0}^{mn-1} a^{-i-1}(d_i + u_i))| \geq 2^{R'-1}(|a| + \delta)^{(m-s_m-1)n} \alpha^{(1+s_m)n} \Delta_0) \Bigg) \\ &= P\Bigg(\bigcap_{m=1}^{k-1} (|(x_0 + \sum_{i=0}^{mn-1} a^{-i-1}(d_i + u_i))| \geq \frac{2^{R'-1}\alpha^n}{|a^n|} (\frac{|a| + \delta}{|a|})^{(m-1)n} (\frac{\alpha}{|a| + \delta})^{(s_m)n} \Delta_0) \Bigg). \end{aligned} \quad (19)$$

In the above, s_m is the number of errors in the transmissions that have taken place until (but not including) time m, except for the one at time 0. An error at time 0 would possibly lead to a further enlargement of the bin size with nonzero probability, whereas no error at time 0 leads to a strict decrease in the bin size.

The study for the number of errors is crucial for analyzing the stopping time distributions. In the following, we will condition on the number of erroneous transmissions for k successive block codings for the under-zoomed phase. Suppose that for $k > 1$ there are s_k total erroneous transmissions in the time stages $\{n, 2n, \ldots, (k-1)n\}$ when the state is in fact under-zoomed, but the controller interprets the received transmission as a successful one. Thus, we take $s_1 = 0$.

Let $\zeta_1, \zeta_2, \ldots, \zeta_{s_{k-1}}$ be the time stages when errors take place, such that

$$\zeta_{t+1} : \min(\min(m > \zeta_t : c'_{nm} \neq c_{nm}), k-1), \quad \zeta_0 = 0$$

so that $\zeta_{s_{k-1}+1} = k - 1$ or $\zeta_{s_{k-1}} = k - 1$ and define $\eta_t = \zeta_{t+1} - \zeta_t$.

In the time interval $[\zeta_t n + 1, \zeta_{t+1} n - 1]$ the system is open loop, that is, there is no control signal applied, as there is no misinterpretation by the controller. However, there will be a nonzero control signal at times $\{\zeta_k n,\ k \geq 0\}$. These are, however, upper bounded by the ranges of the quantizers at the corresponding time stages. That is, when an erroneous interpretation by the controller arises, the control applied $-(a^n/b)u_{(\zeta_z+1)n-1}$ lives in the set: $\{a^n(-2^{R'-1} + k - (1/2))\Delta_{\zeta_z}, 1 \leq k \leq 2^{R'}\}$.

From (19), we obtain (20) which is shown at the bottom of the next page. Regarding (20), let us now observe the following:

$$\begin{aligned} &P\Bigg(\Bigg| a^{\zeta_m n}(x_0 + \sum_{i=0}^{\zeta_m n-1} a^{-i-1} d_i + \sum_{i=0}^{m-1} a^{(-\zeta_i - 1)n} u_{(\zeta_i+1)n-1}) \Bigg| \geq 2^{R'-1}(|a| + \delta)^{(\zeta_m - s_m - 1)n} \alpha^{(1+s_m)n} \Delta_0 \Bigg) \\ &\leq P\Bigg(|\sum_{i=0}^{\zeta_m n-1} a^{-i-1} d_i| \geq 2^{R'-1}(\frac{|a| + \delta}{|a|})^{(\zeta_m - s_m - 1)n}(\frac{\alpha}{|a|})^{(1+s_m)n} \Delta_0 - |x_0 + \sum_{i=0}^{m-1} a^{(-\zeta_i - 1)n} u_{(\zeta_i+1)n-1}| \Bigg). \end{aligned}$$

Since the control signal $u_{(\zeta_i+1)n-1}$ lives in: $\{a^n(-2^{R'-1} + k - (1/2))\Delta_{(\zeta_i)n}, 1 \leq k \leq 2^{R'}\}$, conditioned on having s_{k-1} errors in the transmissions, the bound writes as (22), which is shown at the bottom of the next page, where $\bar{d} = \sum_{i=0}^{\infty} a^{-i-1} d_i$ is a zero-mean Gaussian random variable with variance $\frac{E[d^2]a^{-2}}{1-a^{-2}}$. Here, (21), shown at the bottom of the next page, considers the worst case when even if the quantizer is zoomed, the controller incorrectly picks the worst case control signal and the chain rule for the total probability: for two events A, B:

$P(A,B) \leq \min(P(A),P(B))$. The last inequality follows since $P(|\bar{d}| \geq \alpha_B) \geq P(|\sum_{i=0}^{\zeta_m n-1} a^{-i-1} d_i| \geq \alpha_B)$ for any $\alpha_B \in \mathbb{R}$.

Now, let us consider $m = k-1$. In this case, (23), shown at the bottom of the next page, follows, where in the last inequality we observe that $|x_0| \leq 2^{R'-1}\Delta_0$, since the state is zoomed at this time.

In bounding the stopping time distribution, we will consider the condition that

$$(k-1)-(s_{k-1}+1)(\log_{\frac{(|a|+\delta)}{|a|}}(\frac{|a|+\delta}{\alpha})) > \alpha_A(s_{k-1}+1) \quad (24)$$

for some arbitrarily small but positive α_A, to be able to establish that

$$\left(1 - \sum_{i=1}^{m-1}(\frac{|a|+\delta}{|a|})^{(\zeta_i-\zeta_m)n}(\frac{\alpha}{|a|})^{(s_i-s_m)n}(1-2^{-R'})\right) > 0 \quad (25)$$

and that $(\frac{|a|+\delta}{|a|})^{(\zeta_{k-1}-s_{k-1})n}(\frac{\alpha}{|a|+\delta})^{(s_{k-1})n} > 2$ for sufficiently large n. Now, there exists m such that

$$(\frac{|a|+\delta}{|a|})^{\zeta_m n}(\frac{\alpha}{|a+\delta|})^{s_m n} \geq (\frac{|a|+\delta}{|a|})^{(\zeta_{k-1}-s_{k-1})n}(\frac{\alpha}{|a|+\delta})^{(s_{k-1})n}$$

$$\begin{aligned}
&P\Big(\bigcap_{m=1}^{k-1}(|h_{mn}|>1)|\text{Type I-A error at time } 0\Big)\\
&\leq P\Big(\bigcap_{m=1}^{k-1}\Big(|a^{mn}(x_0+\sum_{i=0}^{mn-1}a^{-i-1}(d_i+u_i))| \geq 2^{R'-1}(|a|+\delta)^{(m-s_m-1)n}\alpha^{(1+s_m)n}\Delta_0\Big)\Big)\\
&\leq P\Big(\bigcup_{p=0}^{k-2}\Big(\{s_{k-1}=p\}\bigcap\Big\{\bigcap_{m=1}^{p}(|a^{\zeta_m n}(x_0+\sum_{i=0}^{\zeta_m n-1}a^{-i-1}d_i+\sum_{i=0}^{m-1}a^{(-\zeta_i-1)n}u_{(\zeta_i+1)n-1})|\\
&\qquad\qquad \geq 2^{R'-1}(|a|+\delta)^{(\zeta_m-s_m-1)n}\alpha^{(1+s_m)n}\Delta_0)\Big\}\Big)\Big)\\
&=\sum_{p=0}^{k-2}\binom{k-2}{p}(P^e_{g|\mathcal{Z}})^p(1-(P^e_{g|\mathcal{Z}}))^{k-1-p}1_{\{s_{k-1}=p\}}P\Big(\bigcap_{m=1}^{p}(|a^{\zeta_m n}(x_0+\sum_{i=0}^{\zeta_m n-1}a^{-i-1}d_i+\sum_{i=0}^{m-1}a^{(-\zeta_i-1)n}u_{(\zeta_i+1)n-1})|\\
&\qquad\qquad \geq 2^{R'-1}(|a|+\delta)^{(\zeta_m-s_m-1)n}\alpha^{(1+s_m)n}\Delta_0\Big|s_{k-1}=p\Big)
\end{aligned} \quad (20)$$

$$\begin{aligned}
&P\Big\{\bigcap_{m=1}^{p}\Big(|a^{\zeta_m n}(x_0+\sum_{i=0}^{\zeta_m n-1}a^{-i-1}d_i+\sum_{i=0}^{m-1}a^{(-\zeta_i-1)n}u_{(\zeta_i+1)n-1})| \geq 2^{R'-1}(|a|+\delta)^{(\zeta_m-s_m-1)n}\alpha^{(1+s_m)n}\Delta_0 \mid s_{k-1}=p\Big)\Big\}\\
&\leq P\Big\{\bigcap_{m=1}^{p}\Big(|\sum_{i=0}^{\zeta_m n-1}a^{-i-1}d_i| \geq 2^{R'-1}a^{-\zeta_m n}(|a|+\delta)^{(\zeta_m-s_m-1)n}\alpha^{(1+s_m)n}\Delta_0\\
&\qquad\qquad -|x_0|-|\sum_{i=0}^{m-1}a^{(-\zeta_i-1)n}u_{(\zeta_i+1)n-1}|\Big|s_{k-1}=p\Big)\Big\}\\
&\leq \min_{0\leq m\leq s_{k-1}}\Big\{P\Big(|\sum_{i=0}^{\zeta_m n-1}a^{-i-1}d_i| \geq 2^{R'-1}(\frac{|a|+\delta}{|a|})^{(\zeta_m-s_m-1)n}(\frac{\alpha}{|a|})^{(1+s_m)n}\Delta_0-|x_0|-(2^{R'-1}-1/2)\Delta_0\\
&\qquad\qquad -\sum_{i=1}^{m-1}|a|^n(\frac{|a|+\delta}{|a|})^{(\zeta_i-s_i-1)n}(\frac{\alpha}{|a|})^{(s_m+1)n}(2^{R'-1}-1/2)\Delta_0\Big|s_{k-1}=p\Big)\Big\} \qquad (21)\\
&\leq \min_{0\leq m\leq s_{k-1}}\Big\{P\Big(|\bar{d}| \geq 2^{R'-1}(\frac{|a|+\delta}{|a|})^{(\zeta_m-s_m-1)n}(\frac{\alpha}{|a|})^{(1+s_m)n}\Delta_0-|x_0|-(2^{R'-1}-1/2)\Delta_0\\
&\qquad\qquad -\sum_{i=1}^{m-1}(\frac{|a|+\delta}{|a|})^{(\zeta_m-s_m-1)n}(\frac{\alpha}{|a|})^{(1+s_m)n}(2^{R'-1}-1/2)\Delta_0\Big|s_{k-1}=p\Big)\Big\} \qquad (22)
\end{aligned}$$

and for this m

$$\left(1-\sum_{i=1}^{m-1}(\frac{|a|+\delta}{|a|})^{n(\zeta_i-\zeta_m-(\log_{\frac{(|a|+\delta)}{|a|}}(\frac{|a|+\delta}{\alpha}))(s_i-s_m))}\right) \geq (1-\sum_{i=1}^{m-1}(\frac{|a|+\delta}{|a|})^{-\alpha_A in})>0. \quad (26)$$

This follows from considering a conservative configuration among an increasing subsequence of times $\{\zeta_i,\ldots,\zeta_m\}$, such that for all elements of this sequence

$$(\frac{|a|+\delta}{|a|})^{\zeta_i n}(\frac{\alpha}{|a+\delta|})^{s_i n} \geq (\frac{|a|+\delta}{|a|})^{(\zeta_{k-1}-s_{k-1})n}(\frac{\alpha}{|a|+\delta})^{(s_{k-1})n}$$

and for every element up until time m, $(\zeta_m-\zeta_i-(s_m-s_i)(\log_{\frac{(|a|+\delta)}{|a|}}(\frac{|a|+\delta}{\alpha}))\geq\alpha_A(s_m-s_i)$. Such an ordered sequence provides a conservative configuration which yet satisfies (26), by considering if needed, m to be an element in the sequence with a lower index value for which this is satisfied. This has to hold at least for one time ζ_m, since $k-1$ satisfies (24) . Such a construction ensures that (25) is uniformly bounded from below for every k since $\sum_{i=1}^{\infty}(\frac{|a|+\delta}{|a|})^{-\alpha_A in}<1$ for n large enough.

Hence, by (24), for some constant $B_b>0$, the following holds:

$$P\left(|\bar{d}|\geq B_b\Delta_0\left((\frac{|a|+\delta}{a})^{\zeta_m n}(\frac{\alpha}{(|a|+\delta)})^{(s_m+1)n}\right)\right) \leq 2\frac{\sigma'}{B_b\Delta_0\left((\frac{|a|+\delta}{a})^{\zeta_m n}(\frac{\alpha}{|a|+\delta})^{(s_m+1)}\right)} \times e^{-\left(B_b\Delta_0(\frac{|a|+\delta}{a})^{\zeta_m n}(\frac{\alpha}{(|a|+\delta)})^{(s_m+1)n}\right)^2/2\sigma'^2}. \quad (27)$$

The above follows from bounding the complementary error function by the following: $\int_q^\infty\mu(dx)\leq\int_q^\infty\frac{x}{q}\mu(dx)$, for $q>0$ when μ is a zero-mean Gaussian measure. In the aforementioned derivation $\sigma'^2=E[d_1^2]|a|^{-2}/(1-|a|^{-2})$. The left-hand side of (27) can be further upper bounded by, for any $r>0$

$$M_r(\Delta_0)r^{-\left((\frac{|a|+\delta}{a})^{\zeta_m n}(\frac{\alpha}{|a|+\delta})^{(s_m+1)n}\right)} +\left(1-1_{\{\zeta_m-(s_m+1)(\log_{\frac{(|a|+\delta)}{|a|}}(\frac{|a|+\delta}{\alpha}))>\alpha_A(s_m+1)\}}\right) \quad (28)$$

with $M_r(\Delta_0)\to 0$ as $\Delta_0\to\infty$ exponentially

$$\frac{M_r(\Delta_0)}{\Delta_0^{-p}}\to 0 \quad (29)$$

for any $p\in\mathbb{N}_+$, due to the exponential dependence of (27) in Δ_0. Thus, combined with (24), conditioned on having s_{k-1} errors and a Type I-A error at time 0, we have the following bound on (22):

$$M_r(\Delta_0)r^{-\left((\frac{|a|+\delta}{a})^{(k-1-s_{k-1})n}(\frac{\alpha}{|a|})^{(s_{k-1}+1)n}\right)} + 1_{\{\zeta_{k-1}\leq\frac{(s_{k-1}+1)}{\kappa}\}} \quad (30)$$

with

$$\kappa=\frac{1}{\log_{\frac{(|a|+\delta)}{|a|}}(\frac{|a|+\delta}{\alpha})+\alpha_A}. \quad (31)$$

We observe that the number of errors needs to satisfy the following relation for the aforementioned bound in (28) to be less than 1:

$$k-1>(1+s_{k-1})/\kappa.$$

Finally, the probability that the number of incorrect transmissions exceeds $\kappa(k-1)-1$ is exponentially low, as we observe in the following. Let, as before, $P_e(n)=P^e_{g|\mathcal{Z}}(n)$. We consider now Chernoff–Sanov's theorem: the sum of Bernoulli error events leads to a binomial distribution. Let for $1>\zeta>0$,

$$\begin{aligned}
&P\Bigg(|\bar{d}|\geq 2^{R'-1}(\frac{|a|+\delta}{|a|})^{(\zeta_m-s_m-1)n}(\frac{\alpha}{|a|})^{(1+s_m)n}-|x_0|-(2^{R'-1}-1/2)\Delta_0 \\
&\qquad -\sum_{i=1}^{m-1}(\frac{|a|+\delta}{|a|})^{(\zeta_m-s_m-1)n}(\frac{\alpha}{|a|})^{(1+s_m)n}(2^{R'-1}-1/2)\Delta_0\Bigg|s_{k-1}=p\Bigg) \\
&=P\Bigg(|\bar{d}|\geq 2^{R'-1}(\frac{|a|+\delta}{|a|})^{(\zeta_m-s_m)n}(\frac{\alpha}{|a|})^{(s_m)n}(\frac{\alpha}{|a|+\delta})^n\left(1-\sum_{i=1}^{m-1}(\frac{|a|+\delta}{|a|})^{(\zeta_i-\zeta_m)n}(\frac{\alpha}{|a|})^{(s_i-s_m)n}(1-2^{-R'})\right)\Delta_0 \\
&\qquad -|x_0|-(2^{R'-1}-1/2)\Delta_0\Bigg|s_{k-1}=p\Bigg) \\
&\leq P\Bigg(|\bar{d}|\geq 2^{R'-1}(\frac{|a|+\delta}{|a|})^{(\zeta_m-s_m)n}(\frac{\alpha}{|a|})^{(s_m)n}(\frac{\alpha}{|a|+\delta})^n\left(1-\sum_{i=1}^{m-1}(\frac{|a|+\delta}{|a|})^{(\zeta_i-\zeta_m)n}(\frac{\alpha}{|a|})^{(s_i-s_m)n}(1-2^{-R'})\right)\Delta_0 \\
&\qquad -2^{R'}\Delta_0\Bigg|s_{k-1}=p\Bigg)
\end{aligned} \quad (23)$$

$D(\zeta, P_e) = \zeta \log(\zeta/P_e) + (1-\zeta)\log(\frac{1-\zeta}{1-P_e})$. Then, the following upper bound holds [19], for every $k > 3$:

$$\begin{aligned} &P\Big(\sum_{t=1}^{k-2} 1_{\{\text{Type II Error}\}} \geq \kappa(k-1) - 1\Big) \\ &= P\Big(\sum_{t=1}^{k-2} 1_{\{\text{Type II Error}\}} \geq (k-2)(\kappa - \frac{1-\kappa}{k-2})\Big) \\ &\leq e^{-(k-2)D((\kappa - \frac{1-\kappa}{k-2}), P_e)}. \end{aligned} \tag{32}$$

Hence

$$P\Big(\sum_{t=1}^{k-2} 1_{\{\text{Type II Error}\}} \geq (\kappa - \frac{1-\kappa}{k-2})(k-2)\Big) \leq (e^{H((\kappa - \frac{1-\kappa}{k-2}))} P_e^{(\kappa - \frac{1-\kappa}{k-2})})^{k-2} \tag{33}$$

with $H(z) = -z\log(z) - (1-z)\log(1-z) \leq 1$. Hence

$$P\Big(\sum_{t=1}^{k-2} 1_{\{\text{Type II Error}\}} \geq (\kappa - \frac{1-\kappa}{k-2})(k-2)\Big) \leq (eP_e^{(\kappa - \frac{1-\kappa}{k-2})})^{k-2}. \tag{34}$$

We could bound the following summation as follows:

$$\sum_{s_{k-1}=0}^{\lfloor \kappa(k-1)\rfloor - 1} \binom{k-2}{s_{k-1}} M_r(\Delta_0) r^{-\Big((k-1)-(s_{k-1}+1)/\kappa\Big)n} \times (P_e)^{s_{k-1}}(1-P_e)^{k-1-s_{k-1}} \tag{35}$$

$$\leq M_r(\Delta_0)(1-P_e)^{k-1}\Big(\sum_{s_{k-1}=0}^{\lfloor(\kappa - \frac{1-\kappa}{k-2})(k-2)\rfloor} \binom{k-2}{s_{k-1}}\Big) \times (\frac{P_e}{1-P_e})^{\kappa(k-1)-1} \tag{36}$$

$$\begin{aligned} &\leq M_r(\Delta_0) 2^{(k-2)} (P_e)^{(\kappa - \frac{1-\kappa}{k-2})(k-2)} \\ &= M_r(\Delta_0)(2P_e^{(\kappa - \frac{1-\kappa}{k-2})})^{(k-2)} \end{aligned} \tag{37}$$

where (35) – (36) hold since r can be taken to be $r > (\frac{1-P_e}{P_e})^\kappa$ by taking Δ_0 to be large enough and in the summations s_{k-1} taken to be $\kappa(k-1) - 1$. We also use the inequality

$$\sum_{s=0}^{\lfloor \kappa(k-1)\rfloor - 1} \binom{k-2}{s} \leq 2^{k-2} \tag{38}$$

and that $\kappa(k-1) - 1 \leq k-2$.

Thus, from (20) we have computed a bound on the stopping time distributions through (34) and (37) . Following similar steps for the Type I-B error and no error cases at time 0, we obtain the bounds on the stopping time distributions as follows.

1) Conditioned an error in the granular region (Type I-A) at time 0, the condition for the number of errors is that

$$k - 1 > (1 + s_{k-1})\frac{1}{\kappa}$$

and by adding (33) and (37), the stopping time is dominated by

$$\begin{aligned} &P(\tau_1 \geq kn) \\ &\leq M_r(\Delta_0)(2P_e^{(\kappa - \frac{1-\kappa}{k-2})})^{(k-2)} + (eP_e^{(\kappa - \frac{1-\kappa}{k-2})})^{k-2} \\ &\leq \Xi(\Delta_0)(eP_e^{(\kappa - \frac{1-\kappa}{k-2})})^{k-2} \end{aligned} \tag{39}$$

for $\Xi(\Delta) = M_r(\Delta) + 1$ which goes to 1 as $\Delta \to \infty$.

2) Conditioned on the error that $\mathcal{Z}$ is the decoding output at time 0, in the above, the condition for the number of errors is that

$$k - 1 > s_{k-1}\frac{1}{\kappa}$$

and we may replace the exponent term $(\kappa - \frac{1-\kappa}{k-2})$ with $(\kappa + \frac{\kappa}{k-2})$ and the stopping time is dominated by

$$P(\tau_1 \geq kn) \leq \Xi(\Delta_0)(eP_e^{(\kappa + \frac{\kappa}{k-2})})^{(k-2)} \tag{40}$$

for $\Xi(\Delta) = M_r(\Delta) + 1$ which goes to 1 as $\Delta \to \infty$.

3) Conditioned on no error at time 0 and the rate condition $R' > \log_2(|a|/\alpha)$, the condition for the number of errors is that

$$k - 1 > 1 + \frac{s_{k-1}}{\kappa}$$

and we may replace the exponent term $(\kappa - \frac{1-\kappa}{k-2})$ with κ. The reason for this is that $|x_0 - \hat{x}_0| \leq \Delta_0/2$ and the control term applied at time n reduces the error.

As a result, (22) writes as (41), shown at the bottom of the next page, in this case. Since $2^{R'-1}(\frac{\alpha}{|a|})^n > 1$, the effect of the additional 1 in the exponent for $(\frac{\alpha}{|a|})^{s_m+1}$ can be excluded, unlike the case with $P^g_{e|e}$ above in (23) .

As a result, the stopping time is dominated by

$$P(\tau_1 \geq kn) \leq \Xi(\Delta_0)(eP_e^{\kappa})^{k-2} \tag{42}$$

for $\Xi(\Delta) = M_r(\Delta) + 1$ which goes to 1 as $\Delta \to \infty$.

This completes the proof of the lemma. □

C. Proof of Theorem 2.2

Once we have the Markov chain by Lemma 5.1, and the bound on the distribution of the sequence of stopping times defined in (15) by Lemma 5.2, we will invoke Theorem 7.2 or Theorem 7.3 with Lyapunov functions $V(x, \Delta) = \log_2(\Delta^2)$, $f(x, \Delta)$ taken as a constant and C a compact set.

As mentioned in Remark 2.2, for a DMC with block length n Shannon's random coding method satisfies

$$\begin{aligned} P_e(n) := &\max_{c \in \{1,2,\ldots,M(n)\}} P(c' \neq c | c \quad \text{is transmitted}) \\ &\leq e^{-nE(R)+o(n)} \end{aligned}$$

with c' being the decoder output. Here, $\frac{o(n)}{n} \to 0$ as $n \to \infty$ and $E(R) > 0$ for $0 < R < C$. Thus, by Lemma 5.2, we observe that

$$E[\tau_1 | x_0, \Delta_0] = \sum_{k=1}^{\infty} P(\tau_1 \geq k) \leq K'_{\Delta_0}(n) < \infty \tag{43}$$

for some finite number $K'_{\Delta_0}(n)$. The finiteness of this expression follows from the observation that for $k-2 > \frac{1-\kappa}{\kappa}$, the exponent in $e^{-n(\kappa - \frac{1-\kappa}{k-2})(E(R) - \frac{o(n)}{n})}$ becomes negative. Furthermore, $K'_{\Delta_0}(n)$ is monotone decreasing in Δ_0 since $M_r(\Delta)$ is decreasing in Δ.

We now apply the random-time drift result in Theorem 7.2 and Corollary 7.1. First, observe that the probability that $\tau_{z+1} \neq \tau_z + n$ is upper bounded by the probability

$$\begin{aligned} &P^e_{g|g} \\ &+ (1 - P^e_{g|g} - P^e_{\mathcal{Z}|g}) 2P\left(\bar{d} \geq (2^{R'}(\frac{\alpha}{|a|})^n - 1)\Delta_0/2\right) \\ &\quad + 2P^e_{\mathcal{Z}|g} P\left(\bar{d} > (2^{R'-1}(|a|+\delta)^n)\Delta_0 - |a^n x_0|\right) \\ &\leq P^e_{g|g} \\ &+ (1 - P^e_{g|g} - P^e_{\mathcal{Z}|g}) 2P\left(\bar{d} \geq (2^{R'}(\frac{\alpha}{|a|})^n - 1)\Delta_0/2\right) \\ &\quad + P^e_{\mathcal{Z}|g} 2P\left(\bar{d} > 2^{R'-1}((|a|+\delta)^n - |a|^n)\Delta_0\right) \\ &=: \Upsilon(\Delta_{\tau_0}). \end{aligned} \tag{44}$$

Observe that, provided that $R'(n) > n\log_2(|a|/\alpha)$

$$\lim_{\Delta_0 \to \infty} \Upsilon(\Delta_{\tau_0}) = P^e_{g|g}.$$

We now pick the Lyapunov function $V(x,\Delta) = \log_2(\Delta^2)$ and $f(x,\Delta)$ a constant to obtain (46), shown at the bottom of the next page, where $\chi > 0$ is an arbitrarily small positive number. In (45), shown at the bottom of the next page, we use the fact that zooming out for all time stages after $\tau_z + n$ provides a worst case sequence and that by Hölder's inequality for a random variable X and an event $\mathbb{A}$ the following holds:

$$\begin{aligned} &E[X 1_{\mathbb{A}}] \\ &\leq (E[|X|^{1+\chi}])^{\frac{1}{1+\chi}} (E[1_{\mathbb{A}}^{\frac{1+\chi}{\chi}}])^{\frac{\chi}{1+\chi}} \\ &= (E[|X|^{1+\chi}])^{\frac{1}{1+\chi}} (P(\mathbb{A}))^{\frac{\chi}{1+\chi}}. \end{aligned} \tag{47}$$

Now, the last term in (45) will converge to zero with n large enough and $\Delta_{\tau_z} \to \infty$ for some $\chi > 0$, since by Lemma 5.2 $P(\tau_{z+1} = \tau_z + kn)$ is bounded by a geometric measure and the expectation of $((\tau_{z+1} - \tau_z - 1)\log_2(|a| + \delta))^{1+\chi}$ is finite and monotone decreasing in Δ_0. The second term in (46) is negative with $P^e_{\mathcal{Z}|g}$ sufficiently small.

These imply that, for some sufficiently large F, the equation

$$E[\log(\Delta^2_{\tau_{z+1}}) | \Delta_{\tau_z}, h_{\tau_z}] \leq \log(\Delta^2_{\tau_z}) - b_0 + b_1 1_{\{|\Delta_{\tau_z}| \leq F\}} \tag{48}$$

holds for some positive b_0 and finite b_1. Here, b_1 is finite since $K'(n)$ is finite. With the uniform boundedness of (43) over the sequence of stopping times, this implies by Theorem 7.3 that $\{(x,\Delta) : |\Delta| \leq F, |\frac{x}{2^{R'-1}\Delta}| \leq 1\}$ is a recurrent set. □

D. Proof of Theorem 2.3

The process (x_{tn}, Δ_{tn}) is a Markov chain, as was observed in Lemma 5.1. In this section, we establish irreducibility of this chain and the existence of a small set (see Section B) to be able to invoke Theorem 7.2, in view of (48) . The following generalizes the analysis in [97] and [94].

Let the values taken by

$$\log_2(\bar{Q}(\Delta_{tn}, c'_{(t+1)n-1}))/s$$

$$\begin{aligned} &P\left(\bigcap_{m=1}^{p} (|a^{\zeta_m n}(x_0 + \sum_{i=0}^{\zeta_m n - 1} a^{-i-1} d_i + \sum_{i=0}^{m-1} a^{(-\zeta_i - 1)n} u_{(\zeta_i+1)n-1})| \right. \\ &\qquad\qquad \left. \geq 2^{R'-1}(|a|+\delta)^{(\zeta_m - s_m - 1)n} \alpha^{(1+s_m)n} | s_{k-1} = p\right) \\ &\leq \min_{0 \leq m \leq s_{k-1}} \left\{ P\left(|\sum_{i=0}^{\zeta_m n - 1} a^{-i-1} d_i| \geq 2^{R'-1}(\frac{|a|+\delta}{|a|})^{(\zeta_m - s_m - 1)n} (\frac{\alpha}{|a|})^{(1+s_m)n} \right.\right. \\ &\qquad \left.\left. - |x_0 - \hat{x}_0| - \sum_{i=1}^{m-1} |a|^n (\frac{|a|+\delta}{|a|})^{\zeta_i - s_i - 1} (\frac{\alpha}{|a|})^{(s_i+1)n} (2^{R'-1} - 1/2)\Delta_0 \right) \right\} \\ &\leq \min_{0 \leq m \leq s_{k-1}} \left\{ P\left(|\bar{d}| \geq (2^{R'-1}(\frac{|a|+\delta}{|a|})^{(\zeta_m - s_m - 1)n} (\frac{\alpha}{|a|})^{(1+s_m)n} - 1/2)\Delta_0 \right.\right. \\ &\qquad \left.\left. - \sum_{i=1}^{m-1} (\frac{|a|+\delta}{|a|})^{(\zeta_i - s_i - 1)n} (\frac{\alpha}{|a|})^{(1+s_i)n} (2^{R'-1} - 1/2)\Delta_0 \right) \right\} \\ &\leq \min_{0 \leq m \leq s_{k-1}} \left\{ P\left\{ |\bar{d}| \geq \left(2^{R'-1}(\frac{\alpha}{|a|})^n \left((\frac{|a|+\delta}{|a|})^{(\zeta_m - s_m - 1)n} (\frac{\alpha}{|a|})^{s_m n} - 2^{-R'} (\frac{\alpha}{|a|})^{-n} \right) \Delta_0 \right.\right.\right. \\ &\qquad \left.\left.\left. - \sum_{i=1}^{m-1} (2^{R'-1}(\frac{\alpha}{|a|})^n (\frac{|a|+\delta}{|a|})^{(\zeta_i - s_i - 1)n} (\frac{\alpha}{|a|})^{s_i n} (1 - 2^{-R'})\Delta_0) \right) \right\} \right\} \end{aligned} \tag{41}$$

be $\{-\tilde{A}, 0, \tilde{B}\}$. Here. $\tilde{A}, B$ are relatively prime. Let $\mathbb{L}_{z_0,\tilde{A},\tilde{B}}$ be defined as

$$\{n \in \mathbb{N}, n \geq \log_2(L')/s : \exists N_A, N_B, n = -N_A\tilde{A} + N_B\tilde{B} + z_0\}$$

where $z_0 = \log_2(\Delta_0)/s$ is the initial condition of the parameter for the quantizer. We note that since $\tilde{A}, \tilde{B}$ are relatively prime, by Bézout's lemma (see [1]) the communication class will include the bin sizes whose logarithms are integer multiples of a constant except those leading to $\Delta < L'$: since we have $\Delta_{(t+1)n} = \bar{Q}(\Delta_{tn}, c'_{(t+1)n-1})\Delta_{tn}$, it follows that

$$\log_2(\Delta_{(t+1)n})/s = \log_2(\bar{Q}(\Delta_{tn}, c'_{(t+1)n-1}))/s + \log_2(\Delta_{tn})/s$$

is also an integer. Furthermore, since the source process $\{x_{tn}\}$ is "Lebesgue irreducible" (the system noise admits a probability density function that is positive everywhere), and there is a uniform lower bound L' on bin sizes, the error process takes values in any of the admissible quantizer bins with nonzero probability. Consider two integers $k, l \geq \frac{\log_2(L')}{s}$. For all $l, k \in \mathbb{L}_{z_0,\tilde{A},\tilde{B}}$, there exist $N_A, N_B \in \mathbb{N}$ such that $l - k = -N_A\tilde{A} + N_B\tilde{B}$. We can show that the probability of N_A occurrences of perfect zoom, and N_B occurrences of under-zoom phases is bounded away from zero. This set of occurrences includes the event that in the first N_A time stages perfect-zoom occurs and later, successively, N_B times under-zoom phase occurs. Considering worst possible control realizations and errors, the probability of this event is lower bounded by

$$\left(P\left(\tilde{d} \in [-2^{R'(n)-1}L' - |a|^n L', 2^{R'(n)-1}L' - |a|^n L']\right) \times (P^e(\mathcal{Z}|i))\right)^{N_B} \times \left(P\left(\tilde{d} \in [-(\alpha^n 2^{R'} - a^n)L'/2, (\alpha^n 2^{R'} - a^n)L'/2]\right) \times (1 - P_e)\right)^{N_A} > 0 \tag{49}$$

where $\tilde{d} = \sum_{i=0}^{n-1} a^{n-i-1} w_i$ is a Gaussian random variable. The above follows from considering the sequence of zoom-ins and zoom-outs and the behavior of $a^n(x_{tn} - \hat{x}_{tn}) + \tilde{d}$. In the aforementioned discussion, $P^e(\mathcal{Z}|i)$ is the conditional error on the zoom symbol given the transmission of granular bin i, with the lowest error probability. (If the lowest such an error probability is zero, an alternative sequence of events can be provided through events concerning the noise variables leading to zooming.) Thus, for any two such integers k, l and for some $r > 0$, $P(\log_2(\Delta_{(t+r)n}) = ls \mid \log_2(\Delta_{tn}) = ks) > 0$.

We can now construct a small set and make the connection with Theorems 2.2 and 7.2. Define

$$C_x \times C'_\Delta = \{(x, \Delta) : L' \leq \Delta \leq F, |h| \leq 1, \frac{\log_2(\Delta)}{s} \in \mathbb{Z}\}.$$

We will show that the recurrent set $C_x \times C'_\Delta$ is small.

$$\begin{aligned}
&E[\log(\Delta^2_{\tau_{z+1}})|x_{\tau_z}, \Delta_{\tau_z}] \\
&= E[\log(\Delta^2_{\tau_{z+1}})(1_{\{\text{Type I-A error at } \tau_z\}} + 1_{\{\text{Type I-B error at } \tau_z\}} + 1_{\{\text{no error at } \tau_z\}})|x_{\tau_z}, \Delta_{\tau_z}] \\
&\leq (1 - P^e_{\mathcal{Z}|g} - P^e_{g|g})\left(n\log_2(\alpha) + nE[\log_2((|a+\delta)^{2(\tau_{z+1}-1)})1_{\{\tau_{z+1} > \tau_z + n\}})|\text{no error}]\right) \\
&\quad + P^e_{\mathcal{Z}|g}\left(n\log_2(|a|+\delta) + nE[\log_2((|a+\delta)^{2(\tau_{z+1}-1)})1_{\{\tau_{z+1} > \tau_z + n\}})|\text{Type I-B error}]\right) \\
&\quad + P^e_{g|g}\Big(n\log_2(\alpha) + nE[\log_2((|a+\delta)^{2(\tau_{z+1}-1)})1_{\{\tau_{z+1} > \tau_z + n\}})|\text{Type I-A error}] \\
&\quad + \log_2(\Delta^2_{\tau_z}) \\
&= \log(\Delta^2_{\tau_z}) + n\left((1 - P^e_{\mathcal{Z}|g})\log_2(\alpha) + P^e_{\mathcal{Z}|g}\log_2(|a|+\delta)\right) \\
&\quad + nE\left[\log_2\left((|a+\delta)^{2(\tau_{z+1}-1)}\right)1_{\{\tau_{z+1} > \tau_z + n\}})\right] \\
&\leq \log(\Delta^2_{\tau_z}) + n\left((1 - P^e_{\mathcal{Z}|g})\log_2(\alpha) + P^e_{\mathcal{Z}|g}\log_2(|a|+\delta)\right) \\
&\quad + \left(P(\tau_{z+1} > \tau_z + n)\right)^{\frac{\chi}{1+\chi}} n\left(\sum_{k=2}^{\infty} P(\tau_{z+1} = \tau_z + kn)((k-1)\log_2(|a|+\delta))^{1+\chi}\right)^{\frac{1}{1+\chi}} &(45) \\
&\leq \log(\Delta^2_{\tau_z}) + n\left((1 - P^e_{\mathcal{Z}|g})\log_2(\alpha) + P^e_{\mathcal{Z}|g}\log_2(|a|+\delta)\right) \\
&\quad + (\Upsilon(\Delta_{\tau_z}))^{\frac{\chi}{1+\chi}} n\left(\sum_{k=2}^{\infty} P(\tau_{z+1} = \tau_z + kn)((k-1)\log_2(|a|+\delta))^{1+\chi}\right)^{\frac{1}{1+\chi}} &(46)
\end{aligned}$$

Toward this end, we first establish irreducibility. For some distribution $\mathcal{K}$ on positive integers, $E \subset \mathbb{R}$, and Δ an admissible bin size

$$\sum_{n\in\mathbb{N}_+} \mathcal{K}(n)P\left((x_n,\Delta_n)\in(E\times\{\Delta\})\,\Big|\,x_0,\Delta_0\right) \geq K_{\Delta_0,\Delta}\psi(E,\Delta).$$

Here, $K_{\Delta_0,\Delta}$, denoting a lower bound on the probability of visiting Δ from Δ_0 in some finite time, is nonzero by (49) and ψ is a positive map as the following argument shows. Let $t>0$ be a time stage for which $\Delta_{tn}=\Delta$ and thus, with $|h_{(t-1)n}|\leq 1$: $|ax_{(t-1)n}+bu_{(t-1)n}|\leq |a|^n\Delta_{(t-1)n}/2=(|a|/\alpha)^n\frac{\Delta}{2}$. Thus, it follows that for $A_1,B_1\in\mathbb{R}$, $A_1<B_1$

$$\begin{aligned}
&P\left(x_{tn}\in[A_1,B_1]\,\Big|\,|a^n x_{(t-1)n}+bu_{(t-1)n}| \leq |a|\Delta_{(t-1)n}/2\right)\\
&=P\left(\tilde{d}_{t-1}\in[A_1-(a^n x_{(t-1)n}+bu_{(t-1)n}),B_1-(a^n x_{(t-1)n}+bu_{(t-1)n})]\,\Big|\,|a^n x_{(t-1)n}+bu_{(t-1)n}|\leq |a|^n\Delta_{(t-1)n}/2\right)\\
&\geq \min_{|z|\leq\frac{\Delta}{2}(|a|/\alpha)^n}\left(P(\tilde{d}_{t-1}\in[A_1-z,B_1-z]\right). \qquad (50)
\end{aligned}$$

Thus, in view of (50), ψ satisfies for $\alpha<\beta,\alpha,\beta\in\mathbb{R}$

$$\psi([A_1,B_1],\Delta) \geq \min_{|z|\leq\frac{\Delta}{2}(|a|/\alpha)^n}\left(P(\tilde{d}_{t-1}\in[A_1-z,B_1-z]\right)>0.$$

The chain satisfies the recurrence property that

$$P_{(x,\Delta)}(\tau_{C_x\times C'_\Delta}<\infty)=1$$

for any admissible (x,Δ). This follows from the construction of

$$\Theta_k(\Delta,x):=P(\mathcal{T}_1\geq kn\mid x,\Delta)$$

where

$$\mathcal{T}_1=\inf(kn>0:|x_k|\leq 2^{R'-1}\Delta_k,x_0=x,\Delta_0=\Delta)$$

and observing that $\Theta_k(\Delta,x)$ is majorized by a geometric measure with similar steps as in Section V-B. Once a state which is perfectly zoomed, that is which satisfies $|x_t|\leq 2^{R'-1}\Delta_t$, is visited, the stopping time analysis can be used to verify that from any initial condition the recurrent set is visited in finite time with probability 1.

We will now establish that the set $C_x\times C'_\Delta$ is small. By [62, Th. 5.5.7], under aperiodicity and irreducibility, every petite set is small. To establish the petite set property, we will follow an approach taken by Tweedie [88] which considers the following test, which only depends on the one-stage transition kernel of a Markov chain: if a set S is such that the following *uniform countable additivity* condition

$$\lim_{k\to\infty}\sup_{x\in S}P(x,B_k)=0$$

is satisfied for every sequence $B_k\downarrow\emptyset$, and if the Markov chain is irreducible, then S is petite (see [88, Lemma 4] and [62, Proposition 5.5.5(iii)]).

Now, the set $C_x\times C'_\Delta$ satisfies the uniform countable additivity condition since for any given bin size Δ' in the countable space constructed above, (51) shown at the bottom of the page holds. This follows from the fact that the Gaussian random variable $\tilde{d}$ satisfies

$$\lim_{k\to\infty}\sup_{\tilde{d}\in C_0}P(\tilde{d}\in A_k)=0$$

uniformly over a compact set C_0, for any sequence $A_k\downarrow\emptyset$, since a Gaussian measure admits a uniformly bounded density function. Hence, $C_x\times C'_\Delta$ is petite.

Finally, aperiodicity of the sampled chain follows from the fact that the smallest admissible state for the quantizer L' can be visited in subsequent time sampled time stages since

$$P(\tilde{d}\in[-2^{R'-1}L'/|a|^n-L',-2^{R'-1}L'/|a|^n+L'])>0.$$

Thus, the sampled chain is positive Harris recurrent. ☐

E. *Proof of Theorem 2.4*

By Kolmogorov's extension theorem, it suffices to check that the property holds for finite-dimensional cylinder sets, since these sets generate the σ-algebra on which the stochastic process measure is defined. Suppose first that the sampled Markov chain is stationary. Consider two elements

$$\begin{aligned}
&P(x_{t+1+n}\in A_1,x_{t+2+n}\in A_2)\\
&=\int_{x_{\lfloor\frac{t+1+n}{n}\rfloor n}}P(dx_{\lfloor\frac{t+1+n}{n}\rfloor n},x_{t+1+n}\in A_1,x_{t+2+n}\in A_2)\\
&=\int_{x_{\lfloor\frac{t+1+n}{n}\rfloor n}}P(x_{t+1+n}\in A_1,x_{t+2+n}\in A_2|x_{\lfloor\frac{t+1+n}{n}\rfloor n})\times P(dx_{\lfloor\frac{t+1+n}{n}\rfloor n})\\
&=\int_{x_{\lfloor\frac{t+1}{n}\rfloor n}}P(x_{t+1}\in A_1,x_{t+2}\in A_2|x_{\lfloor\frac{t+1}{n}\rfloor n})P(dx_{\lfloor\frac{t+1}{n}\rfloor n}).
\end{aligned}$$

The above follows from the fact that the marginals $P(dx_{\lfloor\frac{t+1}{n}\rfloor n})$ and $P(dx_{\lfloor\frac{t+1+n}{n}\rfloor n})$ are equal since the sampled Markov chain

$$\begin{aligned}
&\lim_{k\to\infty}\sup_{(x,\Delta)\in C_x\times C'_\Delta}P((x_{(t+1)n},\Delta_{(t+1)n})\in(B_k\times\Delta')|x_{tn}=x,\Delta_{tn}=\Delta)\\
&=\lim_{k\to\infty}\sup_{(x,\Delta)\in C_x\times C'_\Delta}P\left((\tilde{d},\Delta_{(t+1)n})\in\left((B_k-(a^n x_{tn}+bu_{(t+1)n-1}))\times\Delta'\right)\Big|x_{tn}=x,\Delta_{tn}=\Delta\right)=0 \qquad (51)
\end{aligned}$$

is positive Harris recurrent and assumed to be stationary, and the dynamics for interblock times are time homogeneous Markov. The above is applicable for any finite-dimensional set, thus for any element in the sigma field generated by the finite-dimensional sets, on which the stochastic process is defined. Now, let for some event A, $T^{-n}A = A$, where T denotes the shift operation (see Section A). Then

$$P(A) = \lim_{k\to\infty} P(A \cap T^{-kn}A) = \lim_{k\to\infty} P(A)P(T^{-kn}A|A).$$

Note that a positive Harris recurrent Markov chain admits a unique invariant distribution and for every $x_0 \in \mathbb{R}$

$$\lim_{k\to\infty} P(x_{kn} \in A|x_0) = \pi(A)$$

where $\pi(\cdot)$ is the unique invariant probability measure. Since such a Markov chain forgets its initial condition, it follows that for $A = T^{-n}A$

$$P(A \cap T^{-kn}A) = P(A \cap A) = P(A)P(A).$$

Thus, $P(A) \in \{0,1\}$, and the process is n-ergodic. □

F. Proof of Theorem 2.5

We begin with the following result, which is a consequence of Theorem 7.2.

Lemma 5.3: Under the conditions of Theorem 2.3, we have that if for some $\gamma > 0$, $b < \infty$, the following holds:

$$\gamma E[\sum_{k=0}^{(\tau_1/n)-1} \Delta_{kn}^2 | x_0, \Delta_0] \leq \Delta_0^2 - E[\Delta_{\tau_1}^2 | x_0, \Delta_0] + b1_{\{(\Delta_0,h_0)\in(C_x'\times C_h)\}}$$

then $\lim_{k\to\infty} E[\Delta_{kn}^2] < \infty$. □

Now, under the hypotheses of Theorem 2.3, and observing that Type I-B and I-A errors are worse than the no error case at time 0 for the stopping time tail distributions, we obtain (52) for some finite ζ_1. In (52), we use the property that $H(\kappa) \leq 1$ and (39)–(42).

We now establish that

$$\lim_{\Delta_0\to\infty} \frac{E[\Delta_{\tau_1}^2 \mid x_0, \Delta_0]}{\Delta_0^2} < 1.$$

This is a crucial step in applying Theorem 7.2.

$$\begin{aligned}
&E[\sum_{t=0}^{(\tau_1/n)-1} \Delta_{tn}^2 \mid x_0, \Delta_0] \\
&\leq \Delta_0^2 P_{g|g}^e \Bigg(\sum_{l=2}^{\infty} P(\tau_1 = ln | \text{Type I-A error}) \sum_{k=1}^{(l-1)} (|a|+\delta)^{2(k-1)n}\alpha^{2n} \Bigg) \\
&\quad + \Delta_0^2 P_{\mathcal{Z}|g}^e \Bigg(\sum_{l=2}^{\infty} P(\tau_1 = ln | \text{Type I-B error}) \sum_{k=1}^{l-1} (|a|+\delta)^{2kn} \Bigg) \\
&\quad + \Delta_0^2 (1 - P_{g|g}^e - P_{\mathcal{Z}|g}^e) \Bigg(\sum_{l=2}^{\infty} P(\tau_1 = ln | \text{no error at time } 0) \sum_{k=1}^{l-1} (|a|+\delta)^{2(k-1)n}\alpha^{2n} \Bigg) \\
&\qquad + \Delta_0^2 P(\tau_1 = n) \\
&\leq \Delta_0^2 P_{g|g}^e \Bigg(\sum_{l=2}^{\infty} P(\tau_1 = ln | \text{Type I-A error}) \frac{(|a|+\delta)^{2(l-1)n}}{(|a|+\delta)^{2n}-1}\alpha^{2n} \Bigg) \\
&\quad + \Delta_0^2 P_{\mathcal{Z}|g}^e \Bigg(\sum_{l=2}^{\infty} P(\tau_1 = ln | \text{Type I-B error}) (|a|+\delta)^{2n} \frac{(|a|+\delta)^{2(l)n}}{(|a|+\delta)^{2n}-1} \Bigg) \\
&\quad + \Delta_0^2 (1 - P_{g|g}^e - P_{\mathcal{Z}|g}^e) \Bigg(\sum_{l=2}^{\infty} P(\tau_1 = ln | \text{no error at time } 0) \frac{(|a|+\delta)^{2(l-1)n}}{(|a|+\delta)^{2n}-1}\alpha^{2n} \Bigg) \\
&\qquad + \Delta_0^2 P(\tau_1 = n) \\
&\leq \Delta_0^2 P_{g|g}^e \frac{(|a|+\delta)^{(2/\kappa)n}}{(|a|+\delta)^{2n}-1} \Xi(\Delta_0) \Bigg(\sum_{l=2}^{\infty} (e^{(l-2)}) P_e^{(\kappa)(l-1-\frac{1}{\kappa})} (|a|+\delta)^{2n(l-1-\frac{1}{\kappa})}\alpha^{2n} \Bigg) \\
&\quad + \Delta_0^2 P_{\mathcal{Z}|g}^e \frac{(|a|+\delta)^{2n}}{1-(|a|+\delta)^{-2n}} \Xi(\Delta_0) \Bigg(\sum_{l=2}^{\infty} (eP_e^{\kappa})^{l-2} (|a|+\delta)^{2(l-2)n} \Bigg) \\
&\quad + \Delta_0^2 (1 - P_{g|g}^e - P_{\mathcal{Z}|g}^e)(|a|+\delta)^{2n} \Xi(\Delta_0) \Bigg(\sum_{l=2}^{\infty} (eP_e^{\kappa})^{l-2} \frac{(|a|+\delta)^{2(l-2)n}}{(|a|+\delta)^{2n}-1}\alpha^{2n} \Bigg) \\
&\qquad + \Delta_0^2 P(\tau_1 = n) \\
&\leq \zeta_1 \Delta_0^2
\end{aligned} \tag{52}$$

Following similar steps as in (52) shown at the bottom of the previous page, the following upper bound on $(E[\Delta_{\tau_1}^2 \mid x_0, \Delta_0]/\Delta_0^2)$ is obtained:

$$
\begin{aligned}
&(1 - P^e_{g|g} - P^e_{\mathcal{Z}|g}) \\
&\quad\times \left(\alpha^{2n} + \frac{1}{\Delta_0^2} E[\Delta_{\tau_1}^2 1_{\{\tau_1 > n\}} | \text{no error at time } 0]\right) \\
&+ (P^e_{g|g})\Bigg(\alpha^{2n}(1 + (|a|+\delta)^{2n} + \dots (|a|+\delta)^{2(\lfloor \frac{1}{\kappa} \rfloor)n}) \\
&\quad + \sum_{k=\lceil \frac{1}{\kappa} \rceil + 1}^{\infty} e^{k-2} P_e^{(\kappa(k-1-\frac{1}{\kappa}))} (|a|+\delta)^{2(k-1-\frac{1}{\kappa})n} (\alpha)^{2n} \\
&\qquad \times (|a|+\delta)^{(2/\kappa)n} \Xi(\Delta_0)\Bigg) \\
&+ P^e_{\mathcal{Z}|g}\left((|a|+\delta)^{2n} + \Xi(\Delta_0) \frac{P_e^{\kappa}(|a|+\delta)^{2n}}{1 - P_e^{\kappa}(|a|+\delta)^{2n}}\right).
\end{aligned} \tag{53}
$$

Note now that

$$\lim_{\Delta_0 \to 0} P(\tau_1 > n | \text{no error at time } 0, x_0, \Delta_0) = 0$$

uniformly in x_0 with $|h_0| \leq 1$ and given the rate condition $R'(n) > n \log_2(|a|/\alpha)$ by (44) . Therefore, the first term in (53) converges to 0 in the limit of large Δ_0, since $\lim_{n\to\infty} \kappa \frac{1}{n} \log(P_e) + 2\log_2(|a|+\delta) < 0$ and we have the following upper bound:

$$\left((|a|+\delta)^{2n} \sum_{k=2}^{\infty} (eP_e^{\kappa})^{k-2} (|a|+\delta)^{2n(k-2)} (\alpha)^{2n}\right) < \infty$$

for sufficiently large n.

For the second term in (53), the convergence of the first expression is ensured with $\lim_{n\to\infty} P^e_{g|g}(|a|+\delta)^{(2/\kappa)n}\alpha^{2n} \to 0$ and $P_e(|a|+\delta)^{(2/\kappa)n} \to 0$ as $n \to \infty$. By combining the second and the third terms, the desired result is obtained.

To show that $\lim_{m\to\infty} E[x_{mn}^2] < \infty$, we first show that for some $\kappa > 0$

$$\kappa E[\sum_{m=0}^{(\tau_1/n)-1} x_{mn}^2 \mid x_0, \Delta_0] \leq \Delta_0^2 2^{2(R'-1)}. \tag{54}$$

Now

$$
\begin{aligned}
&E[\sum_{m=0}^{(\tau_1/n)-1} x_{mn}^2 \mid x_0, \Delta_0] \\
&= E\Bigg[\sum_{t=0}^{(\tau_1/n)-1} a^{2tn}\left((x_0 + \sum_{i=0}^{tn-1} a^{-i-1} d_i) + (\sum_{i=0}^{tn-1} a^{-i-1} u_i)\right)^2 \\
&\qquad\qquad \Bigg| x_0, \Delta_0\Bigg] \\
&\leq 2E[\sum_{t=0}^{(\tau_1/n)-1} a^{2tn}(x_0 + \sum_{i=0}^{tn-1} a^{-i-1} d_i)^2 | x_0, \Delta_0] \\
&\quad + 2E[\sum_{t=0}^{(\tau_1/n)-1} a^{2tn}(\sum_{i=0}^{tn-1} a^{-i-1} u_i)^2 | x_0, \Delta_0]
\end{aligned} \tag{55}
$$

which follows from the observation that for X, Y random variables, $E[(X+Y)^2] \leq 2E[X^2] + 2E[Y^2]$.

Let us first consider the component: $(a^t(x_0 + \sum_{i=0}^{tn-1} a^{-i-1} d_i))^2$

$$
\begin{aligned}
&E[\sum_{t=0}^{(\tau_1/n)-1} (a^{tn}(x_0 + \sum_{i=0}^{tn-1} a^{-i-1} d_i))^2 | x_0, \Delta_0] \\
&= E[\sum_{t=0}^{\infty} 1_{\{t < \tau_1/n\}} (a^{tn}(x_0 + \sum_{i=0}^{tn-1} a^{-i-1} d_i))^2 | x_0, \Delta_0] \\
&\leq \sum_{t=0}^{\infty} \left(E[(1_{\{t<\tau_1/n\}})^{1+\chi} | x_0, h_0]\right)^{\frac{1}{1+\chi}} \\
&\qquad \times \left(E[(a^{tn}(x_0 + \sum_{i=0}^{tn-1} a^{-i-1} d_i))^{2(\frac{1+\chi}{\chi})} | x_0, \Delta_0]\right)^{\frac{\chi}{1+\chi}}
\end{aligned} \tag{56}
$$

for some $\chi > 0$, by Hölder's inequality.

Moreover, for some $B_2 < \infty$

$$
\begin{aligned}
&E[a^{2tn(\frac{1+\chi}{\chi})}(x_0 + \sum_{i=0}^{tn-1} a^{-i-1} d_i)^{2(\frac{1+\chi}{\chi})} | x_0, \Delta_0] \\
&= |a|^{2tn(\frac{1+\chi}{\chi})} E[(x_0 + \sum_{i=0}^{tn-1} a^{-i-1} d_i))^{2\frac{1+\chi}{\chi}} | x_0, \Delta_0] \\
&\leq |a|^{2tn(\frac{1+\chi}{\chi})} E[(x_0 + \sum_{i=0}^{\infty} a^{-i-1} d_i)^{2\frac{1+\chi}{\chi}} | x_0, \Delta_0] \\
&= |a|^{2tn(\frac{1+\chi}{\chi})} (2^{R'-1}\Delta_0)^{2\frac{1+\chi}{\chi}} \\
&\qquad \times E[(\frac{x_0 + \sum_{i=0}^{\infty} a^{-i-1} d_i}{2^{R'-1}\Delta_0})^{2\frac{1+\chi}{\chi}} | x_0, \Delta_0] \\
&= |a|^{2tn(\frac{1+\chi}{\chi})} (2^{R'-1}\Delta_0)^{2\frac{1+\chi}{\chi}} \\
&\qquad \times E[(h_0 + \frac{\sum_{i=0}^{\infty} a^{-i-1} d_i}{2^{R'-1}\Delta_0})^{2\frac{1+\chi}{\chi}} | x_0, \Delta_0] \\
&< B_2 (2^{R'-1}\Delta_0)^{2\frac{1+\chi}{\chi}} |a|^{2tn(\frac{1+\chi}{\chi})}
\end{aligned} \tag{57}
$$

where the last inequality follows since for every fixed $|h_0| \leq 1$, the random variable $h_0 + (\sum_{i=0}^{\infty} a^{-i-1} d_i)/(2^{R'-1}\Delta_0)$ has a Gaussian distribution with finite moments, uniform on $\Delta_0 \geq L'$. Thus

$$
\begin{aligned}
&E[\sum_{t=0}^{(\tau_1/n)-1} a^{2tn}(x_0 + \sum_{i=0}^{tn-1} a^{-i-1} d_i)^2 | x_0, \Delta_0] \\
&\leq \sum_{t=0}^{\infty} \left(E[(1_{\{t<\tau_1/n\}})^{1+\chi} | x_0, \Delta_0]\right)^{\frac{1}{1+\chi}} \\
&\qquad \times \left(B_2 (2^{R'(n)-1}\Delta_0)^{2\frac{1+\chi}{\chi}} |a|^{2tn(\frac{1+\chi}{\chi})}\right)^{\frac{\chi}{1+\chi}} \\
&= \sum_{t=0}^{\infty} \left(\Xi(\Delta_0)(eP_e^{(\kappa - \frac{1-\kappa}{t-1})})^{t-1}\right)^{\frac{1}{1+\chi}} \left(B_2^{\frac{\chi}{1+\chi}} \Delta_0^2 |a|^{2tn}\right) \\
&= \sum_{t=0}^{\infty} \left(\Xi(\Delta_0)(eP_e^{(\kappa - \frac{1-\kappa}{t-1})})^{t-1} |a|^{2tn(1+\chi)}\right)^{\frac{1}{1+\chi}} \\
&\qquad \times \left(B_2^{\frac{\chi}{1+\chi}} \Delta_0^2\right) \\
&< \zeta_{B_2} \Delta_0^2
\end{aligned} \tag{58}
$$

for some finite ζ_{B_2} (for a fixed finite n). In the aforementioned discussion, we use the fact that we can pick $\chi > 0$ such that $(P_e)^{\kappa}|a|^{2n(1+\chi)} < 1$. Such χ exists, by the hypothesis that $\lim_{n\to\infty} P_e^{\kappa}(|a|+\delta)^{2n} = 0$.

We now consider the second term in (55) . Since u_i is the quantizer output which is bounded in magnitude in proportion with Δ_i, the second term writes as

$$\begin{aligned} &E[\sum_{t=0}^{(\tau_1/n)-1} a^{2tn}(\sum_{i=0}^{tn-1} a^{(-i-1)}u_i)^2|x_0,\Delta_0] \\ &\le E[\sum_{t=0}^{(\tau_1/n)-1} a^{2tn}(\sum_{i=0}^{t-1} a^{-in}2^{(R'-1)}\Delta_{in})^2|x_0,\Delta_0] \\ &\le E[\sum_{t=0}^{(\tau_1/n)-1} a^{2tn}(\sum_{i=0}^{t-1} a^{-in}2^{(R'-1)}(|a|+\delta)^{in}\Delta_0)^2 \\ &\qquad |x_0,\Delta_0] \\ &\le E[\sum_{t=0}^{(\tau_1/n)-1} a^{2tn}\left(2^{(R'-1)}(\frac{|a|+\delta}{|a|})^{tn}\Delta_0\right)^2|x_0,\Delta_0] \\ &\qquad \times\left(\frac{1}{(1-(\frac{|a|+\delta}{|a|})^n)}\right)^2 \\ &\le \Delta_0^2\tilde{\zeta}_B \end{aligned} \tag{59}$$

for some finite $\tilde{\zeta}_B$, by the bound on the stopping time and arguments presented earlier.

Now, with (53), (55), (58) – (59), we can apply Theorem 7.2: with some $0<\epsilon$ [whose existence is justified by (53)]

$$\delta(x,\Delta) = \epsilon\Delta^2, \quad f(x,\Delta) = \frac{\epsilon}{2\zeta_{B_2}+2\tilde{\zeta}_B}x^2$$

C a compact set and $V(x,\Delta)=\Delta^2$, Theorem 7.2 applies and $\lim_{t\to\infty} E[x_{tn}^2] < \infty$.

Thus, with average rate strictly larger than $\log_2(|a|)$, stability with a finite second moment is achieved. Finally, the limit is independent of the initial distribution since the sampled chain is irreducible, by Theorem 2.3. Now, if the sampled process has a finite second moment, the average second moment for the state process satisfies

$$\lim_{N\to\infty}\frac{1}{N}E[\sum_{k=0}^{N-1} x_k^2] = \frac{1}{n}E_{\pi}[\sum_{k=0}^{n-1} x_k^2|x_0,\Delta_0]$$

which is also finite, where E_π denotes the expectation under the invariant probability measure for x_0, Δ_0. By the ergodic theorem for Markov chains (see Theorem 7.2), the above holds almost surely, and as a result

$$\lim_{N\to\infty}\frac{1}{N}\left(\sum_{k=0}^{N-1} x_k^2\right) = \frac{1}{n}E_{\pi}[\sum_{k=0}^{n-1} x_k^2|x_0,\Delta_0] < \infty \quad \text{a.s.}$$

□

G. Proof of Theorem 2.6

Proof follows from the observation that the number of errors in channel transmission when the state is under-zoomed s is zero. No errors take place in the phase when the quantizer is being zoomed out.

Following (53), the only term which survives is

$$\alpha^{2n} + P^e_{g|g}\left(\alpha^{2n}\left(1+(|a|+\delta)^{2n}+\dots(|a|+\delta)^{2(\lfloor\frac{1}{\kappa}\rfloor)n}\right)\right)$$

which is to be less than 1. We can take $\kappa > 1/2$ for this case. Now

$$\begin{aligned} &\lim_{\Delta\to\infty} P(\tau\ge kn|x_0,\Delta_0) \\ &\qquad \le P(\bar{d} > \frac{(|a|+\delta)^{(k-1)n}\alpha^n}{|a|^{kn}}\Delta_0 2^{R'-1}) = 0 \end{aligned}$$

for $k > \frac{1}{\kappa}$. Hence, $\lim_{n\to\infty} P^e_{g|g}(|a|+\delta)^{2n} \to 0$ is sufficient, since $2 > \frac{1}{\kappa}$. The proof is complete once we recognize $\bar{P}_e$ as $P^e_{g|g}$. □

H. Proof of Theorem 4.2

We provide a sketch of the proof since the analysis follows from the scalar case, except for the construction of an adaptive vector quantizer and the associated stopping time distribution.

Consider the following system:

$$\begin{bmatrix} x^1_{t+1} \\ x^2_{t+1} \\ \vdots \\ x^N_{t+1} \end{bmatrix} = \Lambda \begin{bmatrix} x^1_t \\ x^2_t \\ \vdots \\ x^N_t \end{bmatrix} + \tilde{B}u_t + \tilde{G}d_t \tag{60}$$

where $\Lambda = \mathrm{Diag}(\lambda^i)$ is a diagonal matrix, obtained via a similarity transformation $\Lambda = U^{-1}AU$ and $\tilde{B} = U^{-1}B, \tilde{G} = U^{-1}G$, where U consists of the eigenvectors of the matrix A. We can assume that, without any loss, $\tilde{B}$ is invertible since otherwise, by the controllability assumption, we can sample the system with a period of at most N to obtain an invertible control matrix.

The approach now is quantizing the components in the system according to the adaptive quantization rule provided earlier, except for a joint mapping for the overflow region. We modify the scheme in (5) as follows: let for $i=1,2,\dots,n$, $R_i'(n) = \log_2(2^{R_i(n)}-1) = \log_2(K_i(n))$. The vector quantizer quantizes uniformly the marginal variables and we define the overflow region as the quantizer outside the granular region: $\prod_{i=1}^N[-2^{R_i'(n)-1}\Delta^i, 2^{R_i'(n)-1}\Delta^i]$ and for $i=1,2,\dots,N$

$$Q^{\Delta^i_t}_{K_i}(x) = \mathcal{Z} \quad \text{if} \quad x \notin \prod_{k=1}^N[-2^{R_k'(n)-1}\Delta^k, 2^{R_k'(n)-1}\Delta^k]$$

and for $x \in \prod_{i=1}^N[-2^{R_i'(n)-1}\Delta^i, 2^{R_i'(n)-1}\Delta^i]$, the quantizer quantizes the marginal according to (5) . Hence, here Δ^i is the bin size of the quantizer in the direction of the eigenvector x^i, with rate $R_i'(n)$. For $1\le i\le N$

$$\begin{aligned} u_t &= -1_{\{t=(k+1)n-1\}}\tilde{B}^{-1}\Lambda^n\hat{x}_{kn} \\ \hat{x}^i_t &= Q^{\Delta^i_t}_{K_i}(x^i_t) \\ \Delta^i_{t+1} &= \Delta^i_t\bar{Q}^i(\Delta^i_t, c'_{(t+1)n-1}) \end{aligned} \tag{61}$$

with $\delta^i > 0$, $\alpha^i < 1$, and $L^i > 0$ such that

$$\bar{Q}^i(\Delta^i, c') = (|\lambda^i| + \delta)^n \quad \text{if} \quad c' = \mathcal{Z}$$
$$\bar{Q}^i(\Delta^i, c') = (\alpha^i)^n \quad \text{if} \quad c' \neq \mathcal{Z}, \Delta \geq L^i$$
$$\bar{Q}^i(\Delta^i, c') = 1 \quad \text{if} \quad c' \neq \mathcal{Z}, \Delta < L^i$$

and $R'_i(n) > n \log_2(|\lambda^i|/\alpha^i)$.

Instead of (15), the sequence of stopping times is defined as follows. With $\tau_0 = 0$, define

$$\tau_{z+1} = \inf\{kn > \tau_z : |h^i_{kn}| \leq 1, i = 1, 2 \ldots, N\}, \quad k, z \in \mathbb{Z}_+$$

where $h^i_t = \frac{x^i_t}{\Delta^i_t 2^{R'_i - 1}}$. Now, we observe that N-dimensional system

$$\begin{aligned} &P(\tau_1 > kn | x_0, \Delta_0) \\ &= P\left(\bigcap_{t=1}^{k} \{ \bigcup_{i=1}^{N} (|h^i_t| > 1) \} \mid x_0, \Delta_0 \right) \\ &\leq P\left(\bigcup_{i=1}^{N} (|h^i_{kn}| > 1) | \text{zoom until} \quad k) \mid x_0, \Delta_0 \right) \quad (62) \\ &\leq \sum_{i=1}^{N} P(|h^i_{kn}| > 1 | \text{zoom until} \quad k, x_0, \Delta_0) \quad (63) \end{aligned}$$

where we apply the chain rule for probability in (62) and the union bound in (63). However, for each of the dimensions, $P(|h^i_{kn}| > 1 | \text{zoom until} \quad kn, x_0, \Delta_0)$ is dominated by an exponential measure, and so is the sum. Furthermore, $P(\tau_1 > n | x_0, \Delta_0)$ still converges to 0 provided the rate condition $R'_i(n) > \log_2(|\lambda^i|/\alpha^i)$ is satisfied for every i, since $P(\tau_1 > n | x_0, \Delta_0) \leq \sum_{i=1}^{N} P(|h^i_n| > 1 | x_0, \Delta_0)$. Therefore, analogous results to (45)–(48) are applicable. Once one imposes a countability condition for the bin size spaces as in Theorem 2.3, the desired ergodicity properties are established. □

VI. Concluding Remarks

This paper considered stochastic stabilization of linear systems driven by unbounded noise over noisy channels and established conditions for asymptotic mean stationarity. The conditions obtained are tight with an achievability and a converse. This paper also obtained conditions for the existence of finite second moments. When there is unbounded noise, the result we obtained for the existence of finite second moments required further conditions on reliability for channels when compared with the bounded-noise case considered by Sahai and Mitter. We do not have a converse theorem for the finite second moment discussion; it would be interesting to obtain a complete solution for this setup.

We observed in the development that three types of errors were critical. These bring up the importance of unequal error coding schemes with feedback. Recent results in the literature [10] have focused on fixed-length schemes without feedback, and variable length with feedback and further research could be useful for networked control problems.

The value of information channels in optimization and control problems (beyond stabilization) is an important problem in view of applications in networked control systems. Further research from the information theory community for nonasymptotic coding results will provide useful applications and insight for such problems. These can also be useful to tighten the conditions for the existence of finite second moments. Moderate channel lengths [71], [73], [74], [76] and possible presence of noise in feedback [22] are crucial issues needed to be explored better in the analysis and in the applications of random-time state-dependent drift arguments [94].

Finally, we note that the assumption that the system noise is Gaussian can be relaxed. For the second moment stability, a sufficiently light tail which would provide a geometric bound on the stopping times as in (39) through (27) will be sufficient. For the AMS property, this is not needed. For a noiseless DMC, [97] established that a finite second moment for the system noise is sufficient for the existence of an invariant probability measure. We require, however, that the noise admits a density which is positive everywhere for establishing irreducibility.

A. Variable-Length Coding and Agreement Over a Channel

Let us consider a channel where agreement on a binary event in finite time is possible between the encoder and the decoder. By binary events, we mean, for example, synchronization of encoding times and agreement on zooming times. It turns out that if the following assumption holds, then such agreements are possible in finite expected time: the channel is such that there exist input letters x_1, x_2, x_3, x_4 where $D(P(\cdot|x_1)||P(\cdot|x_2)) = \infty$ and $D(P(\cdot|x_3)||P(\cdot|x_4)) = \infty$. Here, x_1 can be equal to x_4 and x_2 can be equal to x_3. For example, the erasure channel satisfies this property. Note that the aforementioned condition is weaker than having a nonzero zero-error capacity, but stronger than what Burnashev's [14], [93] method requires, since there are more hypotheses to be tested.

In such a setting, one could use variable-length encoding schemes. Such a design will allow the encoder and the decoder to have transmission in three phases: zooming, transmission, and error confirmation. Using random-time, state-dependent stochastic drift, we may find alternative schemes for stochastic stabilization.

Appendix
Stochastic Stability of Dynamical Systems

A) Stationary, Ergodic, and AMS Processes: In this section, we review ergodic theory, in the context of information theory (that is with the transformations being specific to the shift operation). A comprehensive discussion is available in [80] and [32], [33].

Let $\mathbb{X}$ be a complete, separable, metric space. Let $\mathcal{B}(\mathbb{X})$ denote the Borel sigma field of subsets of $\mathbb{X}$. Let $\Sigma = \mathbb{X}^\infty$ denote the sequence space of all one-sided or two-sided infinite sequences drawn from $\mathbb{X}$. Thus, for a two-sided sequence space if $x \in \Sigma$, then $x = \{\ldots, x_{-1}, x_0, x_1, \ldots\}$ with $x_i \in \mathbb{X}$. Let $X_n : \Sigma \to \mathbb{X}$ denote the coordinate function such that $X_n(x) = x_n$. Let T denote the shift operation on Σ, that is $X_n(Tx) = x_{n+1}$. That is, for a one-sided sequence space $T(x_0, x_1, x_2, \ldots) = (x_1, x_2, x_3, \ldots)$.

Let $\mathcal{B}(\Sigma)$ denote the smallest sigma field containing all cylinder sets of the form $\{x : x_i \in B_i, m \leq i \leq n\}$ where $B_i \in \mathcal{B}(\mathbb{X})$, for all integers m, n. Observe that $\cap_{n\geq 0} T^{-n}\mathcal{B}(\Sigma)$ is the tail σ-field: $\cap_n \sigma(x_n, x_{n+1}, \cdots)$, since $T^{-n}(A) = \{x : T^n x \in A\}$.

Let μ be a stationary measure on $(\Sigma, \mathcal{B}(\Sigma))$ in the sense that $\mu(T^{-1}B) = \mu(B)$ for all $B \in \mathcal{B}(\Sigma)$. The sequence of random variables $\{x_n\}$ defined on the probability space $(\Sigma, \mathcal{B}(\Sigma), \mu)$ is a stationary process.

Definition 7.1: Let P be the measure on a process. This random process is ergodic if $A = T^{-1}A$ implies that $P(A) \in \{0, 1\}$.

That is, the events that are unchanged with a shift operation are trivial events. Mixing is a sufficient condition for ergodicity. Thus, a source is ergodic if $\lim_{n\to\infty} P(A \cap T^{-n}B) = P(A)P(B)$, since the process forgets its initial condition. Thus, when one specializes to Markov sources, we have the following: a positive Harris recurrent Markov chain is ergodic, since such a process is mixing and stationary. We will discuss this further in the next section.

Definition 7.2: A random process is N-stationary, (cyclostationary or periodically stationary with period N) if the process measure P satisfies $P(T^{-N}B) = P(B)$ for all $B \in \mathcal{B}(\Sigma)$, or equivalently for any $n \in \mathbb{N}$ samples $t_1, t_2, \ldots, t_n$

$$\begin{aligned} &P(x_{t_1} \in A_1, x_{t_2} \in A_2, \ldots, x_{t_n} \in A_n) \\ &\quad = P(x_{t_1+N} \in A_1, x_{t_2+N} \in A_2, \ldots, x_{t_n+N} \in A_n). \end{aligned}$$

Definition 7.3: A random process is N-ergodic if $A = T^{-N}A$ implies that $P(A) \in \{0, 1\}$.

Definition 7.4: A set $A \in \mathcal{B}(\mathbb{X})$ is coordinate-recurrent if for some $m \in \mathbb{Z}_+$

$$\sum_{m=0}^{\infty} 1_{\{X_m(x)\in A\}} = \infty, \quad \text{a.s.}$$

Definition 7.5: A process on a probability space $(\Omega, \mathcal{F}, P)$ is AMS if there exists a probability measure $\bar{P}$ such that

$$\lim_{N\to\infty} \frac{1}{N} \sum_{k=0}^{N-1} P(T^{-k}F) = \bar{P}(F)$$

for all events F. Here, $\bar{P}$ is called the stationary mean of P, and is a stationary measure.

$\bar{P}$ is stationary since, by definition $\bar{P}(F) = \bar{P}(T^{-1}F)$, for all events F in the tail sigma field for the shift. A cyclostationary process is AMS. See, for example, [9], [33] or [32] (Theorem 7.3.1), that is N-stationarity implies the AMS property.

Asymptotic mean stationarity is a very important property.

1) The Shannon–McMillan–Breiman theorem (the entropy ergodic theorem) applies to finite alphabet AMS sources [33] (see an extension for a more general class [3]). In this case, the ergodic decomposition of the AMS process leads to almost sure convergence of the conditional entropies.
2) Birkhoff's ergodic theorem applies for bounded measurable functions f, if and only if the process is AMS [33].

Let

$$F = \{x : \lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} f(T^i x) \quad \text{exists}\}.$$

It follows that for an AMS process $m(F) = 1$, with m being the stationary mean of the process, Birkhoff's almost-sure ergodic theorem states the following: if a dynamical system is AMS with stationary mean m, then all bounded measurable functions f have the ergodic property, and with probability 1

$$\lim_{N\to\infty} \frac{1}{N} \sum_{i=0}^{N-1} f(T^i x) = E_{m_x}[f], \quad x \in F$$

where E_{m_x} denotes the expectation under measure m_x and m_x is the resulting ergodic measure with initial state x in the ergodic decomposition of the asymptotic mean ([31] *Theorem* 1.8.2): $m(A) = \int m_x(A)m(dx)$. Furthermore

$$\lim_{N\to\infty} \frac{1}{N} E[\sum_{i=0}^{N-1} f(T^i x)] = E_m[f], \quad x \in F.$$

In fact, the above applies for all integrable functions (integrable with respect to the asymptotic mean).

Definition 7.6: A random process is second-moment stable if the following holds:

$$\lim_{N\to\infty} \frac{1}{N} E[\sum_{m=0}^{N-1} (X_m(x))^2] < \infty.$$

Definition 7.7: A random process is quadratically stable (almost surely) if the following limit exists and is finite almost surely:

$$\lim_{N\to\infty} \frac{1}{N} \sum_{m=0}^{N-1} (X_m(x))^2 < \infty.$$

B) Stochastic Stability of Markov Chains and Random-Time State-Dependent Drift Criteria: In this section, we review the theory of stochastic stability of Markov chains. The reader is referred to [62] for a detailed discussion. The results on random-time stochastic drift follows from [94] and [96].

We let $\phi = \{\phi_t, t \geq 0\}$ denote a Markov chain with state space $\mathbb{X}$. The basic assumptions of [62] are adopted: it is assumed that $\mathbb{X}$ is a complete separable metric space, that is locally compact; its Borel σ-field is denoted $\mathcal{B}(\mathbb{X})$. The transition probability is denoted by P, so that for any $\phi \in \mathbb{X}$, $A \in \mathcal{B}(\mathbb{X})$, the probability of moving in one step from the state ϕ to the set A is given by $P(\phi_{t+1} \in A \mid \phi_t = \phi) = P(\phi, A)$. The n-step transitions are obtained via composition in the usual way, $P(\phi_{t+n} \in A \mid \phi_t = \phi) = P^n(\phi, A)$, 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

$$\begin{aligned} Pf(\phi) &:= \int_{\mathbb{X}} P(\phi, dy) f(y), \quad \phi \in \mathbb{X} \\ \mu P(A) &:= \int_{\mathbb{X}} \mu(d\phi) P(\phi, A), \quad A \in \mathcal{B}(\mathbb{X}). \end{aligned}$$

A probability measure π on $\mathcal{B}(\mathbb{X})$ is called invariant if $\pi P = \pi$. That is

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

For any initial probability measure ν on $\mathcal{B}(\mathbb{X})$, we can construct a stochastic process with transition law P, and satisfying $\phi_0 \sim \nu$. We let P_ν denote the resulting probability measure on sample space, with the usual convention for $\nu = \delta_\phi$ when the initial state is $\phi \in \mathbb{X}$. When $\nu = \pi$, the resulting process is stationary.

There is at most one stationary solution under the following irreducibility assumption. For a set $A \in \mathcal{B}(\mathbb{X})$, we denote

$$\tau_A := \min(t \geq 1 : \phi_t \in A). \tag{64}$$

Definition 7.8: Let φ denote a sigma-finite measure on $\mathcal{B}(\mathbb{X})$.

i) The Markov chain is called *φ-irreducible* if for any $\phi \in \mathbb{X}$, and any $B \in \mathcal{B}(\mathbb{X})$ satisfying $\varphi(B) > 0$, we have

$$P_\phi\{\tau_B < \infty\} > 0\,.$$

ii) A φ-irreducible Markov chain is *aperiodic* if for any $\phi \in \mathbb{X}$, and any $B \in \mathcal{B}(\mathbb{X})$ satisfying $\varphi(B) > 0$, there exists $n_0 = n_0(\phi, B)$ such that

$$P^n(\phi, B) > 0 \qquad \text{for all } n \geq n_0.$$

iii) A φ-irreducible Markov chain is *Harris recurrent* if $P_\phi(\tau_B < \infty) = 1$ for any $\phi \in \mathbb{X}$, 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.

Tied to φ-irreducibility is the existence of *small* or *petite* sets. A set $A \in \mathcal{B}(\mathbb{X})$ is small if there is an integer $n_0 \geq 1$ and a positive measure μ satisfying $\mu(\mathbb{X}) > 0$ and

$$P^{n_0}(\phi, B) \geq \mu(B), \quad \textit{for all } \phi \in A, \text{ and } B \in \mathcal{B}(\mathbb{X}).$$

A set $A \in \mathcal{B}(\mathbb{X})$ is petite if there is a probability measure J on the nonnegative integers $\mathbb{N}$, and a positive measure μ satisfying $\mu(\mathbb{X}) > 0$ and

$$\sum_{n=0}^{\infty} P^n(\phi, B)\mathrm{J}(n) \geq \mu(B), \quad \text{for all } \phi \in A, \text{ and } B \in \mathcal{B}(\mathbb{X}).$$

Theorem 7.1 [63]: Suppose that $\boldsymbol{X}$ is a φ-irreducible Markov chain, and suppose that there is a set $A \in \mathcal{B}(\mathbb{X})$ satisfying the following.

i) A is μ-petite for some μ.
ii) A is *recurrent*: $P_\phi(\tau_A < \infty) = 1$ for any $x \in \mathbb{X}$.
iii) A is *finite mean recurrent*: $\sup_{\phi \in A} E_\phi[\tau_A] < \infty$.

Then, $\boldsymbol{X}$ is positive Harris recurrent. □

Let $\mathcal{T}_z,\ z \geq 0$ be a sequence of stopping times, measurable on a filtration generated by the state process with $\mathcal{T}_0 = 0$.

Theorem 7.2: [94], [96]: Suppose that ϕ 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, and a constant $b \in \mathbb{R}$, such that the following holds:

$$\begin{aligned} &E[V(\phi_{\mathcal{T}_{z+1}}) \mid \mathcal{F}_{\mathcal{T}_z}] \leq V(\phi_{\mathcal{T}_z}) - \delta(\phi_{\mathcal{T}_z}) + b1_{\{\phi_{\mathcal{T}_z} \in C\}} \\ &E\left[\sum_{k=\mathcal{T}_z}^{\mathcal{T}_{z+1}-1} f(\phi_k) \mid \mathcal{F}_{\mathcal{T}_z}\right] \leq \delta(\phi_{\mathcal{T}_z}) \qquad z \geq 0. \end{aligned} \tag{65}$$

Then the following hold.

i) ϕ is positive Harris recurrent, with unique invariant distribution π.
ii) $\pi(f) := \int f(\phi)\, \pi(d\phi) < \infty$.
iii) For any function g that is bounded by f, in the sense that $\sup_\phi |g(\phi)|/f(\phi) < \infty$, we have convergence of moments in the mean, and the law of large numbers holds

$$\begin{aligned} &\lim_{t\to\infty} E_\phi[g(\phi_t)] = \pi(g) \\ &\lim_{N\to\infty} \frac{1}{N} \sum_{t=0}^{N-1} g(\phi_t) = \pi(g) \qquad \text{a.s.}, \ \ \phi \in \mathbb{X}. \end{aligned}$$

□

Remark 7.1: We note that the condition $f : \mathbb{X} \to [1, \infty)$ can be relaxed to $f : \mathbb{X} \to [0, \infty)$ provided that one can show that there exists an invariant probability measure.

We conclude by stating a simple corollary to Theorem 7.2, obtained by taking $f(\phi) = 1$ for all $\phi \in \mathbb{X}$.

Corollary 7.1: [94], [96]: Suppose that ϕ is a φ-irreducible Markov chain. Suppose, moreover, that there is a function $V : \mathbb{X} \to (0, \infty)$, a small set C, and a constant $b \in \mathbb{R}$, such that the following hold:

$$\begin{aligned} &E[V(\phi_{\mathcal{T}_{z+1}}) \mid \mathcal{F}_{\mathcal{T}_z}] \leq V(\phi_{\mathcal{T}_z}) - 1 + b1_{\{\phi_{\mathcal{T}_z} \in C\}} \\ &\sup_{z\geq 0} E[\mathcal{T}_{z+1} - \mathcal{T}_z \mid \mathcal{F}_{\mathcal{T}_z}] < \infty. \end{aligned} \tag{66}$$

Then, ϕ is positive Harris recurrent. □

The following is a useful result for this paper.

Theorem 7.3: [62]: Without an irreducibility assumption, if (66) holds for a measurable set C, a function $V : \mathbb{X} \to (0, \infty)$, with $\sup_{x\in C} V(x) < \infty$, then C satisfies $\sup_{x\in C} E[\tau_C] < \infty$.

We have the following results. A positive Harris recurrent Markov process (thus with a unique invariant distribution on the state space) is also ergodic in the sense of ergodic theory (the ergodic theorem for Markov chains has typically a more specialized meaning with the state process being a coordinate process in the infinite-dimensional space $\mathbb{X}^\infty$; see [44]), which however, implies the definition in the more general sense. This follows from the fact that it suffices to test ergodicity on the sets which generate the sigma algebra (that is the finite dimensional sets), which in turn can be verified by the recurrence of the individual sets; probabilistic relations in arbitrary finite sets characterize the properties in the infinite collection, and that, mixing leads to ergodicity.

ACKNOWLEDGMENT

Discussions with Prof. S. P. Meyn, Prof. R. M. Gray, Prof. N. Martins, and Prof. T. Linder on the contents of this paper are gratefully acknowledged. The incisive reviews of two anonymous reviewers and the suggestions of the associate editor have led to significant improvement in the presentation.

REFERENCES

[1] G. E. Andrews, *Number Theory*. New York: Dover, 1995.
[2] R. Bansal and T. Basar, "Simultaneous design of measurement and control strategies for stochastic systems with feedback," *Automatica*, vol. 25, pp. 679–694, 1989.
[3] A. R. Barron, "The strong ergodic theorem for densities: Generalized Shannon McMillan Breiman theorem," *Ann. Probab.*, vol. 13, pp. 1292–1303, 1985.
[4] E. R. Berlekamp, "Block Coding with Noiseless Feedback," Ph.D. dissertation, Dept. Elect. Eng., Massachusetts Inst. Technol., , 1964.
[5] T. Berger, *Rate-Distortion Theory*. Englewood Cliffs, NJ: Prentice-Hall, 1971.
[6] T. Berger, "Information rates of Wiener processes," *IEEE Trans. Inf. Theory*, vol. 16, no. 2, pp. 134–139, Mar. 1970.
[7] P. Billingsley, *Probability and Measure*, 3rd ed. New York: Wiley, 1995.
[8] D. Blackwell, "The comparison of experiments," in *Proc. 2nd Berkeley Symp. Math. Statistics Probability*, Berkeley, 1951, pp. 93–102.
[9] R. A. Boyles and W. A. Gardner, "Cycloergodic properties of discrete-parameter non-stationary stochastic processes," *IEEE Trans. Inf. Theory*, vol. IT-29, no. 1, pp. 105–114, Jan. 1983.
[10] S. Borade, B. Nakiboglu, and L. Zheng, "Unequal error protection: An information-theoretic perspective," *IEEE Trans. Inf. Theory*, vol. 55, no. 12, pp. 5511–5539, Dec. 2009.
[11] V. S. Borkar, "Convex analytic methods in Markov decision processes," in *Handbook of Markov Decision Processes: Methods and Applications*. Boston, MA: Kluwer.
[12] V. S. Borkar and S. K. Mitter, "LQG control with communication constraints," in *Kailath Festschrift*. Boston: Kluwer, 1997, pp. 365–373.
[13] R. Brockett and D. Liberzon, "Quantized feedback stabilization of linear systems," *IEEE Trans. Autom. Control*, vol. 45, no. 7, pp. 1279–1289, Jul. 2000.
[14] M. V. Burnashev, "Data transmission over a discrete channel with feedback. Random transmission time," *Probl. Pered. Inf.*, vol. 12, pp. 10–30, 1976.
[15] C. D. Charalambous, A. Farhadi, D. Denic, and F. Rezaei, "Robust control over uncertain communication channels," in *Proc. IEEE Med. Conf. Control Autom.*, Cyprus, Jun. 2005, pp. 737–742.
[16] G. Como, F. Fagnani, and S. Zampieri, "Anytime reliable transmission of real-valued information through digital noisy channels," *SIAM J. Control Optimization*, vol. 48, pp. 3903–3924, 2010.
[17] L. Coviello, P. Minero, and M. Franceschetti, "Stabilization over Markov feedback channels: The general case," presented at the IEEE Conf. Decision Control, FL, 2011.
[18] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Channels*. Cambridge, U.K.: Cambridge Univ. Press, 2011.
[19] T. Cover and J. Thomas, *Elements of Information Theory*. New York: Wiley, 1991.
[20] R. Dabora and A. Goldsmith, "On the capacity of indecomposable finite-state channels with feedback," in *Proc. Allerton Conf. Commun. Control Comput.*, Urbana, IL, Sep. 2008, pp. 1045–1052.
[21] R. L. Dobrushin, "An asymptotic bound for the probability error of information transmission through a channel without memory using the feedback," *Probl. Kibernetiki*, vol. 8, pp. 161–168, 1962.
[22] S. Draper and A. Sahai, "Noisy feedback improves communication reliability," in *Proc. IEEE Int. Symp. Inf. Theory*, Seattle, WA, Jul. 2006, pp. 69–73.
[23] A. G. D'yachkov, "Upper bounds on the error probability for discrete memoryless channels with feedback," *Probl. Perdachi Inf.*, vol. II, pp. 13–28, 1975.
[24] N. Elia, "When Bode meets Shannon: Control-oriented feedback communication schemes," *IEEE Trans. Autom. Control*, vol. 49, no. 9, pp. 1477–1488, Sep. 2004.
[25] N. Elia and S. K. Mitter, "Stabilization of linear systems with limited information," *IEEE Trans. Autom. Control*, vol. 46, no. 9, pp. 1384–1400, Sep. 2001.
[26] F. Fagnani and S. Zampieri, "Stability analysis and synthesis for scalar linear systems with a quantized feedback," *IEEE Trans. Aut. Control*, vol. 48, no. 9, pp. 1569–1584, Sep. 2003.
[27] R. G. Gallager, "A simple derivation of the coding theorem and some applications," *IEEE Trans. Inf. Theory*, vol. IT-11, no. 1, pp. 3–18, Jan. 1965.
[28] M. Gastpar, B. Rimoldi, and M. Vetterli, "To code, or not to code: Lossy source-channel communication revisited," *IEEE Trans. Inf. Theory*, vol. 49, no. 5, pp. 1147–1158, May 2003.
[29] D. J. Goodman and A. Gersho, "Theory of an adaptive quantizer," *IEEE Trans. Commun.*, vol. 22, no. 8, pp. 1037–1045, Aug. 1974.
[30] R. M. Gray, "Information rates of autoregressive processes," *IEEE Trans. Inf. Theory*, vol. 16, no. 4, pp. 412–421, Jul. 1970.
[31] R. M. Gray, *Entropy and Information Theory*. New York: Springer, 2011.
[32] R. M. Gray, *Probability, Random Processes, and Ergodic Properties*. New York: Springer-Verlag, 1988.
[33] R. M. Gray and J. C. Kieffer, "Asymptotically mean stationary measures," *Ann. Prob.*, vol. 8, pp. 962–973, 1980.
[34] R. M. Gray and T. Hashimoto, "A note on rate-distortion functions for nonstationary Gaussian autoregressive processes," *IEEE Trans. Inf. Theory*, vol. 54, no. 3, pp. 1319–1322, Mar. 2008.
[35] R. M. Gray and F. Saadat, "Block source coding theory for asymptotically mean stationary sources," *IEEE Trans. Inf. Theory*, vol. IT-30, no. 1, pp. 64–67, Jan. 1984.
[36] P. W. Wong and R. M. Gray, "Sigma-delta modulation with i.i.d. Gaussian inputs," *IEEE Trans. Inf. Theory*, vol. 36, no. 4, pp. 784–778, Jul. 1990.
[37] P. Grover and A. Sahai, "Implicit and explicit communication in decentralized control," presented at the Annu. Allerton Conf. Commun., Control, Comput., Monticello, IL, Sep. 2010.
[38] A. Gurt and G. Nair, "Internal stability of dynamic quantised control for stochastic linear plants," *Automatica*, vol. 45, pp. 1387–1396, Jun. 2009.
[39] E. A. Haroutunian, "A lower bound of the probability of error for channels with feedback," *Probl. Peredachi Inf.*, vol. 13, pp. 36–44, 1977.
[40] T. Hashimoto and S. Arimoto, "On the rate-distortion function for the nonstationary Gaussian autoregressive process," *IEEE Trans. Inf. Theory*, vol. IT-26, no. 4, pp. 478–480, Jul. 1980.
[41] M. Horstein, "Sequential decoding using noiseless feedback," *IEEE Trans. Inf. Theory*, vol. IT-9, no. 3, pp. 136–143, Jul. 1963.
[42] O. C. Imer, S. Yüksel, and T. Basar, "Optimal control of LTI systems over unreliable communication links," *Automatica*, vol. 42, pp. 1429–1440, 2006.
[43] H. Ishii and B. A. Francis, *Limited Data Rate in Control Systems with Networks*, ser. Lecture Notes in Control and Information Sciences. Berlin: Springer, 2002, vol. 275.
[44] O. Hernandez-Lerma and J. B. Lasserre, *Markov Chains and Invariant Probabilities*. Basel: Birkhäuser Verlag, 2003.
[45] W. S. Wong and R. W. Brockett, "Systems with finite communication bandwidth constraints—Part II: Stabilization with limited information feedback," *IEEE Trans. Autom. Control*, vol. 42, no. 9, pp. 1294–1299, May 1997.
[46] A. Johnston and S. Yüksel, "Stochastic Stabilization of Partially Observed and Multi-Sensor Systems Driven by Gaussian Noise Under Fixed-Rate Information Constraints" to appear in IEEE Conf. Decision and Control, Hawaii, Dec., 2012.
[47] M. Kac, "On the notion of recurrence in discrete stochastic processes," *Bull. Amer. Math. Soc.*, vol. 53, pp. 1002–1010, 1947.
[48] J. C. Kieffer, "Stochastic stability for feedback quantization schemes," *IEEE Trans. Inf. Theory*, vol. IT-28, no. 2, pp. 248–254, Mar. 1982.
[49] J. C. Kieffer, "On the transmission of Bernoulli sources over stationary channels," *Ann. Probab.*, vol. 8, no. 5, pp. 942–961, 1980.
[50] J. C. Kieffer and M. Rahe, "Markov channels are asymptotically mean stationary," *SIAM J. Math. Anal.*, vol. 12, no. 3, pp. 293–305, 1981.
[51] J. C. Kieffer, "Analysis of dc input response for a class of one-bit feedback encoders," *IEEE Trans. Commun.*, vol. 38, no. 3, pp. 337–340, Mar. 1990.
[52] J. C. Kieffer, "On sliding block coding for transmission of a source over a stationary nonanticipatory channel," *Inf. Control*, vol. 35, no. 1, pp. 1–19, Mar. 1977.
[53] J. C. Kieffer and J. G. Dunham, "On a type of stochastic stability for a class of encoding schemes," *IEEE Trans. Inf. Theory*, vol. IT-29, no. 6, pp. 793–797, Nov. 1983.
[54] A. N. Kolmogorov, "On the Shannon theory of information transmission in the case of continuous signals," *IRE Trans. Inf Theory*, vol. IT-2, no. 4, pp. 102–108, Dec. 1956.

[55] N. C. Martins, M. A. Dahleh, and N. Elia, "Feedback stabilization of uncertain systems in the presence of a direct link," *IEEE Trans. Autom. Control*, vol. 51, no. 3, pp. 438–447, Mar. 2006.

[56] N. C. Martins and M. A. Dahleh, "Feedback control in the presence of noisy channels: 'Bode-Like' fundamental limitations of performance," *IEEE Trans. Autom. Control*, vol. 53, no. 7, pp. 1604–1615, Aug. 2008.

[57] N. C. Martins, M. A. Dahleh, and J. C. Doyle, "Fundamental limitations of disturbance attenuation in the presence of side information," *IEEE Trans. Autom. Control*, vol. 52, no. 1, pp. 56–66, Jan. 2007.

[58] A. S. Matveev, "State estimation via limited capacity noisy communication channels," *Math. Control, Signals, Syst.*, vol. 20, pp. 135–135, 2008.

[59] A. S. Matveev and A. V. Savkin, *Estimation and Control Over Communication Networks*. Boston, MA: Birkhauser, 2008.

[60] A. S. Matveev and A. Savkin, "An analogue of Shannon information theory for detection and stabilization via noisy discrete communication channels," *SIAM J. Control Optim*, vol. 46, pp. 1323–1367, 2007.

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

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

[63] S. P. Meyn and R. Tweedie, "Stability of Markovian processes: I. Criteria for discrete-time chains," *Adv. Appl. Probab.*, vol. 24, pp. 542–574, 1992.

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

[65] P. Minero, M. Franceschetti, S. Dey, and G. Nair, "Data rate theorem for stabilization over time-varying feedback channels," *IEEE Trans. Autom. Control*, vol. 54, no. 2, pp. 243–255, Feb. 2009.

[66] 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.

[67] B. Nakiboglu and L. Zheng, "Upper bounds to error probability with feedback," in *Proc. IEEE Int. Symp. Inf. Theory*, 2009, pp. 1515–1519.

[68] E. Nummelin, *General Irreducible Markov Chains and Non-negative Operators,* 1984.

[69] J. K. Omura, "On general Gilbert bounds," *IEEE Trans. Inf. Theory*, vol. 19, no. 5, pp. 661–666, Sep. 1973.

[70] R. Ostrovsky, Y. Rabani, and L. Schulman, "Error-correcting codes for automatic control," *IEEE Trans. Inf. Theory*, vol. 55, no. 7, pp. 2931–2941, Jul. 2009.

[71] H. Palaiyanur and A. Sahai, "An upper bound for the block coding error exponent with delayed feedback," in *Proc. IEEE Int. Symp. Inf. Theory*, Jun. 2010, pp. 246–250.

[72] H. H. Permuter, T. Weissman, and A. J. Goldsmith, "Finite state channels with time-invariant deterministic feedback," *IEEE Trans. Inf. Theory*, vol. 55, no. 2, pp. 644–662, Feb. 2009.

[73] Y. Polyanskiy, H. V. Poor, and S. Verdú, "Feedback in the non-asymptotic regime," *IEEE Trans. Inf. Theory*, vol. 57, no. 8, pp. 4903–4925, Aug. 2011.

[74] Y. Polyanskiy, H. V. Poor, and S. Verdú, "Channel coding rate in the finite blocklength regime," *IEEE Trans. Inf. Theory*, vol. 56, no. 5, pp. 2307–2359, May 2010.

[75] A. Sahai, "Coding unstable scalar Markov processes into two streams," in *Proc. IEEE Int. Symp. Inf. Theory*, 2004, pp. 462–462.

[76] A. Sahai, "Why delay and block length are not the same thing for channel coding with feedback," presented at the Inf. Theory Appl., San Diego, CA, Jan. 2006.

[77] A. Sahai, "Any-time information theory," Ph.D. dissertation, Massachusetts Inst. Technol., Cambridge, MA, 2001.

[78] A. Sahai and S. Mitter, "The necessity and sufficiency of anytime capacity for stabilization of a linear system over a noisy communication link. Part I: scalar systems," *IEEE Trans. Inf. Theory*, vol. 52, no. 8, pp. 3369–3395, Aug. 2006.

[79] L. Schenato, B. Sinopoli, M. Franceschetti, K. Poolla, and S. S. Sastry, "Foundations of control and estimation over lossy networks," *Proc. IEEE*, vol. 95, no. 1, pp. 163–187, Jan. 2007.

[80] P. C. Shields, *The Ergodic Theory of Discrete Sample Paths*. Providence, RI: American Mathematical Society, 1996.

[81] R. T. Sukhavasi and B. Hassibi, Linear error correcting codes with anytime reliability, 2011.

[82] N. Şen, F. Alajaji, and S. Yüksel, "Feedback capacity of a class of symmetric finite-state Markov channels," *IEEE Trans. Inf. Theory*, vol. 56, no. 7, pp. 4110–4122, Jul. 2011.

[83] T. Simsek and P. Varaiya, "Noisy data-rate limited estimation: Renewal codes," in *Proc. IEEE Conf. Decision Control*, Dec. 2003, pp. 3149–3154.

[84] S. Tatikonda, A. Sahai, and S. Mitter, "Stochastic linear control over a communication channels," *IEEE Trans. Autom. Control*, vol. 49, no. 9, pp. 1549–1561, Sep. 2004.

[85] S. Tatikonda, "Control under communication constraints," Ph.D. dissertation, MIT, Cambridge, MA, 2000.

[86] S. Tatikonda and S. Mitter, "Control under communication constraints," *IEEE Trans. Autom. Control*, vol. 49, no. 7, pp. 1056–1068, Jul. 2004.

[87] S. Tatikonda and S. Mitter, "The capacity of channels with feedback," *IEEE Trans. Inf. Theory*, vol. 55, no. 1, pp. 323–349, Jan. 2009.

[88] R. L. Tweedie, "Drift conditions and invariant measures for Markov chains," *Stochastic Proc. Their Appl.*, vol. 92, pp. 345–354, 2001.

[89] S. Vembu, S. Verdú, and Y. Steinberg, "The source-channel separation theorem revisited," *IEEE Trans. Inf. Theory*, vol. 41, no. 1, pp. 44–54, Jan. 1995.

[90] S. Verdú and T. S. Han, "A general formula for channel capacity," *IEEE Trans. Inf. Theory*, vol. 40, no. 4, pp. 1147–1157, Jul. 1994.

[91] J. C. Walrand and P. Varaiya, "Optimal causal coding-decoding problems," *IEEE Trans. Inf. Theory*, vol. IT-19, no. 11, pp. 814–820, Nov. 1983.

[92] H. S. Witsenhausen, "On the structure of real-time source coders," *Bell Syst. Tech. J*, vol. 58, pp. 1437–1451, Jul./Aug. 1979.

[93] H. Yamamoto and K. Itoh, "Asymptotic performance of a modified Schalkwijk-Barron scheme for channels with noiseless feedback," *IEEE Trans. Inf. Theory*, vol. 25, no. 6, pp. 729–733, Nov. 1979.

[94] 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*.

[95] S. Yüksel and T. Basar, "Control over noisy forward and reverse channels," *IEEE Trans. Autom. Control*, vol. 56, no. 5, pp. 1014–1029, May 2011.

[96] S. Yüksel, "A random time stochastic drift result and application to stochastic stabilization over noisy channels," presented at the Annu. Allerton Conf., IL, Oct. 2009.

[97] S. Yüksel, "Stochastic stabilization of noisy linear systems with fixed-rate limited feedback," *IEEE Trans. Autom. Control*, vol. 55, no. 12, pp. 2847–2853, Dec. 2010.

[98] S. Yüksel, "A tutorial on quantizer design for networked control systems: Stabilization and optimization," *Appl. Comput. Math.*, vol. 11, pp. 365–403, 2011.

[99] K. S. Zigangirov, "Upper bounds for the error probability for channels with feedback," *Probl. Perdachi Inf.*, vol. 6, pp. 87–92, 1970.

Serdar Yüksel received his B.Sc. degree in Electrical and Electronics Engineering from Bilkent University in 2001; 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 postdoctoral researcher at Yale University for a year before joining Queen's University as an Assistant Professor of Mathematics and Engineering at the Department of Mathematics and Statistics. His research interests are on stochastic and decentralized control, information theory, and applied probability.