Control Over Noisy Forward and Reverse Channels

Serdar Yüksel, *Student Member, IEEE*, and Tamer Başar, *Fellow, IEEE*

Abstract—We consider the problem of remotely controlling a continuous-time linear time-invariant system driven by Brownian motion process, when communication takes place over noisy memoryless discrete- or continuous-alphabet channels. What makes this class of remote control problems different from most of the previously studied models is the presence of noise in both the forward channel (connecting sensors to the controller) and the reverse channel (connecting the controller to the plant). For stability of the closed-loop system, we look for the existence of an invariant distribution for the state, for which we show that it is necessary that the entire control space and the state space be encoded, and that the reverse channel be at least as reliable as the forward channel. We obtain necessary conditions and sufficient conditions on the channels and the controllers for stabilizability. Using properties of the underlying sampled Markov chain, we show that under variable-length coding and some realistic channel conditions, stability can be achieved over discrete-alphabet channels even if the entire state and control spaces are to be encoded and the number of bits that can be transmitted per unit time is strictly bounded. For control over continuous-alphabet channels, however, a variable rate scheme is not necessary. We also show that memoryless policies are rate-efficient for Gaussian channels.

Index Terms—Information theory, networked control systems, stochastic control, stochastic stability.

I. Introduction

THE use of digital and wireless channels such as the Internet or bus lines (as in a Controller Area Network (CAN)) in control systems has become common place. Some salient examples include vehicle control systems, large-scale printers and aerospace applications. The presence of such channels brings up many challenges, since having a heterogeneous communication link between the actuator and the sensor or the sensor and the controllers as well as between the controller and the plant makes the traditional control approaches such as separation of estimation and control [17], Kalman filtering and linear control [1] design inapplicable or inefficient.

There are generally three types of agents involved in a remote control system (see later Fig. 1): sensors, controllers, and actuators (plants), one or more of which might be connected through

communication channels on the network. Further, there could be more than one of each, with limited communication among them, thus possibly leading to a decentralized structure. Various forms of this problem have been introduced and studied in the networked control literature, see for example [1] and [6], among others, which, however, have addressed the situation when the channels that connect the controller to the plant (*reverse channels*) are noiseless. In this paper, on the other hand, we deal with networked (linear) systems where the reverse channels (as well as the forward ones—those that connect the sensors to the controller) are noisy. Both discrete- and continuous-alphabet channels are considered, and in both cases they are taken to be memoryless. Before discussing the contents of this paper, and its contributions, let us first provide a brief overview of the existing results in the literature on the general topic of control over noisy channels.

One of the earliest papers on the topic is [1], which has shown that for a scalar discrete-time linear Gaussian system controlled over a Gaussian channel, the encoder and the controllers with noiseless causal feedback which jointly minimize a quadratic objective functional are all linear. This was perhaps the first paper that used information theory along with stochastic control in the analysis of a control system, whose results were subsequently extended to continuous-time systems in [3]. References [45] and [44] studied the optimal causal coding problem over respectively a noiseless channel and a noisy channel with noiseless feedback. Reference [10] showed the chaotic nature of quantization in control in one of the first papers to bring in quantization as a design limitation; [6], on the other hand, studied the trade-off between delay and reliability, and formulated relevant and challenging problems; the questions that were posed there led to an accelerated pace of research efforts on the topic: Significant progress on the connection between information theory and control has been achieved through study of the minimum information rate requirements needed for stabilizability over noisy channels with noiseless feedback, under various assumptions on the noise models, as well as control over noiseless channels—as reported in [41], [46], and [29], where [29] also considered a class of quantizer policies for systems driven by noise with unbounded support set for its probability measure. Several studies in the literature have focused on noiseless discrete channels with time-invariant encoders, where the main issue becomes one of design of an invariant quantizer; see [12] and [13]. The first of these, [12], adopts a Lyapunov-based approach to stabilize a system with limited information, and shows that the coarsest quantizer achieving stability is logarithmic and that the design is universal, i.e., it has the same base for construction regardless of the sampling interval. We will show in this paper that this property regarding sampling carries over to stochastic systems as well. The second reference, [13], quantifies the relationship between the rate of convergence and communication rate for scalar systems, and provides conditions for set invariance as a measure
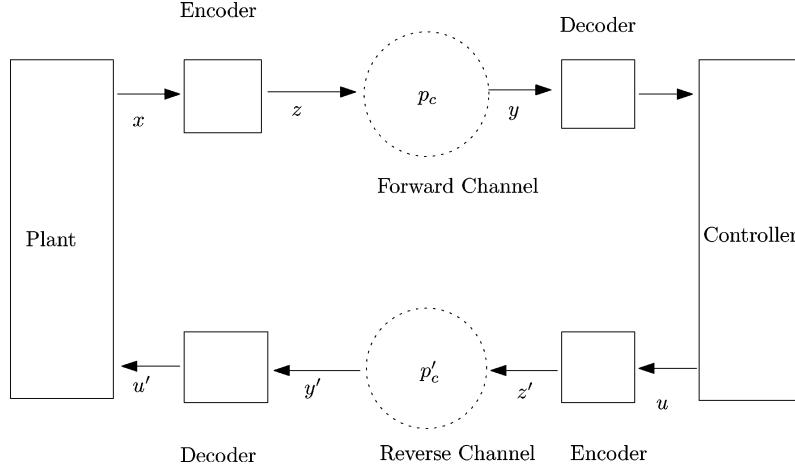
Fig. 1. Control over noisy channels.

of stability. An information theoretic approach to quantizer design leading to stabilizability was presented in [51]. The Bode integral formula arguments were presented in [11] and [24] as alternative and insightful derivations to interpret the information theoretic approach to the derivation of the fundamental data rates for stabilization and disturbance rejection. In [19], a Lyapunov theoretic quantizer design leading to stability is provided, with further results reported in [18], and time-varying quantizer design presented also in [36]. Reference [9] studied control over channels with partial observations, and adopted an information theoretic approach, also with a robust formulation on the class of sources. Entropy approach to disturbance rejection problems was earlier reported in [20]. For noisy binary forward channels with noiseless feedback, coding schemes were presented for the forward channel with noiseless feedback in [40], and linear time-invariant as well as time-varying control policies on continuous-alphabet channels modeled as SNR channels were considered in [7].

It should be noted that the requirements on the channels depend on what the underlying objective is. For instance, the objective can be one of moment stabilizability, or almost sure stabilizability, or positive recurrence of the chain. In [35], moment stabilizability was considered and the notion of *anytime capacity* introduced as a relevant measure when there is noiseless feedback. This paper investigates a similar framework as in [35], with however two main differences. First, we consider systems driven by unbounded noise, which makes the almost-sure tracking of the state by the controller within a bounded confidence interval impossible. Second, we take the reverse channel to be noisy. The price we pay, however, is the redundancy in the transmission of the correlated information in the sensor messages, since innovation encoding is not possible. Another related reference, [25], studies stability over noisy discrete channels. There, the plant is noise-free, and the reverse channel is noiseless; for such a system, it is argued that capacity is a sufficient measure. In our case, however, the plant and the reverse channels are also noisy. Reference [25] also argues that almost sure stabilizability is not possible for control over DMCs and that it is the zero-error capacity which is the relevant notion. We will observe that for noiseless plants, if the number of channel uses is penalized, then capacity is not a sufficient measure except for

noiseless discrete channels. Furthermore, we provide structural results on coding and decoding schemes for stabilizability. Reference [29] has studied systems driven by unbounded noise, but for only noiseless channels and a perfect feedback channel.

Regarding noisy reverse channels, there have been only a few studies: An early work is [42], which has studied the Gaussian channel case, with no encoding in the reverse channel, in the context of relaxation of the noiseless feedback channel. References [17] and [39] have studied optimal control policies for control over communication networks with UDP and TCP type protocols with packet losses in the feedback channel as well as the forward one. The case with delayed communications has been studied in [38].

An information structure with noiseless, instantaneous feedback leads to a tractable problem, as the information at the controller is nested at the sensor on the plant side. For such a system, one can introduce an extended Markov chain where the state space is the space of conditional distributions on the real line with the topology of weak convergence. Further, one can formulate an optimal control problem of choosing the quantizer bin edges so as to minimize a long-term average cost function, as it has been done in [48], which provides existence results for optimal sensing and control. When noiseless feedback is present, the streaming coding schemes in [34] could be used, exploiting the nested structure. Along these lines, when the channel is Gaussian, [1] provided a comprehensive study on the optimality of linear coding (innovation) and control policies. Reference [28] considered time-varying channels.

The presence of a noisy channel with no explicit feedback leads to a *non-classical* information structure (see e.g., [1]), since the agents (controller, encoders and decoders) do not have nested information. Furthermore, the *dual role* of control is present, as the control policy might affect the estimation error of the controller with respect to the state of the system. As such, any stochastic optimal control formulation leads to an intractable distributed control problem, and there is no systematic design of practical value. To alleviate the difficulty and obtain a tractable problem, one could consider placing restrictions on the memory at the controller [23]; however, even with such memory restrictions, the issues encountered in this paper would still be present. Another possibility is the

zooming technique introduced in [22], and its extension for systems driven by unbounded noise [47], which, however, also requires synchronization between the encoder and the decoder. Further, the issues studied here would still come up regardless of the degree of memory involved at the decoder on the plant side, if the memory is finite. Finally, toward a general theory for constructing codes for dynamic systems taking values in a finite set, a graph-based, trajectory tracking code construction has been presented in [32].

To recap, in contradistinction with most of the previous work on this topic, we study here systems where both (the forward and the reverse) channels are noisy, and further we consider the case where the system noise has unbounded support. We use indirect methods, information theory, and Markov stability theory, to arrive at a set of necessary as well as a set of sufficient conditions. We also provide a new code design via variable-rate coding through state-dependent sampling, which leads to stochastic stability of the Markov chain. In fact, such a stochastic drift type approach seems to be a powerful tool; for a related result, the reader is referred to [49].

A precise problem formulation and a summary of the main results are provided in the next section, followed by some background material on Markov chains and coding as relevant to the developments in the paper. Section IV provides some general necessary conditions for stability, which are further developed for DMCs in the next two sections, Section V and VI, without and with system noise, respectively. Section VII discusses the significance of the side channel, the use of variable-length coding through variable rate sampling to avoid the need for infinitely many codewords. The last section of the main body of the paper, Section VIII, deals with continuous alphabet channels, and the paper ends with the concluding remarks of Section IX and seven appendices containing formal proofs of the results in the main body.

II. PROBLEM FORMULATION

We consider in this paper stabilization of open-loop unstable continuous-time linear time-invariant (LTI) stochastic systems when communications between the plant and the controller (forward channel), and the controller and the plant (reverse channel) are conducted over noisy channels which are either discrete memoryless or continuous memoryless. As we will argue later, at the conceptual level there is no loss of generality in restricting our attention to scalar systems. This will enable us to discuss the main challenges, introduce the essential ideas, and present the development without unnecessarily cluttering the notation and garbling the discussion. Accordingly, we consider the class of LTI continuous-time scalar systems described by

$$dx'_{t'} = (\mu x'_{t'} + b'u'_{t'})dt' + dB_{t'}, \quad t' \geq 0 \qquad (1)$$

where $x'_{t'}$ is the state; $B_{t'}, t' \geq 0$, is the standard Brownian motion process; $u'_{t'}$ is the (applied) control, which is assumed to be piecewise constant over intervals of length T_s (which is a constant sampling period); the initial state x_0 is a second-order Gaussian random variable, and $\mu > 0$, thus making the open-loop system unstable. After sampling, with period T_s, we have the discrete-time system

$$x_{t+1} = ax_t + bu'_t + d_t, \quad t = 0, 1, \ldots \qquad (2)$$

where t is the discrete-time variable, defined through the relationship $t' = tT_s$; $x_t = x'_{tT_s}$ is the state x' at the sampling times; $\{d_t\}$ is a sequence of zero-mean independent, identically distributed (i.i.d.) Gaussian random variables; $a = e^{\mu T_s}$; $b = b'(e^{\mu T_s} - 1)/\mu$; and $E[d_t^2] = (e^{2\mu T_s} - 1)/2\mu$. We refer to the channel carrying the signal from the plant to the controller as the *forward channel* and the channel carrying the signal from the controller to the plant as the *reverse (feedback) channel* (see Fig. 1), and as we have indicated earlier, the plant is controlled over a reverse channel and the controller receives information over the forward channel, both of which are noisy. Let us first revisit the notions of discrete and continuous (alphabet) memoryless channels.

Definition 2.1: A Discrete Memoryless Channel (DMC) is characterized by a discrete input alphabet $\mathcal{M}$, and a discrete output alphabet $\mathcal{M}'$, and a conditional probability mass function $p(m'|m)$, from $\mathcal{M} \times \mathcal{M}'$ to $\mathbb{R}$. Let $m_0^n \in \mathcal{M}^{n+1}$ be a sequence of input symbols, $\{m_0, m_1, \ldots, m_n\}$, and let $m_0'^n \in \mathcal{M}'^{n+1}$ be a sequence of output symbols, $\{m_0', m_1', \ldots, m_n'\}$, where $m_k \in \mathcal{M}$ and $m_k' \in \mathcal{M}'$ for all k. Let p_{DMC}^{n+1} be the conditional probability 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: $p_{DMC}^{n+1}(m_0'^n|m_0^n) = \prod_{k=0}^{n} p_{DMC}(m_k'|m_k), \; \forall m^n \in \mathcal{M}^{n+1}, m_0'^n \in \mathcal{M}'^{n+1}$, where m_k, m_k' denote the kth components of the vectors $m_0^n, m_0'^n$, respectively. $\diamond$

Definition 2.2: A Continuous Memoryless Channel (CMC) is characterized by a continuous input alphabet $\mathcal{M} \subset \mathbb{R}$, and a continuous output alphabet $\mathcal{M}' \subset \mathbb{R}$, and a conditional probability measure $p(A|m)$, from $\mathcal{M} \times \mathcal{B}(\mathcal{M}')$ to $\mathbb{R}$, where $\mathcal{B}(\mathcal{M}')$ is the Borel σ-algebra of $\mathcal{M}'$. Let $m_0^n \in \mathcal{M}^{n+1}$ be a sequence of input symbols, $\{m_0, m_1, \ldots, m_n\}$, and let A_0^n be a sequence of Borel subsets in $\mathcal{B}(\mathcal{M}')$, $\{A_0, A_1, \ldots, A_n\}$, where $m_k \in \mathcal{M}$ and $A_k \in \mathcal{B}(\mathcal{M}')$ for all k. Let p_{CMC}^{n+1} be the conditional probability measure on the n-tuple of the input and Borel measurable output spaces. A CMC from $\mathcal{M}^{n+1}$ to $\mathcal{M}'^{n+1}$ satisfies the following:

$$p_{CMC}^{n+1}(A_0 \times A_1 \times \cdots A_n|m_0^n)$$
$$= \prod_{k=0}^{n} p_{CMC}(A_k|m_k), \; \forall m^n \in \mathcal{M}^{n+1}$$

where m_k denotes the kth component of the vector m_0^n. $\diamond$

When the forward channel is a DMC, we will let M_f denote the set of sensor symbols, with $|M_f|$ denoting its cardinality, and N_f be the number of channel uses. Then, the coding rate for the forward channel is defined as $R_f = \log(|M_f|)/N_f$. Likewise, for the reverse channel, for the DMC case, the coding rate is given by $R_r = \log(|M_r|)/N_r$ with obvious corresponding meanings for M_r and N_r. The plant output (state) is quantized (by the source coder) and turned into a bit stream for each quantization symbol (by the channel encoder) before inputed to the forward channel (see Fig. 1). Likewise, the controller output is quantized and turned into a bit stream before inputed to the reverse channel.

Our objective is to obtain upper and lower bounds on communication rates such that the overall system, seen as a Markov chain, is stable with an invariant distribution. In addition to these rate bounds, we will also construct new coding schemes. The

specific models of control systems we will consider include discrete-alphabet noisy channels with or without system noise, and continuous-alphabet noisy channels with system noise.

Before stating the main results of the paper, we note that the analysis we will carry out for the scalar system is readily applicable (at least conceptually) to multivariable systems. Consider a multi-dimensional LTI continuous-time system described by

$$dx'_{t'} = (\tilde{A}x'_{t'} + \tilde{B}u'_{t'})dt' + dW_{t'} \quad y'_{t'} = Cx'_{t'} \qquad (3)$$

where, x' is the n-dimensional state process, W is the standard vector Brownian motion process, u' is the m-dimensional (applied) control which is assumed to be piecewise constant over intervals of length T_s, y' is an l–dimensional output or observation process, and the initial state x_0 is a second-order zero-mean random vector with a given covariance matrix. Without any loss of generality, one can take the matrix $\tilde{A}$ to be in Jordan form, which we assume to have at least one unstable eigenvalue, and also $\tilde{B}$ and $\tilde{C}$ to be of full rank. Let $(\tilde{A}, \tilde{B})$ be controllable, and $(\tilde{A}, \tilde{C})$ be observable, so that without control the output is unbounded. Now, following sampling, with period T_s, we have the discrete-time system

$$x_{t+1} = Ax_t + Bu'_t + d_t, \quad y_t = Cx_t \qquad (4)$$

where, as in the scalar case, x_t is the state at time t, and $\{d_t\}$ is a sequence of i.i.d. Gaussian random vectors with covariance matrix Q. In the above, $A := e^{\tilde{A}T_s}$, which is upper-triangular, $B := \int_0^{T_s} e^{\tilde{A}(T_s-\tau)}\tilde{B}d\tau$, and $Q := \int_0^{T_s} e^{\tilde{A}\tau}e^{\tilde{A}^T\tau}d\tau$. We will assume, without any loss of generality (due to the controllability and observability assumptions), that the matrices B and C are both full rank. We note that due to the upper-triangular structure of the sampled system matrix, the analysis for the scalar case will apply to the vector case as well: Using an induction type argument, one can first start with the lowest mode (in the matrix diagonal) of the system, and stabilize that mode so that there is a finite invariant second moment of the state; one can then regard this mode as disturbance for the one upper mode and obtain the conditions for stabilizability for this mode. We note that the random process for the upper mode might not have Markov dynamics, as the marginal probability distribution on the upper mode might not be Markov, due to the statistical coupling with the lower modes, but the joint system consisting of all modes will be Markov. Hence, a sequential scalar coding scheme will be stabilizing, if all of the scalar schemes are stabilizing as individual scalar systems driven by noise (of increasing variance as one moves up in the matrix).

A. Main Results

Before closing this section, we list below the main results and main contributions of this paper in the context of the problem formulated above.

1) For both discrete memoryless channels (DMCs) and continuous memoryless channels (CMCs), we obtain necessary conditions on the channel capacities for the existence of a controller that will lead to $\limsup_{T\to\infty} E[x_T^2] < \infty$.
2) We show that if the underlying closed-loop system is described by an irreducible Markov chain (such as in the case

when the system noise has unbounded support for its probability distribution), then the entire state space and the control space have to be encoded.
3) For the DMC case, we show that variable length coding can stabilize the system. To meet the finite length constraint, variable length codewords are transmitted through variable length sampling intervals. Using properties of the sampled Markov chain, we provide stability conditions for the Markov chain under the criterion of positive Harris recurrence.
4) With DMCs, a two-part code is devised to stabilize the system: One code transmits a timing information through an arbitrarily low-rate non-zero zero-error capacity channel, and the other code encodes the state information in a given, state-dependent range.
5) For CMCs, time-invariant policies can lead to stability and there is no need for variable rate-sampling. Gaussian channels are considered in particular; optimal memoryless linear policies are obtained and shown to be rate-efficient.

III. PRELIMINARIES

Referring to Fig. 1, in our setup both the sensor and the controller act as both transmitters and receivers because of the closed-loop structure. For the DMC case, we model the forward source-channel encoder as a stochastic kernel $p_s(z_t|x_t)$, $x_t \in \mathbb{R}$, $z_t \in \mathcal{Z}$ (with $\mathcal{Z}$ being the channel alphabet), between the source output and the channel input; hence $p_s(z_t|x_t)$ is a collection of (conditional) probability mass functions parametrized by $x_t \in \mathbb{R}$. The forward channel is a memoryless stochastic kernel, $p_c(y_t|z_t)$, between the channel input and the channel output, where $y_t \in \mathcal{Y}$, the output channel alphabet. The channel output is acted on by the controller in a *memoryless* fashion, so that we have another well-defined stochastic kernel, $p(u_t|y_t)$, which is the probability for control at time t to be $u_t \in \mathcal{U}$ given that the output of the forward channel at time t is $y_t \in \mathcal{Y}$.

The reverse channel also has a source-channel encoder, $p'_s(z'_t|u_t), z'_t \in \mathcal{Z}'$, channel mapping $p'_c(y'_t|z'_t), y'_t \in \mathcal{Y}'$, and a channel decoder $p'_d(u'_t|y'_t), u'_t \in \mathcal{U}'$. Appropriate adjustments are made to the interpretations of these different stochastic kernels in the case of CMCs.

For the DMC case, a quantizer is used to obtain a countable representation of the input source. A scalar quantizer Q is a Borel-measurable mapping from the real line to a finite or countable set. We take the quantization bins to be regular. Thus, the bins, $\mathcal{B}_i$, can be taken to be non-overlapping semi-open intervals, $\mathcal{B}_i = [\delta_i, \delta_{i+1})$ for $i > 0$, with $\delta_i < \delta_{i+1}, i = 0, 1, 2, \ldots$, such that δ_0 is at the origin, where $\{\delta_i\}$ are termed "bin edges". We consider "symmetric quantizers," in the sense that if $(\delta_i, \delta_{i+1}]$ is a quantization bin, where $0 < \delta_i < \delta_{i+1}$, then $\mathcal{B}_{-i} := (-\delta_{i+1}, -\delta_i]$ is also a quantization bin. We define the encodable state set $S_x \subset \mathbb{R}$ as the set of elements which are represented by some codeword; $S_x := \bigcup_i \mathcal{B}_i$. Such a definition applies to the encodable control set, S_c, as well. Suppose that the state is within the encodable state set and is in the ith bin of the quantizer. The source coding output at the plant sensor will represent this state as q_i and send the ith index over the channel. After a joint mapping of the channel and the channel decoder, the controller will receive the transmitted index i as index j with probability $p(j|i)$. The controller will apply its control over

index j, computing Q'_j. Thus, the controller decoder, the controller, and the encoder can be regarded as a single (composite) mapping. The controller transmits the control signal through the reverse channel to the plant which would interpret this value as Q'_l with probability $p'(l|j)$, by a mapping through the reverse channel. Given that the state is in the ith bin, the plant will receive the control Q'_l with probability $\sum_j p'(l|j)p(j|i)$. Thus, the applied control will be $u'_t = Q'_l$ with probability $\sum_j p'(l|j)p(j|i)$, the probability of the state being in the ith bin being $p(i) = \mathrm{Prob}(x \in \mathcal{B}_i)$. For CMCs, however, we do not use a quantizer; we denote the joint channel encoder and the channel as a stochastic kernel, $p(A|x)$, for $x \in \mathbb{R}$, and $A \in \mathcal{B}(\mathbb{R})$. The control is a deterministic function of the channel decoder, mapping $\mathbb{R}$ to $\mathbb{R}$. The control signals are sent back to the plant, via a reverse channel encoder and a reverse channel. Upon the arrival of the reverse channel output, the plant decoder obtains the decoded control signals, $u' \in \mathbb{R}$.

A. Stability of Markov Chains

Here, we review some relevant aspects of Markov chains, and list a number of definitions useful for the development in the paper. Let $\{x_t, t \geq 0\}$ be a Markov chain with state space $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, and defined on a probability space $(\Omega, \mathcal{F}, \mathcal{P})$, where $\mathcal{B}(\mathbb{R})$ denotes the Borel σ–field on $\mathbb{R}$, Ω is the sample space, $\mathcal{F}$ a sigma field of subsets of Ω, and $\mathcal{P}$ a probability measure. For $x \in \mathbb{R}$ and $D \in \mathcal{B}(\mathbb{R})$, we let $P(x, D) := P(x_{t+1} \in D | x_t = x)$ denote the transition probability from x to D, that is the probability of the event $\{x_{t+1} \in D\}$ given that $x_t = x$. Hence, the probability of the event $\{x_{t+1} \in D\}$ for any t can be computed recursively by starting at $t = 0$, with $P(x_1 \in D) = \int_{\mathbb{R}} P(x_1 \in D | x_0 = x)P(dx_0)$, and iterating with a similar formula for $t = 1, 2, \ldots$. We now have a number of definitions. See [26] for further details.

Definition 3.1: For a Markov chain with transition probability defined as above, a probability measure π is *invariant* on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ if $\pi(D) = \int_X P(x, D)\pi(dx)$, $\quad \forall D \in \mathcal{B}(\mathbb{R})$. $\quad \diamond$

Definition 3.2: A Markov chain is μ–*irreducible*, if for any Borel set with $\mu(B) > 0$, and $\forall x \in \mathbb{R}$, there exists some integer $n > 0$, possibly depending on B and x, such that $P^n(x, B) > 0$, where $P^n(x, B)$ is the transition probability in n stages, that is $P(x_{t+n} \in B | x_t = x)$. $\quad \diamond$

Definition 3.3: A μ-irreducible Markov chain is *aperiodic* if for any $x \in \mathbb{R}$, and any Borel set B satisfying $\mu(B) > 0$, there exists $n_0 = n_0(x, B)$ such that $P^n(x, B) > 0$ for all $n \geq n_0$.

Definition 3.4: For $B \in \mathcal{B}(\mathbb{R})$, let $\eta_B = \sum_{t=0}^{\infty} 1_{\{x_t \in B\}}$ (with 1_E denoting the indicator function for some event E). The μ-irreducible Markov chain $\{x_t, t \geq 0\}$ is *Harris recurrent* if $\mathrm{Prob}(\eta_B = \infty) = 1$ for all Borel sets with $\mu(B) > 0$. $\quad \diamond$

An alternative characterization is through stopping times. Let us define: $\tau_C := \inf\{t > 0 : x_t \in C\}$.

Definition 3.5: A μ-irreducible Markov chain is *Harris recurrent* if $\mathrm{Prob}(\tau_B < \infty | x_0 = x) = 1$ for any $x \in \mathbb{R}$, and any Borel set B satisfying $\mu(B) > 0$. It is *positive Harris recurrent* if in addition there is an invariant probability measure π. Such a chain which is not recurrent is called *transient*.

We now define the notion of a *petite set*:

Definition 3.6: A set $C \subset \mathbb{R}$ is κ-petite on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ if for some distribution $\mathcal{T}$ on $\mathbb{N}$ (set of natural numbers), and some

nontrivial measure κ, $\sum_{n=0}^{\infty} P^n(x, B)\mathcal{T}(n) \geq \mu(B)$, $\quad \forall x \in C$, and $B \in \mathcal{B}(\mathbb{R})$. $\quad \diamond$

A Markov chain's stability can be characterized by drift conditions [26], [27]. Let the Markov chain be irreducible, aperiodic, and S be a petite set. Let $V(.) : \mathbb{R} \to \mathbb{R}_+$ be a positive-valued functional, and $b \in \mathbb{R}_+$. Consider the following inequality for some $\epsilon \geq 0$, and some function $n(\cdot) : \mathbb{R} \to \mathbb{Z}_+$:

$$\int_{\mathbb{R}} P^{n(x)}(x, dy)V(y) \leq (1 - \epsilon)^{(n(x))}V(x) - n(x) + b1_{\{x \in S\}}$$

$\forall x \in \mathbb{R}$. If the inequality holds with $\epsilon = 0$, then a finite invariant measure exists for the Markov chain. If $\epsilon > 0$, then the chain is exponentially ergodic.

B. Codes: Upper Bounds on Error Probabilities

In this subsection, we discuss reliability of codes. Although longer block codes improve channel reliability, when it comes to control applications, long delays and larger sampling periods are undesirable. There is therefore a tradeoff between having short codelengths and having reliable transmission, and the explicit dependence of error probability on the codelength is characterized by *error exponents* [5]. Let $\mathcal{C} = \{c_0, c_1, \ldots, c_{M-1}\}$ be a codebook of cardinality M, where each codeword is of length N. Let $p(y|c_m)$ be the conditional probability of y being received given that input to the channel is c_m. Suppose that the decoding rule is such that the mth codeword is the output if $p(y|c_m) > p(y|c_n)$, for all $n \neq m$. This rule corresponds to the Maximum Likelihood (ML) decoding [5]. The set of output symbols which would lead to a decoding of c_m is thus given by $R_m = \{y : p(y|c_m) > p(y|c_n), n \neq m\}$. It follows that, if $y \notin R_m$, for at least one $n \neq m$, say n', such that $p(y|c_{n'}) \geq p(y|c_m)$, we have $\sum_{n \neq m} (p(y|c_n)/p(y|c_m)) \geq 1$. The probability of error given that c_m is transmitted is given by

$$p_{e|m} = \sum_{y \notin R_m} p(y|c_m) = \sum_{n \neq m} \sum_{y \in R_n} p(y|c_m). \tag{5}$$

Let $s > 0$. It follows that

$$\sum_{y \in R_n} p(y|c_m)$$

$$\leq \sum_{y \in R_n} p(y|c_m)\left(\frac{p(y|c_n)}{p(y|c_m)}\right)^{s/(s+1)}$$

$$= \sum_{y \in R_n} (p(y|c_m))^{1/(s+1)}(p(y|c_n))^{s/(s+1)}$$

$$= \exp\left(\log\left[\sum_{y \in R_n} (p(y|c_m))^{1/(s+1)}(p(y|c_n))^{s/(s+1)}\right]\right)$$

$$=: \exp\left(-d(c_m, c_n; s)\right) \tag{6}$$

which defines the quantity $d(c_m, c_n; s)$ whose minimum over all codeword pairs and $s > 0$ is called the *minimum distance of a code*, as introduced in [37]. For $s = 1$, we have the Bhattacharyya distance between two symbols c_m and c_n, which we denote (by some abuse of notation) by $d(c_m, c_n)$; note that $d(c_m, c_n; 1) = d(c_m, c_n)$. The minimum of this quantity over all $c_m, c_n \in \mathcal{C}$, $m \neq n$, is called the *minimum Bhattacharyya distance of a code* $\mathcal{C}$ and is denoted by $d(\mathcal{C})$.

We let $E_L(N, R) := (1/N)d(\mathcal{C})$ denote the minimum Bhattacharrya distance for a codebook $\mathcal{C}$ with length N and rate R. The probability of error between two different codewords (i.e., $p(n|m), m \neq n; c_m, c_n \in \mathcal{C}$) can then be upper bounded using $E_L(N, R)$: $p(n|m) \leq e^{-NE_L(N,R)}$. We note that, for low coding rates, the minimum Bhattacharrya distance is closely related to the Gilbert bound; for details see [31], [43]. For rates which are not low, however, a more useful bound is the random coding exponent [14]. The difference is that in the random coding bound the exponent is strictly positive for rates below capacity. This may not be so for the Gilbert bound. With the random coding bound, under Maximum Likelihood decoding, we have the average probability of error, $p_e := \sum_m p(c_m)p(e|m)$, where $c_m \in \mathcal{C}$, to be bounded by $p_e \leq e^{-(N(E_R(R)-o(N)/N))}$ ([14, Chp. 5]). We note that the random coding exponent can be used also as a bound for the errors in the transmission of each individual message (and not only on the average), and the results can be applied for this bound at rates where the Gilbert bound is weaker.

IV. NECESSARY CONDITIONS

A. Conditions on Capacities

For the system under study, there is a relationship between the capacities of the two channels and the existence of an invariant probability measure, as given in Thm. 4.1.

Theorem 4.1: Consider the system described by (2), and let C_f and C_r denote, respectively, the forward and reverse channel capacities. Then, for the existence of an invariant probability measure with a finite second moment, we need $\min(C_f, C_r) \geq \log_2(|a|)$.

Proof: See Appendix A. $\diamond$

Hence, we observe that the reverse channel is as important as the forward channel.

B. Structural Conditions

We now present a result on the structure of the encoder, which has significant practical implications that will be elaborated on later in the paper.

Theorem 4.2: For a discrete-time linear system as in (2), with $|a| > 1$, with channel transitions forming an irreducible Markov chain, if either the encodable control set or the encodable state set is bounded, the Markov chain is transient.

Proof: See Appendix B. $\diamond$

We note that a related result has been presented by Nair and Evans in [29] on the unboundedness of the second moment of the state, when quantizers with bounded range are used.

Corollary 4.1: Consider a noisy linear system as in (2), which is open-loop unstable. Further, suppose that the noise process has a continuous probability distribution function with unbounded support and that there is no communication feedback. Such a system, controlled over discrete memoryless channels with finite capacity, cannot be made to have an invariant probability measure when fixed-length coding is used for the transmission of state and control signals. $\diamond$

The restriction alluded to in Thm. 4.2 above leads to significant complexity on encoding for control over a discrete noisy channel, for there needs to be a matching between the entire state space, which requires a countably infinite number of codewords, and a finite-symbol channel. Such a complexity does not arise, however, in a CMC, as we will see later. This now motivates us to introduce the following.

Definition 4.1: An open-loop unstable system is *escape-free* if all the state symbols are encoded in such a way that given $x \in X$, there exists in the quantizer a reconstruction level, q, such that $x < q$. $\diamond$

Let $\tau_{[-M,M]} := \inf\{t > 0 : |x_t| \in [-M, M]\}$ for $M > 0$. By Thm. 4.2, for the chain to be positive recurrent, i.e., to have $\text{Prob}(\tau_{[-M,M]} < \infty | x_0 = x) = 1, \forall x \in X$, an unstable system has to be escape-free. Such a condition is not required for a stable system, since such a system is always recurrent. A system controlled over a continuous channel can always be made escape-free, and if the capacity is sufficiently large then the system can be stabilized. We will observe that using a dynamic structure, *escape-freeness* can be assured by considering a side channel which can be either continuous or a discrete one with finite capacity that can transmit variable length codes through variable sampling.

V. ASYMPTOTIC STABILITY OVER DMCS IN THE ABSENCE OF SYSTEM NOISE

We first have the following result.

Theorem 5.1: Let $C \subset \mathbb{Z}$, $L < \infty$, and $\delta_i, \forall i \in \mathbb{Z}$, be the bin edges of a symmetric quantizer. Let $1_{\{i \in C\}}$ be the indicator function for $i \in C$. Then, for a discrete channel, if the following drift condition holds for some sufficiently small $\epsilon > 0$, and for all bins

$$\max\left(\left|\left(\sum_l \sum_j p(j|i)p'(l|j)(a\delta_i + bQ'_l)\right)\right|, \left|\left(\sum_l \sum_j p(j|i)p'(l|j)(a\delta_{i+1} + bQ'_l)\right)\right|\right)$$
$$< \delta_i - \epsilon + L1_{\{i \in C\}}. \tag{7}$$

C is a recurrent set, that is

$\sup_{x \in \cup_{i \in C} \mathcal{B}_i} E[\min(t > 0 : x_t \in \cup_{i \in C}\mathcal{B}_i)|x_0 = x] < \infty$.

Proof: See Appendix C. $\diamond$

We now study the case when the channels are discrete and noiseless. If the channels are noiseless, the above leads (with $\epsilon = 0, L = 0$) to Elia-Mitter's logarithmic quantizer [12]:

Corollary 5.1: Consider a symmetric quantizer at the sensor. Let the forward and the reverse channels be noiseless. To lead to a drift towards the origin, quantizer bin edges (on the positive real line) have to satisfy $\delta_{i+1} \leq (1 + 2/|a|)\delta_i$. $\diamond$

Before studying the stability conditions, however, we first note the following relationship between reliability and delay.

Let us fix the forward and reverse channel rates, $R_f = \log_2(|M_f|)/N_f$ and $R_r = \log_2(|M_r|)/N_r$. We penalize the codelengths in the forward and reverse channels by a possibly linear term in the sampling period. It then takes longer to send more bits, that is reliability competes with delay. The following theorem says that if the controller waits long enough, stability can be achieved. To separate out the difficulty that comes about due to the escape-freeness requirement, we consider here first a one-stage problem; the more general system and control setup will be considered subsequently.
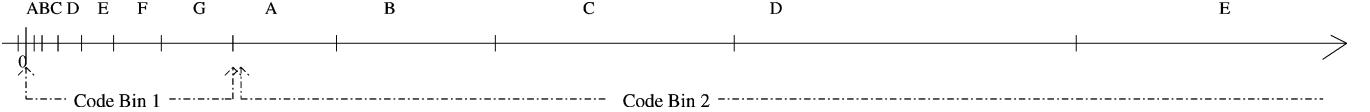
Fig. 2.   Illustration of the binning approach to the joint source channel code. The symbols in a given Codebin are represented by the same channel code, that is, letters such as $A, B, C, \ldots$. The mode symbols, such as $1, 2, 3, \ldots$, are carried by the side channel.

Theorem 5.2: Consider the scalar continuous-time system (1), but without the driving Brownian motion process. Let the probability distribution of the initial state x_0 have a bounded support set and have a finite second moment. Let the sampling period be a function of block lengths: $T_s = \alpha N_f + \beta N_r$; α, β be possibly depending on the codelengths, and the number of symbols in the state and control be $K = |\mathcal{X}'| = |\mathcal{U}| = |\mathcal{U}'|$. Let the rates $R_f = \ln(K)/N_f$ and $R_r = \ln(K)/N_r$ be kept constant as N_f and N_r grow. If the system and channel parameters satisfy the following three conditions,

$$\lim_{N_f \to \infty} (R_f + 2\mu\alpha - E_L^f(N_f, R_f)) + \left(\frac{2\mu\beta R_f}{R_r}\right) < 0,$$

$$\lim_{N_f \to \infty} (R_r + 2\mu\beta - E_L^r(N_r, R_r)) + \left(\frac{2\mu\alpha R_r}{R_f}\right) < 0,$$

$$\frac{\alpha}{R_f} + \frac{\beta}{R_r} < \frac{1}{\mu} \quad (8)$$

then $\lim_{T_s \to \infty} E[x_{T_s}^2] = 0$.

Proof: See Appendix D.   ◇

Let $a = e^{\mu T_s}$ as before. We have the following observations[1] regarding the result of Thm. 5.2 above: If there is no channel noise, the third condition of (8) is the well-studied quantization condition: $K > |a|$. Simple positivity of the error exponents (which is the case when rate is less than the capacity, $R < C$) does not directly lead to stability, and the exponent actually has to be larger than a specific positive quantity. Thus, a more accurate measure is the reliability of the channel, and not necessarily its capacity, as had been observed earlier in [35] and in part in [44]. As an illustration, consider binary symmetric channels with cross-over probabilities υ, for which the Gilbert exponent (which is equivalent to the Bhattacharyya exponent in this case) is $E_L(R) = H^{-1}(1 - R) \ln(2\sqrt{\upsilon(1-\upsilon)})$ for $R \leq \ln(2) - H(1/1 + e)$ [31] where H is the binary entropy function. In this case, a stabilizing rate region is given by the following:

$$\left\{ R_f, R_r : \frac{1}{N_f} \mu(\alpha N_f + \beta N_r) < R_f, \right.$$

$$\frac{R_f + 2\mu\alpha + 2\mu\beta \frac{R_f}{R_r}}{\ln(2\sqrt{\upsilon(1-\upsilon)})} < H^{-1}(1 - R_f),$$

$$\frac{1}{N_r} \mu(\alpha N_f + \beta N_r) < R_r,$$

$$\left. \frac{R_r + 2\mu\beta + 2\mu\alpha \frac{R_r}{R_f}}{\ln(2\sqrt{\upsilon(1-\upsilon)})} < H^{-1}(1 - R_r) \right\}.$$

[1][[4]]We note that one could obtain a converse bound using the sphere packing bound [14]. Sphere packing bound uses the maximum of the minimum distance discussed earlier over all codes used for a channel. This provides a lower bound on the probability of error for any transmitted signal, which can then readily be used in our setting leading to a converse result.

However, as mentioned above, if the control is applied more than once, the escape-freeness does not necessarily hold. Therefore, Thm. 5.2 should only be considered as a prelude to the more comprehensive analysis that will be carried out in the next section. Henceforth we will be using the bound obtained via the Gilbert exponent, but we note that it would be possible to obtain parallel results using the random coding exponent as well.

VI. ASYMPTOTIC STABILITY OVER DMCs IN THE PRESENCE OF SYSTEM NOISE

We now consider the sampled system (2) driven by i.i.d. noise, which is a more realistic scenario, where a finite sampling period is given, and the amount of data to be sent over a sampling period is finite. In this case, asymptotic analysis of Thm. 5.2 becomes inapplicable. We already know from Thm. 4.2 that for stability the encodable set has to be unbounded; however, the number of bits that can be transmitted per unit time over the channel is finite. To resolve this dichotomy, we will introduce a coding scheme based on binning [33], [50], where we will transmit the coset of the code and send the particular bin using an additional side channel.

Suppose that $K = 2^{N_f R_f}$ symbols can be transmitted during each time stage. Partition the entire state space into bins, group K adjacent elements into one larger bin, indexed by I, and represent them by a single channel codebook. We refer to this ensemble of bins as a *Codebin*. Hence, a total of $2^{N_f R_f}$ codewords are used to represent the entire state space (see Fig. 2). Thus, we have $\text{Codebin}(I) = \{x : \delta_{I N_f R_f} \leq x < \delta_{(I+1)N_f R_f}\}$. We denote the bin indexes by δ_{nI+i}, which means that the edge belongs to codebin I and is represented by the ith channel codeword. We say the source code is in mode I, if the state is in Codebin I (see Fig. 3). The reconstruction value of each bin is assumed to be its midpoint, so that $Q_i = (\delta_i + \delta_{i+1})/2$. We define $p^m(J|I)$ as the probability of error of Codebin (mode of the quantizer) transmission from mode I to mode J through the side channel. This means that the mode is erroneously transmitted from the plant to the controller if $J \neq I$. Likewise, for the reverse channel we have $p^{m'}(L|J)$ as the side channel mapping.

We will consider two schemes for the side channel. One is the case where the side channel is noisy but can transmit an infinite number of codewords, and thus leads to an escape-free system. In the second scheme, we consider a time-channel as the side channel, which is assumed to carry binary decisions and to be noiseless; it can, however, send only a limited number of bits per unit time. This scheme will be studied in the next section.

Before proceeding further, one observation is in order. Since the system noise is Gaussian, the Markov chain is Lebesgue-irreducible. Furthermore, if there is a compact set which is recurrent, then all open subsets of this compact set are also recurrent. Now, even though the transition kernel is not continuous for all
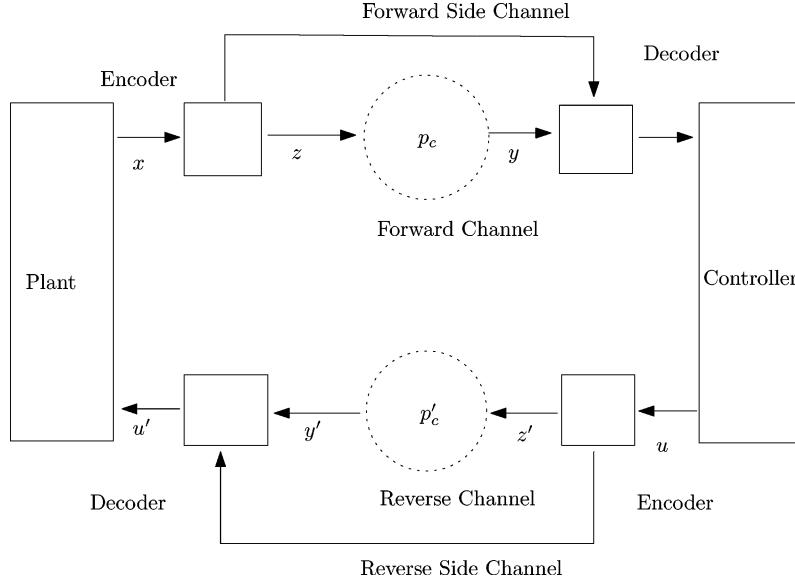
Fig. 3. Side channels can be regarded as separate uses of the forward and the reverse channels sending the modes.

$x \in \mathbb{R}$, for a small set $C \in \mathcal{B}(\mathbb{R})$ in one of the bins close to the origin, $P(x_{t+1} \in A | x_t = x)$, $A \in \mathcal{B}(\mathbb{R})$, will be uniformly continuous in $x \in C$, and bounded from below by some non-trivial measure. As such, such a set can serve as a petite set. With this observation, we have the following result.

Theorem 6.1: Consider the scalar continuous-time system described by (1), which is to be remotely controlled. Let T_s be a sampling period, which is a function of block lengths: $T_s = \alpha N_f + \beta N_r$, with α and β possibly depending on the code-lengths, and let the number of symbols in the state and the control be $K = |\mathcal{X}'| = |\mathcal{U}| = |\mathcal{U}'|$. Suppose that the forward and reverse channel codes are respectively of N_f and N_r bits long, and let the rates be $R_f = \log_2(K)/N_f$ and $R_r = \log_2(K)/N_r$. Let two variables be defined as follows:

$$\Xi(\gamma, I) := \sum_L \sum_J P^{m'}(L|J) P^m(J|I) \gamma^{2N_f R_f(|L-I|)}$$

$$\Upsilon := \left(e^{-N_f E_L^f(N_f, R_f) - N_r E_L^r(N_r, R_r)} 2^{N_f R_f} \right.$$
$$\left. + e^{-N_f E_L^f(N_f, R_f)} + e^{-N_r E_L^r(N_r, R_r)} \right) 2^{N_f R_f}. \quad (9)$$

For some $\gamma > 1$, with $\limsup_{I \to \infty} \Xi(\gamma, I) =: \Xi(\gamma)$, if the following condition is satisfied

$$\gamma < 1 + \sqrt{2} e^{-\mu(\alpha N_f + \beta N_r)} \sqrt{1 - \frac{4((1 + \gamma^{2N_f R_f} \Upsilon) \Xi(\gamma) - 1)}{e^{-2\mu(\alpha N_f + \beta N_r)}}}$$

then drift conditions hold and there exists a coding scheme leading to a finite second moment. A corresponding source coder is a symmetric logarithmic quantizer with expansion ratio γ, i.e., $|\delta_{i+1}| < \gamma |\delta_i|$.

Proof: See Appendix E. ◇

In a control problem, in general the controller is assumed to know the plant model and the dynamics. This information can be viewed as side information at the controller. For instance, if a dynamic system is being tracked, then what is known on the current state and the system model already provides some information on the distribution for the next state, without an explicit feedback. We now note the following.

Proposition 6.1: Suppose that the probability measure on the initial state has a finite support set. Further, suppose that the forward channel error, the system noise, and the reverse channel noise are all uniformly bounded such that there exists $\Delta < \infty$ such that, for a DMC

$$\sup_{(x,y) \in \mathcal{M} \times \mathcal{M}', p(y|x) > 0} |y - x| \le \Delta$$

and for a CMC

$$\sup_{(x,A) \in \mathcal{M} \times \mathcal{B}(\mathcal{M}'), p(A|x) = 1, p(\mathcal{M}' - A|x) = 0} \{ \sup_{y \in A} |y - x| \} \le \Delta.$$

Then, to achieve $\limsup_{t \to \infty} |x_t|^2 < \infty$, a.s., there is no need for a forward side channel.

Proof: Let Δ_0, Δ_f, Δ_s, and Δ_r be the lengths of the support intervals of the distributions for the initial state, forward channel error, systems noise, and the reverse channel error, respectively. Then, the total uncertainty at the first stage will be bounded by $\Delta =: |a|(\Delta_f + \Delta_0) + |b|\Delta_r + \Delta_s$. Clearly for large enough x, using a logarithmic quantizer, there exists a k_0 such that $\forall k > k_0$, $\gamma^{k N_r R_f} \delta_1 > 2\Delta$. Using binning techniques, there will be no error in distinguishing between two codewords with the same coset. In this case if the two nearest bins sharing the same coset are spread out with a distance greater than $|a|(\Delta_f + \Delta_0) + \Delta_s$, then with only the coset information, the controller can deduce the exact value of the bin. ◇

Unlike the transmissions from the controller, in general the plant cannot predict the control signal it will receive, since it does not have access to the decision policy at the controller. Therefore, such a relaxation does not apply to the reverse channel. We have observed that stability might be possible through the use of a side channel. However this side channel has to have infinitely many codewords! In the next section, we will get around this difficulty by a coding scheme which uses variable length coding through variable rate sampling.
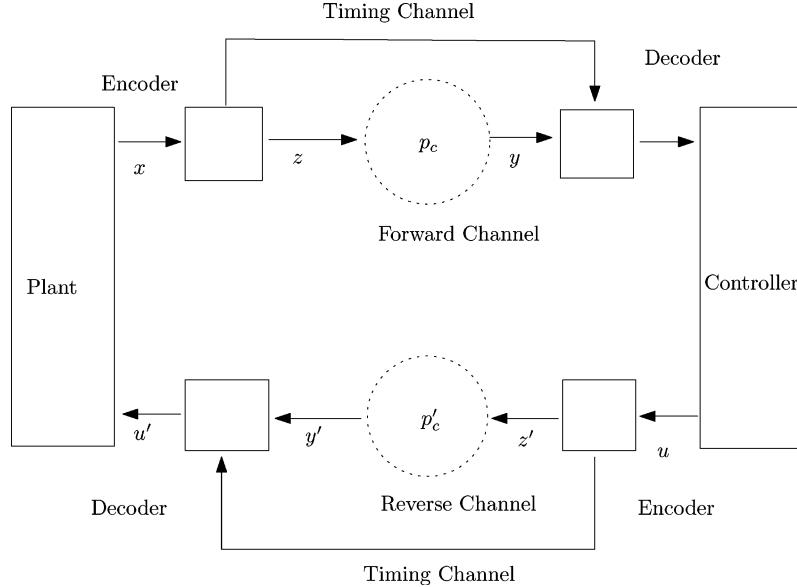
Fig. 4.   Variable encoding via a time-channel as side channel.

VII. VARIABLE LENGTH CODING THROUGH VARIABLE RATE SAMPLING

In practice, there is a strict bound on the actual number of channel uses per time. Markovian stability theory can be used to show that even with such a restriction, stability can be achieved. In case an invariant density exists, the occurrence of high magnitude signals will be rare. Building on this fact, we introduce below a variable length encoding scheme.

The controller and the sensor can send *side channel information*, i.e., the index information of the Codebin, over variable periods by using either explicit variable length codes or a *timing channel* (see Fig. 4). Timing channel is noiseless and carries the binary signal of starting the encoding, effectively carrying the index information. In such a scheme, Codebins are generated according to the number of sampling periods required to send the side channel information. Thus, the effective sampling period will vary. Here, the number of Codebins for a given period will grow exponentially with the sampling period. However, in this case the system will no longer be first-order Markovian.

Given state x, let $n(x)$ be the number of sampling periods it takes for the transmission of the side channel symbol (or timing instant). Note that $n(x)$ is a causally measurable stopping time, which is in fact deterministic. Consider the sampling process given by the dynamics: $T_k = T_{k-1} + n(x_{T_{k-1}})$. The sampled process, x_{T_k}, is also Markovian by the strong Markov property [30]. With this observation, we can obtain a state-dependent drift condition as discussed in Section III, to study stability. Here, we provide only conditions for the existence of an invariant measure; the approach for stronger conditions is identical and the extension is merely technical. If the transmission of the state and the control signals takes k times as long as it does in the fixed length case, the effective sampling period in the variable length encoding scheme is $kT_s, k \in \mathbb{Z}_+$. Thus, the system will be open loop during kT_s seconds. These considerations lead to the following counterpart of Thm. 6.1 in the case of variable-length encoding for side channels.

Theorem 7.1: Consider the scalar continuous-time system described by (1), which is to be remotely controlled over discrete channels. Let the forward and the reverse side channels be noiseless and the side channel symbols be transmitted over variable durations (as described above). Further, let $U_k(\gamma) := \gamma_k^{2(kN_f R_f)}$, and

$$\Upsilon(k) := \left(e^{-kN_f E_L^f(N_f, R_f) - kN_r E_L^r(N_r, R_r)} 2^{kN_f R_f} \right.$$
$$\left. + e^{-kN_f E_L^f(N_f, R_f)} + e^{-kN_r E_L^r(N_r, R_r)} \right) 2^{kN_f R_f}. \quad (10)$$

Suppose $\lim_{k \to \infty} (\Upsilon(k) U_k(\gamma)/e^{-2\mu k T_s}) < 1$. If for a sequence $\{\gamma_k \geq 1, k \in \mathbb{Z}_+\}$, $\exists k_0 > 0$ such that, $\forall k > k_0$, the following holds:

$$\gamma_k < 1 + 2e^{-k\mu(\alpha N_f + \beta N_r)} \sqrt{1 - \frac{\Upsilon(k) U_k(\gamma)}{e^{-2\mu k T_s}}},$$
$$N_f R_f > \log_2(|a|)$$

then there exists a coding scheme leading to positive Harris recurrence for the sampled Markov chain. The source coder is a symmetric logarithmic quantizer with sequentially decreasing expansion ratios γ_k used for symbols transmitted in the kth time stage.

Proof: See Appendix F.   ◇

In the preceding discussion, we required the time-channel to transmit noisefree the integer $k \in \mathbb{Z}_+$ (for kT_s as the open-loop duration during transmission) in k time stages. A prefix-free, uniquely decodable code can be used for the time-channel, such as, with 1 denoting the stopping bit, $1, 01, 001, 0001, \ldots, 0 \ldots 01, \ldots$. Hence, it suffices to send one bit of information to stop decoding. One needs an arbitrarily small, but non-zero zero-error capacity, since for every zero-error capacity of ϵ, there exists a sampling period T_s such that $1/T_s < \epsilon$ and reliable transmission of one bit per time-stage is possible. One can adjust the petite set as a function

of T_s. We note that in Theorem 7.1, the conditions simplify for noiseless channels, since the error probabilities are zero.

If the goal is the existence of a finite moment, more stringent criteria will be needed. Nonetheless, the same techniques can be applied. We also note that the coding construction can be arbitrary, as long as the drift conditions are satisfied.

VIII. CONTINUOUS-ALPHABET CHANNELS

In this section, we consider CMCs and obtain achievable rates for control over continuous alphabet channels. For continuous alphabet channels, there is no restriction on the values transmitted over the channel; there may be constraints over the input distributions, but arbitrary values can be fed into the channel. Thus, the system can always be designed to be escape-free. Here, we will study two special channels: Gaussian and continuous erasure channels.

A. Stabilizing Rates Over Gaussian Channels

For Gaussian channels, we associate power constraints with the encoder outputs, P_f and P_r, for the forward and the reverse channels, respectively. Our objective is to develop coding schemes that will lead to an invariant density with a finite second moment. Here, one does not face the difficulty of explicitly using a finite codelength, for $\mathcal{X}'$ is the entire real line, and rare events are transmitted with higher magnitude signals, whose contribution to the expected power is limited. This was the main difficulty we observed in DMCs in the design of variable length codes in a control context. There, rare events had to be represented in longer codewords to prevent the Markov chain from becoming transient.

In the following theorem, we restrict the encoders, the sensor, and the controller to be scaling their inputs. We further restrict the controller and the decoders to be linear in their arguments, and obtain the decoder and the controller that minimize the invariant second moment of the state. Our aim is to obtain a stabilizing configuration.

Theorem 8.1: Suppose that the sensor encoder and the controller encoder have average power constraints P_f and P_r, respectively. Further suppose that the encoders and decoders, and finally the controller itself, are restricted to be linear and memoryless. Then, the optimal such policy at the input of the plant, which minimizes the steady state variance is

$$ u'_t = -\left(\frac{a}{b}\right)\sqrt{P_f P_r} \left[\frac{\|x_t\|}{\left[\sqrt{P_f + \sigma_w^{f\,2}}(P_r + \sigma_w^{r\,2})\right]} \right] y'_t $$

where $\|x_t\| = \sqrt{E[x_t^2]}$, $y'_t = z'_t + w^r_t$, and $\sigma_w^{f\,2}$ and $\sigma_w^{r\,2}$ are the channel noise variances for the forward and the reverse channels, respectively. If the forward and the reverse channel capacities satisfy

$$ 2^{-2C_f} + 2^{-2C_r} - 2^{-2C_f - 2C_r} < 1/a^2, $$

then the steady state variance is finite.

Proof: See Appendix G. ◇

Observe that the lower bound on the capacities is $\log_2(|a|)$. This leads us to the following Corollary to Theorem 8.1:
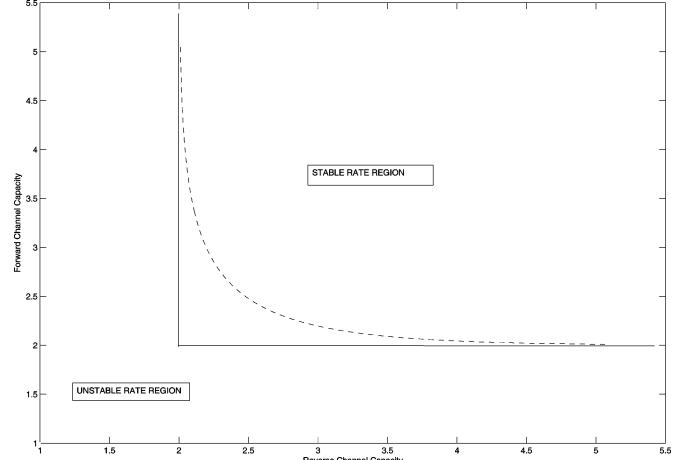

Fig. 5. For the case when $a = 4$, the area above the dashed curve contains stabilizing capacity pairs under linear memoryless policies, the capacity pairs outside the perpendicular edges are not stabilizing under any memoryless scheme.

Corollary 8.1: As C_f (respectively, C_r) $\to \infty$, the condition on C_r (respectively, C_f) becomes $C_r > \log_2(|a|)$ (respectively, $C_f > \log_2(|a|)$). Hence, if either the forward channel or the reverse channel is noiseless, then memoryless coders are as good as any other coder for the minimization of rate (power) requirements for there to be an invariant distribution for the state process. ◇

Hence, linear policies are almost optimal as they meet the lower bound, when one of the channels becomes very reliable. It is known that when noiseless feedback is available, innovation coders are optimal. The Corollary above shows that, if the goal is to have stability, memoryless schemes might as well be used.

[42] studied Gaussian channels in the context of relaxation of the "equimemory" condition. There, control is not encoded and by optimizing over the controller gain it is proven that the rate in the reverse channel does not improve the performance beyond a threshold value. In our case, however, the control is encoded as well, and an increment in power strictly improves the rate restrictions for the forward channel. Note that if one encodes the control, the problem is one of information transmission. If there is no encoding, then control is applied as it is received and there exists an optimal value for the control power. We note that [2] showed that for a one-shot problem, the analysis above leads to an optimal solution.

We plot the achievable (sufficient) and the necessary rate regions in Fig. 5. We observe that use of memoryless policies is actually not very inefficient, since the rate region is not too far off from the lower bounds; especially when one of the channels becomes more reliable, linear policies are almost tight, as can be deduced analytically.

B. Continuous Erasure Channels

Consider forward and reverse erasure channels which lose packets with probabilities p_f and p_r, respectively. Consider also the case where the packets can be sent without a need of quantization, i.e., the erasure channel codebook set is the real line (thus the information theoretic capacity is infinite). The consideration of such an infinite capacity erasure channel model has

been common in the network control literature [17]. In this case we have the following result.

Proposition 8.1: Consider the unstable plant in (2), along with forward and reverse erasure channels. If the forward and the reverse channel packet loss probabilities satisfy the inequality $p_f + p_r - p_f p_r < 1/a^2$, then $\lim_{t\to\infty} E[x_t^2] < \infty$.

Proof: We will have with probability $(1 - p_f)(1 - p_r)$, $x_{t+1} = ax_t - b(a/b)x_t + d_t$. If there is loss in the forward channel as well as in the reverse channel, zero control is applied (as a consequence of the memoryless policy). Then the evolution of the second moment will be $E[x_{t+1}^2] = [1 - (1 - p_f)(1 - p_r)]a^2 E[x_t^2] + E[d_t^2]$, which is clearly stable if and only if the condition in the statement of the proposition holds.   ◇

IX. CONCLUDING REMARKS

In this paper we have analyzed networked control systems over noisy channels, where the system noise can have unbounded support, both the forward and the reverse channels can be noisy, and they can be discrete or continuous.

To recap the main messages and contributions of the paper: We have obtained capacity bounds and achievable forward and reverse rate regions leading to a positive recurrent Markov chain. We observed that control over discrete channels requires an intricate design. We showed the inadequacy of fixed-length encoding schemes and introduced the notion of escape-freeness, which required the design to use variable length encoding. Our design stabilizes the system via variable rate sampling and uses properties of the sampled Markov chain. We observed that continuous alphabet channels are simpler to analyze for they can always be designed to be escape-free, without resorting to variable length sampling. However, since continuous channels are not as widely used as discrete channels, we view the analysis on DMCs to be important in applications.

The stochastic drift approach presented in this paper seems to admit generalizations to many problems where the transmission of the control or sensor signals has a randomness in it. One example of such an approach is reported in [49], for proving stochastic stability of adaptive encoders for Markov sources where random stopping times are the instances when the encoder can transmit information to a controller. It seems that such a random drift approach is a promising way, as variable-length decoding schemes also use stopping time arguments [8] for optimal performance and control over communication channels [35]. Thus, one future direction of research would be to study a combination of Bhattacharyya distance arguments and Burnashev's variable rate coding scheme with very little feedback for use in the reverse channel.

APPENDIX A
PROOF OF THEOREM 4.1

Let P_t be the probability measure at time t, generated as a result of the Markovian recursions $P_t(D) = \int_X P(x, D) P_{t-1}(dx)$, $\forall D \in \mathcal{B}(\mathbb{R})$, $t \geq 1$. Suppose that we have an invariant probability measure P for the Markovian process, with a finite second moment. Then, the entropy of the invariant density is also finite (which is bounded by the entropy of the Gaussian density with the same second moment).

Furthermore, by the irreducibility of the Markovian process, there is a unique invariant probability measure.

Since the system is driven by a Gaussian noise process, it follows that P_t admits a density, which we will refer as p_t, and which can be expressed as a convolution of the Gaussian measure with another probability measure. That is

$$p_t(x) = \int_{\mathbb{R}} \nu(x - z)\tilde{p}_{t-1}(z)dz$$

where $\tilde{p}_{t-1}$ is a piecewise continuous density function and ν is the Gaussian density function. The density function $\tilde{p}_{t-1}$ is piecewise continuous due to the effect of quantization and channel errors. It is thus a simple exercise to show that $p_t(x)$ is uniformly continuous and is such that, for any open set $B \subset \mathbb{R}$, $P_t(x \in B) > 0$. This implies that P_t is absolutely continuous with respect to P (which in turn is absolutely continuous with respect to the Lebesgue measure). Let p be the density corresponding to P.

This consequently implies that the (Kullback-Leibler) divergence $D(p_t\|p) < \infty$, for some $t > 1$, where for two probability density functions S, Q on $\mathbb{R}$, the divergence is defined by the relation $D(S\|Q) = \int_{\mathbb{R}} S(x) \log_2(S(x)/Q(x))dx$.

These lead to the conclusion that, by Theorem 4 of Harremoës and Holst [16] (see also [4]), $\lim_{t\to\infty} D(p_t\|p) = 0$. Furthermore, by the f–norm Ergodic Theorem ([26, Chapter 14]), $\int p_t(x) \log_2(p(x))dx \to \int p(x) \log_2(p(x))dx$. Therefore, $\int p_t(x) \log_2(p_t(x))dx$ converges to $\int p(x) \log_2(p(x))dx$. Hence, $\lim_{t\to\infty} h(x_t)$ exists and is finite.

Since $x_{t+1} = ax_t + bu_t' + d_t$, conditioning does not increase entropy, and $\{d_t\}$ is independent

$$
\begin{aligned}
h(x_{t+1}) &\geq h(x_{t+1}|u_t') = h\left(a\left(x_t - \left(\frac{b}{a}\right)u_t'\right) + d_t|u_t'\right) \\
&= h(ax_t + d_t|u_t') \geq h(ax_t + d_t|u_t', d_t) \\
&= h(ax_t|u_t') = \log_2(|a|) + h(x_t|u_t') \quad (11)
\end{aligned}
$$

which implies $h(x_{t+1}) - h(x_t|u_t') \geq \log_2(|a|)$. Since $I(x_t; u_t') = h(x_t) - h(x_t|u_t')$, we have $I(x_t; u_t') \geq h(x_t) + \log_2(|a|) - h(x_{t+1})$. For $\lim_{t\to\infty} h(x_{t+1}) - h(x_t) = 0$, we have

$$\liminf_{t\to\infty} I(x_t; u_t') \geq \log_2(|a|).$$

We now study the asymptotic behavior of the mutual information $I(x_t; u_t') = h(x_t) + h(u_t') - h(x_t, u_t')$. There exists a limit for $h(x_t)$ from the discussion above. There exists an invariant limit density for u_t' with a finite second moment, since the control is a result of a stationary stochastic kernel. Thus, $h(x_t, u_t')$ is the joint entropy process of a stable stationary pair, and thus has a limit with the arguments above. Hence, $I(x_t; u_t')$ has a limit, and $\lim_{t\to\infty} I(x_t; u_t') \geq \log_2(|a|)$. Therefore, for some arbitrary stationary distribution, the mutual information is lower bounded by $\log_2(|a|)$. From the general alphabet version of the data processing inequality ([15], Chapter 5), we have $I(x_t; u_t') \leq I(z_t; y_t)$ and $I(x_t; u_t') \leq I(z_t'; y_t')$.

Since capacities of memoryless channels are achieved by the supremizing source distributions, $C_f = \sup_{p(z_t)} I(z_t; y_t)$, $C_r = \sup_{p(z_t')} I(z_t'; y_t')$, and the

capacities of the individual channels upper bound the joint capacity, we have $\min(C_f, C_r) \geq \log_2(|a|)$. ⋄

APPENDIX B
PROOF OF THEOREM 4.2

Without any loss of generality, let the bounded encodable control set be given by $S_e = \{u : |u| < M/b'\}$. Since the chain is irreducible, all the open sets are irreducible ([26], Chapter 8). Hence, $\mathrm{Prob}(x_t \in T_k | x_0) > 0$. Due to this observation, the estimation error of the controller, $e_t := x_t - E[x_t | I_{t-1}^c]$ (where I_{t-1}^c is the information at the controller at time $t-1$), with regard to the state will be visiting T_k in finite time with probability one. Now suppose that the initial state is in a set $T_k := [R, 2^k R]$.

We study the exit time of the process x_t from T_k. We have $|b'u_t'| < M, \forall t \geq 0$. Define a process $dv_t = \gamma v_t dt + dB_t$, with $v_0 = x_0 \in T_k$, where $x_0 > R > M/(\mu - \gamma)$, and $\mu > \gamma > 0$ and B_t having the same sample path as the disturbance in the original system. Let $\tau^N := \min(\inf\{t : x_t \leq R\}, N)$, and $\tau' := \min(\inf\{t : v_t \leq R\}, N)$. Further, define $\tau_k'^N := \min(\inf\{t : v_t \notin T_k\}, N)$, and further define $e_t := x_t - v_t$. Since $(\mu x_t + b'u_t') > \gamma v_t$, for $0 \leq t \leq \min(\tau_k^N, \tau_k'^N)$, e_t is almost surely positive, for $0 \leq t \leq \min(\tau^N, \tau'^N)$. Since $x_t > v_t$ almost surely, the exit time $\tau_k'^N \leq \tau_k^N$ almost surely. Now, let $f(x) = e^{-2\gamma x}$. Note that $f(x)$ is continuously differentiable and bounded over the set of interest. Hence, we can apply Dynkin's formula [30], from which it follows that

$$E_{x_0}[f(v_{\tau_k'^N})] = f(x_0) + E_{x_0}\left[\int_0^{\tau_k'^N} Af(v_s)ds\right]$$

where A is the generator function [30], given by $Af(x) = \gamma f_x + (1/2)f_{xx}$, where f_x denotes the partial derivative of f with respect to x, and f_{xx} is its second partial derivative. Since $A(e^{-2\gamma x}) = 0$, and $E[\tau_k'^N] < \infty$, we have the expectation $E_{x_0}[f(v_{\tau_k'^N})] = f(x_0)$. With $N \to \infty$, $\tau_k'^N \to \tau_k'$ and by the Monotone Convergence Theorem, it follows that $E_{x_0}[f(v_{\tau_k'})] = f(x_0)$. Hence, $p_R e^{-2\gamma R} + (1 - p_R)e^{-\gamma 2^{k+1}R} = e^{-2\gamma x_0}$. We have that p_R is bounded in k, and $\gamma > 0$. Hence, it follows that $\lim_{k\to\infty} p_R(k) = e^{-2\gamma x_0}/e^{-2\gamma R} < 1$. Thus, the probability of the exit time for the newly defined process satisfies $p(\tau' < \infty) < 1$. Hence, $\mathrm{Prob}(\tau' < \infty) < 1$, and $\mathrm{Prob}(\tau < \infty) < 1$. Thus, the Markov chain is transient. Due to the open-loop instability of the dynamics, the return time of the process to a compact set around the origin has probability less than one. The proof for the encodable state set follows from similar arguments. ⋄

APPENDIX C
PROOF OF THEOREM 5.1

The expectation $\sum_l \sum_j p(j|i)p'(l|j)[ax + bQ_l']$ is linear in x; therefore, the maximum value of $|\sum_l \sum_j p(j|i)p'(l|j)[ax + bQ_l']|$ for $x \in [\delta_i, \delta_{i+1}]$ is achieved at one of the end points of each quantization bin (as $|f(x)|$ is a convex function, if $f(x)$ is linear). Thus, by ensuring the drift condition for the bin edges, a uniform decrease in the Lyapunov value for all x will be attained.

Let $V(x) = |x|$. The result of the theorem follows from the construction of a Supermartingale sequence as follows: Let $M_0 = V(x_0)$ and for $n \geq 1$, $M_n = V(x_n) + \sum_{k=0}^{n-1}(\epsilon - L1_{\{x_k \in \cup_{i \in C} \mathcal{B}_i\}})$. The sequence $\{M_n\}$ forms a Supermartingale: For any finite n, $E[M_{n+1}|\sigma(x_1, x_2, \ldots, x_n)] \leq M_n$. Let $\tau = \min(k > 0 : x_k \in \cup_{i \in C} \mathcal{B}_i)$. Let for $n \in \mathbb{Z}$, $\tau^n = \min(\tau, \min(t > 0 : t + V(x_t) \geq n))$ be a stopping time. Then, $E[M_{\tau^n}|x_0 = x] \leq M_0$, and hence $(\sup_{x \in \cup_{i \in C} \mathcal{B}_i} |x|)E[\tau^n|x_0 = x] \leq (M_0 + L)/\epsilon$, and by the Monotone Convergence Theorem, it follows that $\sup_{x \in \cup_{i \in C} \mathcal{B}_i} E[\tau|x_0 = x] < \infty$. ⋄

APPENDIX D
PROOF OF THEOREM 5.2

Suppose that the continuous-time system is sampled with a period T_s, to lead to: $x_{t+1} = ax_t + bu_t'$, where $a = e^{\mu T_s}$ and b is the sampled data control coefficient, as introduced earlier.

Let $Q_i' = E[x|x \in [\delta_i, \delta_{i+1}]]$, which is the center of moment of the corresponding bin (centroid). Let $\pi_0(x)$ be the probability measure on the initial state. Then, the distortion is given by

$$D = \sum_{l=1}^K \sum_{j=1}^K \sum_{i=1}^K p(j|i)p'(l|j) \int_{\delta_i}^{\delta_{i+1}} a^2 \left(x - \left(\frac{b}{a}\right)Q_l'\right)^2 \pi_0(dx)$$

which can be upper bounded as a function of the codelengths in the forward and reverse channels

$$\begin{aligned}
D \leq & \ e^{(2\mu\alpha - E_L^f(N_f, R_f))N_f} e^{(2\mu\beta - E_L^r(N_r, R_r))N_r} \\
& \times \sum_{l=1, l \neq j}^K \sum_{j=1, j \neq i}^K \sum_{i=1}^K \int_{\delta_i}^{\delta_{i+1}} \left(x - \left(\frac{b}{a}\right)Q_l'\right)^2 \pi_0(dx) \\
& + e^{(2\mu\alpha)N_f} e^{(2\mu\beta - E_L^r(N_r, R_r))N_r} \\
& \times \sum_{i=1}^K \sum_{l=1, l \neq i}^K \int_{\delta_i}^{\delta_{i+1}} \left(x - \left(\frac{b}{a}\right)Q_l'\right)^2 \pi_0(dx) \\
& + e^{(2\mu\alpha - E_L^f(N_f, R_f))N_f} e^{(2\mu\beta N_r)} \\
& \times \sum_{j=1}^K \sum_{i=1, i \neq j}^K \int_{\delta_i}^{\delta_{i+1}} \left(x - \left(\frac{b}{a}\right)Q_j'\right)^2 \pi_0(dx) \\
& + (e^{2\mu})^{\alpha N_f + \beta N_r} \sum_{i=1}^K \int_{\delta_i}^{\delta_{i+1}} \left(x - \left(\frac{b}{a}\right)Q_i'\right)^2 \pi_0(dx). \quad (12)
\end{aligned}$$

The conditions $(R_f + 2\mu\alpha - E_L^f(N_f, R_f)) < 0$ and $(R_r + 2\mu\beta - E_L^r(N_r, R_r)) < 0$ guarantee the convergence of the first term above. Note that the last term in (12) is just the quantization error, and using asymptotic quantization theory [51], the distortion is inversely proportional with the square of the number of symbols under a Lloyd-Max quantizer, and this needs to compensate the growth in T_s, which is satisfied by the third condition in (8). We note here that the third condition in (8) can also be expressed as $\lim_{N_f \to \infty}(K/e^{\mu(\alpha N_f + \beta N_r)}) > 1$.

What remains to be done is the analysis of the cross terms (second and the third terms in the summation). For the term

$$e^{(2\mu\alpha)N_f + (2\mu\beta - E_L^r(N_r, R_r))N_r}$$

$$\cdot \sum_{i=1}^K \sum_{l=1, l \neq i}^K \int_{\delta_i}^{\delta_{i+1}} \left(x - \left(\frac{b}{a}\right)Q_l'\right)^2 \pi_0(dx)$$

using the centroid property of $Q'_i, \forall i$, we obtain

$$e^{(2\mu\alpha)N_f + (2\mu\beta - E_L^r(N_r, R_r))N_r} \sum_{i=1}^{K} \sum_{l=1, l\neq j}^{K}$$

$$\int_{\delta_i}^{\delta_{i+1}} \left(\left(x - \left(\frac{b}{a}\right)Q'_i \right)^2 + \left(\left(\frac{b}{a}\right)Q_i - \left(\frac{b}{a}\right)Q'_l \right)^2 \right) \pi_0(dx)$$

for asymptotic boundedness of which it suffices to have $(2\mu\alpha)N_f + (R_r + 2\mu\beta - E_L^r(N_r, R_r))N_r < 0$. Likewise, for the other cross term, we need $(R_f + 2\mu\alpha - E_L^f(N_f, R_f))N_f + (2\mu\beta N_r) < 0$.

The proof for the random coding exponent follows similar steps. In this case, under Maximum Likelihood decoding, we obtain that

$$D = \sum_{l=1}^{K} \sum_{j=1}^{K} \sum_{i=1}^{K} p(j|i)p'(l|j) \int_{\delta_i}^{\delta_{i+1}} a^2 \left(x - \left(\frac{b}{a}\right)Q'_l \right)^2 \pi_0(dx)$$

$$\leq e^{-(N_f E_R^f(R_f) + N_r E_R^r(R_r))} a^2 M^2 + e^{-N_r E_R^r(R_r)} a^2 M^2$$

$$+ e^{-N_f E_R^f(R_f)} a^2 M^2$$

$$+ \sum_i \int_{\delta_i}^{\delta_{i+1}} a^2 \left(x - \left(\frac{b}{a}\right)Q'_i \right)^2 \pi_0(dx)$$

where M is the maximum value that $|x_0|$ can take. We observe that the stability condition is the stability of each of the individual terms above. This concludes the proof. $\diamond$

APPENDIX E
PROOF OF THEOREM 6.1

Suppose that the continuous-time system is sampled with a fixed sampling period T_s to lead to $x_{t+1} = ax_t + bu'_t + d_t$, where $a = e^{\mu T_s}$ and $b = \gamma T_s$ and d_t is zero-mean Gaussian with variance $E[d_t^2] = D(B) = (e^{2\mu T_s} - 1)/(2\mu)$. Using the drift condition, we need, for some $\epsilon > 0$, for some compact set S and with S^C denoting the region outside S, $E[x_{t+1}^2|x_t] < (1-\epsilon)x_t^2$, $\forall x \in S^C$. We first note that $E[x_{t+1}^2] = a^2 E[(x_t + (b/a)u'_t)^2] + D(B)$. Now, for a given state value, x, the expected distortion, $D(x)$, can be written as

$$\sum_L \sum_J P^{m'}(L|J)P^m(J|I)$$

$$\cdot \sum_{l=1}^{K} \sum_{j=1}^{K} p(j|i)p'(l|j)a^2 \left(x - \left(\frac{b}{a}\right)Q'_{l+NR_fL} \right)^2. \quad (13)$$

In the above expression, we can identify three different sources for error: one in case there is an error in the regular channel code with possible error in the side channels, one in case there is an error in the side channel given that there is no error in the channel code, and one in case there is an error in neither the side channels nor the channel codes. We also note that, using a simple bound via the parallellogram law

$$\left(x - \left(\frac{b}{a}\right)Q'_{i+NR_fL} \right)^2 \leq 2 \left(x - \left(\frac{b}{a}\right)Q'_{i+NR_fI} \right)^2$$

$$+ 2 \left(\left(\frac{b}{a}\right)Q'_{i+NR_fI} - \left(\frac{b}{a}\right)Q'_{i+NR_fL} \right)^2. \quad (14)$$

For a logarithmic quantizer with expansion ratio γ, $(\delta_{i+NR_fI})^2/(\delta_{j+NR_fL})^2 = \gamma^{|2(i-j+NR_f(I-L))|}$. Thus, for a transition from the ith bin to the lth bin, using the parallellogram law, the distortion will be less than $p(l|i)\sum_{L,J} P^{m'}(L|J)P^m(J|I)4\gamma^{2\max(i+NR_fI, l+NR_fL)+1}$. Now summing over all $l \neq i$, and using the lower bound on the error exponent, the above can be upper bounded by

$$4\Upsilon \sum_L \sum_J P^{m'}(L|J)P^m(J|I)\gamma^{2N_fR_f(|L-I|+1)}\delta_1^2 x^2$$

$$= 4\gamma^{2N_fR_f}\Xi(\gamma, I)\Upsilon x^2. \quad (15)$$

We have captured the distortion caused by channel errors and the side channel errors. For the case where there is only side channel error, the distortion divided by x^2 is upper-bounded by

$$\frac{2(\gamma-1)^2}{4} + 4 \sum_{L, L\neq I, J} P^{m'}(L|J)P^m(J|I)\gamma^{2NR_f(|L-I|)}$$

$$= 4(\Xi(\gamma, I) - 1) + \frac{(\gamma-1)^2}{2}. \quad (16)$$

We finally have the source coding error (when there is no error in any of the channels) as the following: $D_s = (x - (b/a)Q'_i)^2$ which can be upper-bounded by $(1/4)(\delta_{i+1} - \delta_i)^2$ or $x^2(\gamma - 1)^2/4$. We will upper bound this by $x^2(\gamma - 1)^2/2$ in the following to add the probability of events conveniently (as this occurs when either of the channels leads to an error). For some compact set S, if the following is satisfied $\forall x \in S^C$, then stability will be achieved

$$e^{2\mu T_s} \left(4\gamma^{2N_fR_f}\Xi(\gamma, I)\Upsilon + 4(\Xi(\gamma, I) - 1) + \frac{(\gamma-1)^2}{2} \right)$$

$$< \left((1-\epsilon) - \frac{e^{2\mu T_s} - 1}{2\mu x^2} \right).$$

The above is implied by the following for sufficiently large $i \in \mathbb{Z}_+$:

$$\frac{1}{2(\gamma-1)^2} < \left(\left((1-\epsilon) - \frac{e^{2\mu T_s} - 1}{2\mu\delta_i^2} \right)(e^{-2\mu})^{T_s} \right.$$

$$\left. - 4((1 + \gamma^{2N_fR_f}\Upsilon)\Xi(\gamma, I) - 1) \right)$$

and this completes the proof. $\diamond$

APPENDIX F
PROOF OF THEOREM 7.1

The proof is analogous to the one for Thm. 6.1. The difference is that the forward channels in the Codebins, are encoded here in k time stages. Also, here the transmissions in the side channels are assumed to be noise free. However, due to the increase in the effective sampling period, the sampled Brownian motion noise has a larger variance. Let C_k be the set of codebins, which are represented by codes that are transmitted over k transmissions. For the existence of an invariant distribution, it suffices that for some $\epsilon > 0$ and $k_0 > 0$

$$E[x_{t+k(x_t)T_s}^2|x_t] - x_t^2 < -k(x_t),$$

$$\forall x_t \in \text{Codebin}(I), \ I \in C_k, \ k \geq k_0. \quad (17)$$

Now, in our setting, we will have that there will be $2^{kN_f R_f}$ codewords which are transmitted in k sampling periods. Here, γ_k will be the logarithmic quantizer ratio for this set of codewords. Let us consider the following condition for some $\epsilon > 0$ and $k_0 > 0$

$$E[x_{t+kT_s}^2 | x_t] < (1-\epsilon)x_t^2 \qquad (18)$$

$\forall x \in \text{Codebin}(I), I \in C_k, k \geq k_0$. We now write $E[x_{t+kT_s}^2] = e^{2kT_s\mu}E[(x_t + (b/a)u_t')^2] + D_k(B)$, where $D_k(B) := (e^{2\mu kT_s} - 1)/2\mu$ is the variance of the standard Wiener process integrated over a period of kT_s. Thus, we need $e^{2\mu kT_s}E[(x_t + (b/a)u_t')^2] < [(1-\epsilon)x_t^2 - D_k(B)], \forall x \in S^C$. We will now bound the distortion term following the results of Thm. 5.2. The distortion is equal to

$$E\left[\left(x_t + \left(\frac{b}{a}\right)u_t'\right)^2\right]$$
$$= \sum_{l=1}^{K}\sum_{j=1}^{K} p(j|i)p'(l|j)e^{2kT_s\mu}\left(x - \left(\frac{b}{a}\right)Q_l'\right)^2.$$

The probability of error consists of the summation of the probability of errors in both channels, or in one of the channels. We can bound the probability of error by the following:

$$\Upsilon(k) = \left(e^{-kN_f E_L^f(R_f) - kN_r E_L^r(R_r)}2^{kN_f R_f}\right.$$
$$\left. + e^{-kN_f E_L^f(R_f)} + e^{-kN_r E_L^r(R_r)}\right)2^{kN_f R_f}.$$

For any i and l, the worst case distortion is upper bounded by $\gamma_k^{2(kN_f R_f)}x_t^2$, and the source coding error is upper-bounded by $(1/4)x_t^2(e^{2k\mu})^{T_s}(\gamma_k - 1)^2$. Hence if we have

$$e^{2\mu kT_s}\left(\Upsilon(k)x_t^2 + \frac{(\gamma_k - 1)^2 x_t^2}{4}\right) < (1-\epsilon)x_t^2 - D_k(B) \quad (19)$$

for $x_t \in C_k, k \geq k_0$, (18) will be satisfied. It now follows that (18) implies (17), since $\lim_{|x|\to\infty} k(x)/x = 0$. This follows since γ_k can be as large as

$$\left(1 + 2\sqrt{e^{-2\mu kT_s} - \Upsilon(k)U_k(\gamma)}\right)$$
$$= \left(1 + 2e^{-\mu kT_s}\sqrt{1 - \frac{\Upsilon(k)U_k(\gamma)}{e^{-2\mu kT_s}}}\right),$$

$\lim_{k\to\infty}(\Upsilon(k)U_k(\gamma)/e^{-2\mu kT_s}) < 1$, and that, as a result

$$\lim_{k\to\infty}\frac{(1 + 2\sqrt{e^{-2\mu kT_s} - \Upsilon(k)U_k(\gamma)})^{2^{kN_f R_f}}}{k} = \infty$$

when $N_f R_f > \log_2(|a|)$. This implies that $\lim_{|x|\to\infty} k(x)/x = 0$. ◇

APPENDIX G
PROOF OF THEOREM 8.1

We have

$$z_t = \frac{\sqrt{P_f}x_t}{\|x_t\|}, \quad y_t = z_t + w_t, \quad x_t' = \alpha_t y_t, u_t = \rho_t x_t',$$
$$z_t' = \frac{\sqrt{P_r}u_t}{\|u_t\|}, \quad y_t' = z_t' + w_t', \qquad u_t' = \beta_t y_t'.$$

The control applied, u_t', can be written as follows:

$$u_t' = \beta_t\sqrt{P_r}\left(\frac{\sqrt{P_f}x_t}{\left(\|x\|_t\sqrt{P_f + \sigma_w^{f^2}}\right)} + \frac{w_t}{\sqrt{P_f + \sigma_w^{f^2}}}\right) + \beta_t w_t'.$$

Using this in the system equation $x_{t+1} = ax_t + bu_t' + d_t$, we obtain

$$E[x_{t+1}^2] = E\left[\left(\left(a + \frac{b\beta_t\sqrt{P_r P_f}}{(\|x_t\|\sqrt{P_f + \sigma_{w^2}^f})}\right)x_t\right)^2\right]$$
$$+ \frac{b^2\beta_t^2\sigma_w^{f^2}P_r}{(P_f + \sigma_w^{f^2})} + b^2\beta_t^2\sigma_w^{r^2} + E[d_t^2]. \quad (20)$$

Define $\rho_t := b\beta_t$. The minimization of $E[x_{t+1}^2]$ with respect to ρ_t yields

$$\rho_t = -\frac{a\|x_t\|\sqrt{P_f P_r}}{\left((P_r + \sigma_w^{r^2})\sqrt{P_f + \sigma_w^{f^2}}\right)}. \qquad (21)$$

From this, the optimal expression for β_t follows. If we plug this optimal β_t value in the expression for $\|x_{t+1}\|^2 = E[x_{t+1}^2]$, we obtain

$$\|x_{t+1}\|^2 = \|x_t\|^2 a^2\left(1 - \frac{P_r P_f}{(P_f + \sigma_w^{f^2})(P_r + \sigma_w^{r^2})}\right)^2$$
$$+ \|x_t\|^2 a^2\left(\frac{\sigma_w^{f^2}P_r^2 P_f}{((P_f + \sigma_w^{f^2})^2(P_r + \sigma_w^{r^2})^2)}\right)$$
$$+ \frac{\|x_t\|^2 a^2 P_r P_f}{((P_f + \sigma_w^{f^2})(P_r + \sigma_w^{r^2})^2)}\sigma_w^{r^2}$$
$$+ E[d_t^2].$$

Upon recognizing the capacity expression in the following:

$$\frac{P_f}{(P_f + \sigma_w^{f^2})} = 1 - 2^{-2C_f}, \qquad \frac{P_r}{(P_r + \sigma_w^{r^2})} = 1 - 2^{-2C_r}$$

we arrive at

$$E[x_{t+1}^2]$$
$$= a^2\left\{(1 - 2^{-2C_f})(1 - 2^{-2C_r})(2^{-2C_f} - 2^{-2C_f - 2C_r} + 2^{-2C_r})\right.$$
$$\left. + \left(1 - (1 - 2^{-2C_f})(1 - 2^{-2C_r})\right)^2\right\}E[x_t^2] + E[d_t^2].$$

Rearranging the first term, it follows that the condition

$$1 - (1 - 2^{-2C_f})(1 - 2^{-2C_r}) < \frac{1}{a^2}$$

implies stability. ◇

an important missing term in one of the expressions in Theorem 5.2.

REFERENCES

[1] R. Bansal and T. Başar, "Simultaneous design of measurement and control strategies for stochastic systems with feedback," *Automatica*, vol. 25, no. 9, pp. 679–694, 1989.

[2] R. Bansal and T. Başar, "Solutions to a class of linear-quadratic-Gaussian LQG stochastic team problems with nonclassical information," *Syst. Contr. Lett.*, vol. 9, pp. 125–130, 1987.

[3] T. Başar and R. Bansal, "Optimum design of measurement channels and control policies for linear-quadratic stochastic systems," *European J. Oper. Res.*, vol. 73, pp. 226–236, 1994.

[4] A. R. Barron, "Limits of information, Markov chains, and projections," in *Proc. IEEE Int. Symp. Inf. Theory*, Sorrento, Italy, 2000, p. 25.

[5] R. E. Blahut, *Principles of Information Theory*. Reading, MA: Addison-Wesley, 1987.

[6] V. S. Borkar and S. K. Mitter, "LQG control with communication constraints," in *Kailath Festschrift*. Norwell, MA: Kluwer, 1997, pp. 365–373.

[7] J. H. Braslavsky, R. H. Middleton, and J. S. Freudenberg, "Feedback stabilization over signal-to-noise ratio constrained channels," *IEEE Trans Autom. Contr.*, vol. 52, pp. 1391–1403, 2007.

[8] M. V. Burnashev, "Data transmission over a discrete channel with feedback. Random transmission time," *Prob. Inf. Transm.*, vol. 12, pp. 10–30, 1976.

[9] C. D. Charalambous, A. Farhadi, S. Denic, and F. Rezaei, "Robust control over uncertain communication channels," in *Proc. IEEE Med. Conf. Contr. Autom.*, Cyprus, Jun. 2005, pp. 737–742.

[10] D. F. Delchamps, "Stabilizing a linear system with quantized state feedback," *IEEE Trans. Autom. Contr.*, vol. 35, no. 8, pp. 916–924, 1990.

[11] N. Elia, "When Bode meets Shannon: Control-oriented feedback communication schemes," *IEEE Trans. Autom. Contr.*, vol. 49, no. 9, pp. 1477–1488, Sep. 2004.

[12] N. Elia and S. K. Mitter, "Stabilization of linear systems with limited information," *IEEE Trans. Autom. Contr.*, vol. 46, no. 9, pp. 1384–1400, Sep. 2001.

[13] F. Fagnani and S. Zampieri, "Stability analysis and synthesis for scalar linear systems with a quantized feedback," *IEEE Trans. Autom. Contr.*, vol. 48, no. 9, pp. 1569–1584, Sep. 2003.

[14] R. G. Gallager, *Information Theory and Reliable Communication*. New York: Wiley, 1968.

[15] R. M. Gray, *Entropy and Information Theory*. New York: Springer Verlag, 1990.

[16] P. Harremoës and K. K. Holst, "Convergence of Markov chains in information divergence," *J. Theoret. Probab.*, vol. 22, pp. 186–202, 2009.

[17] O. C. Imer, S. Yüksel, and T. Başar, "Optimal control of LTI systems over communication networks," *Automatica*, vol. 42, no. 9, pp. 1429–1440, 2006.

[18] H. Ishii and T. Başar, "Remote control of LTI systems over networks with state quantization," *Syst. Contr. Lett.*, vol. 54, no. 1, pp. 15–32, 2005.

[19] H. Ishii and B. A. Francis, *Limited Data Rate in Control Systems with Networks*. Berlin, Germany: Springer, 2002, vol. 275, Lecture Notes in Control and Information Sciences.

[20] E. A. Jonckheere, A. A. Hammad, and B. F. Wu, "Chaotic disturbance rejection — A Kolmogorov-Sinai entropy approach," in *Proc. IEEE Conf. Decision Contr.*, San Antonio, TX, Dec. 1993.

[21] T. Kailath, "The divergence and Bhattacharyya distance measures in signal selection'," *IEEE Trans. Commun. Tech.*, vol. 15, pp. 52–60, Feb. 1967.

[22] D. Liberzon, "On stabilization of linear systems with limited information," *IEEE Trans. Autom. Contr.*, vol. 48, no. 2, pp. 304–307, Feb. 2003.

[23] A. Mahajan and D. Teneketzis, "On the design of globally optimal communication strategies for real-time noisy communication with noisy feedback," *IEEE J. Sel. Areas Commun.*, vol. 26, pp. 580–595, May 2008.

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

[25] A. S. Matveev and A. V. Savkin, *Estimation and Control over Communication Networks*. Boston, MA: Birkhäuser, 2008.

[26] S. P. Meyn and R. Tweedie, *Markov Chains and Stochastic Stability*. New York: Springer Verlag, 1993.

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

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

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

[30] B. Oksendal, *Stochastic Differential Equations*. Berlin, Germany: Springer, 2003.

[31] J. K. Omura, "Expurgated bounds, Bhattacharyya distance, and rate distortion function," *Inf. Contr.*, vol. 24, pp. 358–383, 1974.

[32] 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.

[33] S. S. Pradhan and K. Ramchandran, "Distributed source coding using syndromes (DISCUS): Design and construction," *IEEE Trans. Inf. Theory*, vol. 49, no. 3, pp. 626–643, Mar. 2003.

[34] A. Sahai, "Why do block length and delay behave differently if feedback is present?," *IEEE Trans. Inf. Theory*, vol. 54, no. 5, pp. 1860–1886, May 2008.

[35] 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.

[36] S. V. Sarma, M. A. Dahleh, and S. Salapaka, "On time-varying bit-allocation maintaining stability and performance: A convex parameterization," *IEEE Trans. Autom. Contr.*, vol. 53, no. 5, pp. 1147–1159, Jun. 2008.

[37] C. Shannon, R. G. Gallager, and E. R. Berlekamp, "Lower bounds to error probability for coding on discrete memoryless channels," *Inf. Contr.*, vol. 10, pp. 65–103, 1967.

[38] L. Schenato, "Optimal estimation in networked control systems subject to random delay and packet drop," *IEEE Trans. Autom. Contr.*, vol. 53, no. 6, pp. 1311–1317, Jun. 2008.

[39] 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.

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

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

[42] S. Tatikonda, A. Sahai, and S. Mitter, "LQG control problems under communication constraints," in *Proc. IEEE Conf. Decision Contr.*, Dec. 1998, pp. 1165–1170.

[43] A. J. Viterbi and J. K. Omura, *Principles of Digital Communication and Coding*. New York: McGraw-Hill, Inc., 1979.

[44] J. C. Walrand and P. Varaiya, "Optimal causal coding-decoding problems," *IEEE Trans. Inf. Theory*, vol. 19, pp. 814–820, 1983.

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

[46] W. S. Wong and R. W. Brockett, "Systems with finite communication bandwidth constraints – Part II: Stabilization with limited information feedback," *IEEE Trans. Autom. Contr.*, vol. 42, no. 9, pp. 1294–1299, 1997.

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

[48] S. Yüksel, "Optimal LQG coding and control over communication channels: An existence and an infeasibility result," in *Proc. IEEE Amer. Contr. Conf.*, Seattle, WA, Jun. 2008, pp. 3114–3119.

[49] S. Yüksel, "A random time stochastic drift result and application to stochastic stabilization over noisy channels," in *Proc. Annu. Allerton Conf.*, Oct. 2009, pp. 628–635.

[50] S. Yüksel and T. Başar, "Communication constraints for decentralized stabilizability with time-invariant policies," *IEEE Trans. Autom. Contr.*, vol. 52, no. 6, pp. 1060–1066, Jun. 2007.

[51] S. Yüksel and T. Başar, "Minimum rate coding for LTI systems over noiseless channels," *IEEE Trans. Autom. Contr.*, vol. 51, no. 12, pp. 1878–1887, Dec. 2006.

Serdar Yüksel (S'02) was born in Turkey in 1979. He received the B.Sc. degree in electrical and electronics engineering from Bilkent University, Turkey, in 2001, and the M.S. and Ph.D. degrees in electrical and computer engineering from the University of Illinois at Urbana-Champaign, in 2003 and 2006, respectively.

He was a Postdoctoral Researcher at Yale University, New Haven, CT, before joining Queen's University, London, U.K., as an Assistant Professor of mathematics and engineering at the Department of Mathematics and Statistics in 2007. His research interests are on stochastic control, decentralized control, information theory, source coding theory, and multi-terminal systems.

Dr. Yüksel serves on the IFAC (International Federation of Automatic Control) Committee on Stochastic Systems.

Tamer Başar (S'71–M'73–SM'79–F'83) received the B.S.E.E. degree from Robert College, Istanbul, in 1969, and the M.S., M.Phil, and Ph.D. degrees from Yale University, New Haven, CT, in 1970, 1971, and 1972, respectively.

He is currently with the University of Illinois at Urbana-Champaign (UIUC), where he holds the academic positions of Swanlund Endowed Chair, Center for Advanced Study Professor of Electrical and Computer Engineering, Research Professor at the Coordinated Science Laboratory, and Research Professor at the Information Trust Institute. He joined UIUC in 1981 after holding positions at Harvard University and Marmara Research Institute, Turkey. He is currently the Editor-in-Chief of *Automatica*, Editor of the Birkhäuser Series on *Systems & Control*, Managing Editor of the *Annals of the International Society of Dynamic Games* (ISDG), He has published extensively in systems, control, communications, and dynamic games, and has current research interests in modeling and control of communication networks; control over heterogeneous networks; estimation and control with limited sensing and transmission; resource allocation, management and pricing in networks; mobile and distributed computing; and security and trust in computer systems.

Dr. Başar received several the Medal of Science of Turkey (1993); Distinguished Member Award (1993), Axelby Outstanding Paper Award (1995), and Bode Lecture Prize (2004) of the IEEE Control Systems Society (CSS); Millennium Medal of IEEE (2000); Tau Beta Pi Drucker Eminent Faculty Award of UIUC (2004); the Outstanding Service Award (2005) and the Giorgio Quazza Medal (2005) of the International Federation of Automatic Control (IFAC); the Richard Bellman Control Heritage Award (2006) of the American Automatic Control Council (AACC); honorary doctorate (Doctor Honoris Causa) from Doğuş University (Istanbul; 2007); honorary professorship from Northeastern University (Shenyang; 2008); and Isaacs Award of ISDG (2010). He is a member of the U.S. National Academy of Engineering, a member of the European Academy of Sciences, a Fellow of IFAC, a past President of CSS, a past (founding) President of ISDG, and the current President of AACC.and member of editorial and advisory boards of several international journals in control, wireless networks, and applied mathematics.