STATIONARY AND ERGODIC PROPERTIES OF STOCHASTIC NONLINEAR SYSTEMS CONTROLLED OVER COMMUNICATION CHANNELS*

SERDAR YÜKSEL†

Abstract. This paper is concerned with the following problem: Given a stochastic nonlinear system controlled over a noisy channel, what is the largest class of channels for which there exist coding and control policies so that the closed-loop system is stochastically stable? Stochastic stability notions considered are stationarity, ergodicity or asymptotic mean stationarity. We do not restrict the state space to be compact, for example, systems considered can be driven by unbounded noise. Necessary and sufficient conditions are obtained for a large class of systems and channels. A generalization of Bode's integral formula for a large class of nonlinear systems and information channels is obtained. The findings generalize existing results for linear systems.

Key words. networked control, stochastic stability, ergodicity, nonlinear systems

AMS subject classifications. 93E03, 60J20

DOI. 10.1137/140989686

1. Introduction. Consider an N-dimensional controlled nonlinear system described by the discrete-time equations

$$(1.1) \qquad x_{t+1} = f(x_t, u_t, w_t)$$

for a (Borel measurable) function f, with $\{w_t\}$ being an independent and identically distributed (i.i.d.) system noise process with $w_t \sim \nu$.

This system is connected over a noisy channel with a finite capacity to a controller, as shown in Figure 1. The controller has access to the information it has received through the channel. A source coder maps the source symbols, state values, to corresponding channel inputs. The channel inputs are transmitted through a channel; we assume that the channel is a finite alphabet channel with input alphabet $\mathcal{M}$ and output alphabet $\mathcal{M}'$.

We refer by a *coding policy* Π a sequence of functions $\{\gamma_t^e, t \geq 0\}$ which are causal such that the channel input at time t, $q_t \in \mathcal{M}$, under Π^{comp} is generated by a function of its local information, that is,

$$q_t = \gamma_t^e(\mathcal{I}_t^e),$$

where $\mathcal{I}_t^e = \{x_{[0,t]}, q'_{[0,t-1]}\}$ and $q_t \in \mathcal{M}$, the channel input alphabet given by $\mathcal{M} := \{1, 2, \ldots, M\}$, for $0 \leq t \leq T - 1$. Here, we have the notation for $t \geq 1$: $x_{[0,t-1]} = \{x_s, 0 \leq s \leq t - 1\}$.

The channel maps q_t to q'_t in a stochastic fashion so that $P(q'_t | q_t, q_{[0,t-1]}, q'_{[0,t-1]})$ is a conditional probability measure on $\mathcal{M}'$ for all $t \in \mathbb{Z}_+$. If this expression is equal to $P(q'_t | q_t)$, the channel is said to be a memoryless channel, that is, the past variables

---

†Department of Mathematics and Statistics, Queen's University, Kingston, Ontario, K7L 3N6, Canada (yuksel@mast.queensu.ca).
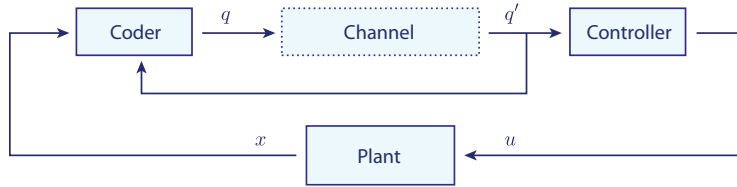
FIG. 1. *Control over a noisy channel with feedback.*

do not affect the channel output q'_t given the current channel input q_t. Even though in this paper we will consider discrete alphabet channels, the analysis is also applicable to a large class of continuous alphabet channels (through an appropriate quantized approximation of the channel; see, e.g., [14]).

The receiver/controller, upon receiving the information from the channel, generates its decision at time t, also causally: An admissible causal controller policy is a sequence of functions $\gamma = \{\gamma_t\}$ such that

$$\gamma_t : \mathcal{M}'^{t+1} \to \mathbb{R}^m, \qquad t \geq 0,$$

so that $u_t = \gamma_t(q'_{[0,t]})$. We call such encoding and control policies *causal* or *admissible*.

In the networked control literature, the goal in the encoder/controller design is typically either to optimize the system according to some performance criterion or stabilize the system. For stabilization, linear systems have been studied extensively where the goal has been to identify conditions so that the controlled state is stochastically stable, as we review briefly later.

This paper is concerned with necessary and sufficient conditions on information channels in a networked control system for which there exist coding and control policies such that the controlled system is stochastically stable in one or more of the following senses: (i) the state $\{x_t\}$ and the coding and control parameters lead to a stable (positive Harris recurrent) Markov chain, (ii) $\{x_t\}$ is asymptotically stationary or asymptotically mean stationary (AMS) and satisfies Birkhoff's sample path ergodic theorem (see section A for a review of these concepts), and (iii) $\{x_t\}$ is ergodic.

In the remainder of this section, we provide a literature review, first for nonlinear systems and then briefly for linear systems in the context of the goals of this paper, and highlight the contributions of the paper. Section 2 develops some supporting results and a generalization of Bode's integral formula for nonlinear systems and general information channels. Section 3 develops conditions for ergodicity and asymptotic mean stationarity of the controlled system. Section 4 establishes conditions for stationarity of the controlled system under structured (stationary) coding and control policies. Section 5 presents an ergodic construction for a nonlinear system driven by additive Gaussian noise and controlled over discrete noiseless channels. Section 6 contains some concluding remarks.

1.1. Some notation and preliminaries. Let x be an $\mathbb{X}$-valued random variable, where $\mathbb{X}$ is countable. The *entropy* of x is defined as $H(x) = -\sum_{z \in \mathbb{X}} p(z) \log_2(p(z))$, where p is the probability mass function of the random variable x. If x is an $\mathbb{R}^n$-valued random variable, and the probability measure induced by x is absolutely continuous with respect to the Lebesgue measure, the *(differential) entropy* of x is defined by $h(x) = -\int_{\mathbb{X}} p(x) \log_2(p(x)) dx$, where $p(\cdot)$ is the probability density function (pdf) of x.

The *mutual information* between a discrete (continuous) random variable x and another discrete (continuous) random variable y, defined on a common probability space, is defined as $I(x;y) = H(x) - H(x|y)$, where $H(x)$ is the entropy of x (differential entropy if x is a continuous random variable), and $H(x|y)$ is the conditional entropy of x given y ($h(x|y)$ is the conditional differential entropy if x is a continuous random variable). For more general settings including when the random variables are continuous, discrete, or a mixture of the two, mutual information is defined as $I(x;y) := \sup_{Q_1,Q_2} I(Q_1(x);Q_2(y))$, where Q_1 and Q_2 are quantizers with finitely many bins (see Chapter 5 in [16]). An important relevant result is the following. Let x be a random variable and Q be a quantizer applied to x. Then, $H(Q(x)) = I(x;Q(x)) = h(x) - h(x|Q(x))$. For a concise overview of relevant information theoretic concepts, we refer the reader to Chapter 5 of [67]. For more complete coverage, see [14] or [8]. When the realization x of a random variable x_t needs to be explicitly mentioned, the event $x_t = x$ will be emphasized. We use the conditional probability (expectation) notation $P_x(\cdot)$ ($E_x[\cdot]$)to denote $P(\cdot|x_0 = x)$ ($E[\cdot|x_0 = x]$). Finally, for a square matrix A, $|A|$ denotes the absolute value of its determinant.

Throughout the paper, all the random variables will be defined on a common probability space $(\Omega, \mathcal{F}, P)$.

1.2. Literature review. In the literature, the study of nonlinear systems has typically considered noise-free controlled systems controlled over discrete noiseless channels. Many of the studies on control of nonlinear systems over communication channels have focused on constructive schemes (and not on converse theorems), primarily for noise-free sources and channels; see, e.g., [2], [29], and [46]. For noise-free systems, it typically suffices to consider only a sufficiently small *invariant* neighborhood of an equilibrium point to obtain stabilizability conditions.

One important problem which has not yet been addressed to our knowledge is to obtain converse (or impossibility) theorems: The question of when an open-loop unstable nonlinear stochastic control system can or cannot be made ergodic or asymptotically mean stationary subject to information constraints has not been addressed.

Entropy based arguments (which are crucial in obtaining fundamental bounds in information theory and ergodic theory) can be used to obtain converse results: The entropy, as a measure of uncertainty growth, of a dynamical system has two related interpretations: A topological (distribution-free/geometric) one and a measure-theoretic (probabilistic) one. Although the analysis in this paper is probabilistic, we provide a short discussion on the topological entropy: The distribution-free entropy notion (see, e.g., [23]) for a dynamical system taking values in a compact metric space is concerned with the time-normalized number of distinguishable paths/orbits by some finite $\epsilon > 0$ the system's paths can take values in as the time horizon increases and $\epsilon \to 0$. With such a distribution-free setup [41] studied the stabilization of deterministic systems controlled over discrete noiseless finite capacity channels: The topological entropy gives a measure of the number of distinct control inputs needed to make a compact set invariant for a noise-free system. Nair et al. [41] extend the notion of topological entropy to controlled dynamical systems and develop the notion of feedback entropy or invariance entropy [7]; see also [6] for related results. Nair et al. [41] define two notions of invariance for a set K. A set can be made *weakly invariant* if there exists $t > 0$ such that for every $x_0 \in K$, there exists a sequence of control actions so that $x_t \in K' \subset \text{interior}(K)$. Strong invariance of K requires that $x_1 \in K'$. With a relaxation of deterministic controls, [52] has studied *invariance entropy* for random dynamical systems, and [35] has generalized the topological entropy theoretic

results to include random dynamical models to obtain an observability condition over discrete channels. For a comprehensive discussion of such a geometric interpretation of entropy in controlled systems, see [24]. The results for deterministic systems pose questions on set stability which are not sufficient to study stochastic setups. Stochasticity also allows for control over general noisy channels and thus is applicable to establish connections with information theory (we note that a distribution-free counterpart for such studies requires one to investigate zero-error capacity formulations [35]; however, many practical channels, including erasure channels, have zero zero-error capacity).

On the other hand, the measure-theoretic (also known as Kolmogorov–Sinai or metric) entropy is more relevant to information-theoretic as well as random noise-driven stochastic contexts since in this case, one considers the *typical* distinguishable paths/orbits of a dynamical system and not all of the sample paths a dynamical system may take (and hence the topological entropy typically provides upper bounds on the measure-theoretic entropy). Measure-theoretic entropy is crucial in the celebrated Shannon–McMillan–Breiman theorem [15] as well as the isomorphism theorem [44], [23]. For further relations between different interpretations of entropy as well as their computations (such as through *Lyapunov exponents* as a result of Pesin's formula), we refer the reader to [62]. Such an entropy notion has operational practical usage in identifying fundamental limits on source and channel coding for stationary sources [51]. However, the findings in the information theory literature have not yet been successfully applied to nonlinear networked control systems in general due to the following reasons: (i) The open-loop system in networked control may be unstable and stabilizable only through a control loop. In the information theory literature, stochastic stability results for coding schemes have been established primarily for (control-free) stable sources and, when nonstationary, have involved only linear Gaussian autoregressive processes [18]. Moreover, such a control-free analysis does not lead to conclusive results for nonlinear controlled sources since nonlinear systems suffer from the *dual-effect*: one cannot decouple estimation from control and control from conditional entropy properties under a stationary probability measure. (ii) The coding schemes for such studies in information theory are noncausal; in networked control systems, coding must be causal (that is, real-time or essentially zero-delay [67]).

There have been few studies which have adopted a measure-theoretic entropic view for the control of nonlinear dynamical systems over communication channels. Relevant contributions include [36] and [63]: Building on [31] and [69], [36] develops an entropy analysis for nonlinear system dynamics. A related entropy analysis for a class of stochastic nonlinear systems has been considered in [63]. Recently [59] and [58] have considered fading and erasure channels between the controller and the actuator and have studied ergodicity properties using Lyapunov theoretic arguments under a class of structures imposed on control policies; these contributions do not consider finite-rate information and coding restrictions which may arise due to the presence of a channel. Other important relevant works which consider deterministic systems are [29] and [28], where stability of zooming schemes, as in [4], have been considered.

Finally, we note an important related discussion in view of Bode's integral formula as extended to a class of nonlinear systems in [69] under somewhat restrictive conditions; see [69, Theorem 9]. Relevant work includes [12], [31], and [43] for linear systems. For nonlinear systems the entropy and mutual information arguments provide the appropriate fundamental bounds instead of a sensitivity integral/transfer function analysis which is commonly used for linear systems as is also advocated in

[69]. An earlier contribution utilizing measure theoretic entropy for the study and classification of controlled stochastic systems is [22]. The findings in our paper provide further generalizations; see Theorem 2.2 and Remark 4.

The stability criteria outlined earlier have been studied extensively for linear systems of the form

$$(1.2) \qquad\qquad x_{t+1} = Ax_t + Bu_t + Gw_t,$$

where $x_t \in \mathbb{R}^N$ is the state at time t, $u_t \in \mathbb{R}^m$ is the control input, and $\{w_t\}$ is a sequence of i.i.d. $\mathbb{R}^d$-valued random vectors (such as Gaussian). Here, (A, B) and (A, G) are controllable pairs.

For noise-free linear systems controlled over discrete-noiseless channels, Wong and Brockett [61], Baillieul [1], and, more generally, Tatikonda and Mitter [56] (see also [55]) and Nair and Evans [40] have obtained the minimum lower bound needed for stabilization over a class of communication channels under various assumptions on the system noise and channels, sometimes referred to as a *data-rate theorem*. This theorem states that for stabilizability under information constraints, in the mean-square sense, a minimum average rate per time stage needed for stabilizability has to be at least $\sum_{i:|\lambda_i|>1} \log_2(|\lambda_i|)$, where $\{\lambda_i, 1 \leq i \leq N\}$ are the eigenvalues of A.

The particular notion of stochastic stability is crucial in characterizing the conditions on the channels and important extensions have been made in the literature, notably by Matveev and Savkin [34], [35], Sahai and Mitter [48], [49], and Martins, Dahleh, and Elia [32]. For a more comprehensive review, see [42], Chapters 5–8 of [67], [31], and [13]. Reference [39] considered erasure channels and obtained necessary and sufficient time-varying rate conditions for control over such channels. Reference [9] considered second moment stability over a class of Markov channels with feedback. Motivated by such problems, [64] and [68] developed a martingale-method for establishing stochastic stability, which later led to a random-time state-dependent drift criterion, leading to the existence of an invariant distribution possibly with moment constraints; these were utilized to obtain policies leading to strong forms of stochastic stability, such as ergodicity or positive Harris recurrence [67], for linear systems driven by additive unbounded noise.

The following definition (see [67, Definition 8.5.1]) will be useful in the analysis later in the paper.

DEFINITION 1.1. *Channels are said to be of Class A type if*
- *they satisfy the Markov chain condition*

$$(1.3) \qquad\qquad q'_t \leftrightarrow q_t, q_{[0,t-1]}, q'_{[0,t-1]} \leftrightarrow \{x_0, w_s, s \geq 0\},$$

   that is, almost surely, for all Borel sets B,

$$P(q'_t \in B | q_t, q_{[0,t-1]}, q'_{[0,t-1]}, x_0, w_s, s \geq 0) = P(q'_t \in B | q_t, q_{[0,t-1]}, q'_{[0,t-1]}),$$

   for all $t \geq 0$, and
- *their capacity with feedback is given by*

$$C = \lim_{T \to \infty} \max_{\{P(q_t | q_{[0,t-1]}, q'_{[0,t-1]}),\ 0 \leq t \leq T-1\}} \frac{1}{T} I(q_{[0,T-1]} \to q'_{[0,T-1]}),$$

   where the directed mutual information is defined by

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

Memoryless channels belong to this class; for such channels, feedback does not increase the capacity [8]. Such a class also includes finite state stationary Markov channels which are indecomposable [45] and non-Markov channels which satisfy certain symmetry properties [10]. Further examples can be found in [57] and in [11].

THEOREM 1.1 (see [67], [65]). *Consider the multidimensional linear system* (1.2). *For such a system controlled over a Class A type noisy channel with feedback, if the channel capacity satisfies*

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

(i) *there does not exist a stabilizing coding and control scheme with the property* $\liminf_{T\to\infty} \frac{1}{T} h(x_T) \le 0$; (ii) *the system cannot be made AMS or ergodic (see section* A).

For sufficiency, assume that A is a diagonalizable matrix (a sufficient condition for which is that its eigenvalues are distinct real).

THEOREM 1.2 (see [67], [65]). *Consider the multidimensional linear system* (1.2) *with a diagonalizable matrix A and Gaussian noise, controlled over a discrete memoryless channel. If the Shannon capacity of the channel satisfies*

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

there exists a stabilizing scheme which makes the process $\{x_t\}$ AMS. If the channel is noiseless, or is a memoryless erasure channel, or is a Gaussian channel, then the process $\{x_t\}$ can be made (asymptotically) stationary and ergodic.

1.3. Contributions of the paper. As stated above, stochastic stabilization of nonlinear systems driven by noise (especially unbounded noise) over communication channels has not been studied to our knowledge where the goal is to establish asymptotic (mean) stationarity, ergodicity, or stationarity of the closed-loop system. We use measure-theoretic entropy analysis and ergodic theoretic tools to arrive at necessary and sufficient conditions. A by-product of the analysis is a generalization of Bode's integral formula to a class of nonlinear systems and arbitrary information channels with memory. The approach in the paper, although building on our earlier work on linear systems, contains significant generalizations in the approach due to the non-linearity of the source. We also consider a construction of a stabilizing coding and control scheme for multidimensional nonlinear sources driven by unbounded noise controlled over a discrete noiseless channel.

2. Sublinear entropy growth and a generalization of Bode's integral formula for nonlinear systems. In the paper, instead of a general $\mathbb{R}^N$-valued nonlinear state model

$$(2.1) \qquad\qquad x_{n+1} = f(x_n, u_n, w_n),$$

we will consider nonlinear systems of the form

$$(2.2) \qquad\qquad x_{n+1} = f(x_n, w_n) + Bu_n,$$
$$(2.3) \qquad\qquad x_{n+1} = f(x_n) + Bu_n + w_n,$$
$$(2.4) \qquad\qquad x_{n+1} = f(x_n, u_n) + w_n.$$

We also will have an occasion to discuss nonlinear systems of the form

$$(2.5) \qquad\qquad x_{n+1} = f(x_n, w_n) + B(x_n)u_n.$$

In all of the models above, x_n is the $\mathbb{R}^N$-valued state, w_n is the $\mathbb{R}^N$-valued noise variable, u_n is $\mathbb{R}^s$-valued, and w_n is assumed to be an independent noise process with $w_n \sim \nu$.

We assume throughout that f is measurable and continuously differentiable in the state variable. For a possibly nonlinear differentiable function $f : \mathbb{R}^n \to \mathbb{R}^m$, the Jacobian matrix of f is an $n \times m$ matrix function consisting of partial derivatives of f such that

$$J(f)(i,j) = \frac{\partial (f(x))_i}{\partial x_j}, \quad 1 \le i \le m, 1 \le j \le n.$$

We will have the following assumption throughout the paper.

Assumption 2.1. In the models considered above $f(\cdot, w) : \mathbb{R}^N \to \mathbb{R}^N$ is invertible for every realization of w.

In the following $|J(f)|$ will denote the absolute value of the determinant of the Jacobian. Furthermore, with $f_w(x) = f(x, w)$, we define $J(f(x, w)) := J(f_w(x))$.

Assumption 2.2. There exist $M_1 \in \mathbb{R}$ and $L_1 \in \mathbb{R}$ so that for all x, w

$$L_1 \le \log_2(|J(f(x, w))|) \le M_1.$$

The following is our first result; it provides conditions for sublinear entropy growth (in time) which implies quadratic stability. The result will also be used in the next section and its proof leads to a generalization of Bode's integral formula as discussed further below. Let $\pi_t(B) = P(x_t \in B)$ for all Borel B.

THEOREM 2.1. *Consider the networked control problem over a Class* A *channel.* (i) *Let f have the form in* (2.2), (ii) *let Assumptions* 2.1 *and* 2.2 *hold, and* (iii) *let x_0 have finite differential entropy.* (a) *If there is an admissible coding and control policy such that*

$$\liminf_{t \to \infty} h(x_t)/t \le 0,$$

it must be that

$$(2.6) \qquad C \ge \liminf_{T \to \infty} \frac{1}{T} \sum_{t=0}^{T-1} \int \pi_t(dx) \left(\int \nu(dw) \log_2(|J(f(x, w))|) \right).$$

(b) *If there is an admissible coding and control policy such that*

$$\limsup_{t \to \infty} h(x_t)/t \le 0,$$

it must be that

$$(2.7) \qquad C \ge \limsup_{T \to \infty} \frac{1}{T} \sum_{t=0}^{T-1} \int \pi_t(dx) \left(\int \nu(dw) \log_2(|J(f(x, w))|) \right).$$

In either case, if $L := \inf_{x,w} \log_2 |J(f(x, w))|$, then $C \ge L$.

Remark 1. The condition $\limsup_{t \to \infty} h(x_t)/t \le 0$ is a weak condition. For example, a stochastic process whose second moment grows subexponentially in time so that $\limsup_{T \to \infty} \frac{\log(E[x_T^2])}{T} \le 0$ satisfies this condition. Hence, *quadratic stability* implies this condition.

Remark 2. In the theorem, we would have obtained the same results if we had replaced $\lim \sup_{t \to \infty} h(x_t)/t \leq 0$ with $\lim \sup_{t \to \infty} \frac{1}{t} h(x_t | q'_{[0,t-1]})/t \leq 0$. This condition would be more relevant for state estimation problems, where the goal is not necessarily to make the state stable but to make the estimation error stable (where u_t would be the state estimate and $x_t - u_t$ would be the estimation error). Since $h(x_t | q'_{[0,t-1]}) \leq h(x_t)$, it is evident that the condition $h(x_t)/t \leq 0$ implies that $h(x_t | q'_{[0,t-1]})/t \leq 0$

Proof of Theorem 2.1. Recall that for channels of the type Class A (which includes the discrete memoryless channels as a special case), the capacity is given by

$$C = \lim_{T \to \infty} \max_{\{P(q_t | q_{[0,t-1]}, q'_{[0,t-1]})\}} \frac{1}{T} I(q_{[0,T-1]} \to q'_{[0,T-1]}),$$

where

$$I(q_{[0,T-1]} \to q'_{[0,T-1]}) = \sum_{t=1}^{T-1} I(q_{[0,t]}; q'_t | q'_{[0,t-1]}) + I(x_0; q'_0).$$

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

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

Here, (2.8) follows from the assumption that the channel is of Class A type.

(a) Consider the following:

$$(2.10) \qquad \lim_{T \to \infty} R_T \geq \lim \sup_{T \to \infty} \frac{1}{T} \left(\sum_{t=1}^{T-1} I(x_t; q'_t | q'_{[0,t-1]})) + I(x_0; q'_0) \right)$$

$$= \lim \sup_{T \to \infty} \frac{1}{T} \left(\sum_{t=1}^{T-1} \left(h(x_t | q'_{[0,t-1]}) - h(x_t | q'_{[0,t]}) \right) + I(x_0; q'_0) \right)$$

$$= \lim \sup_{T \to \infty} \frac{1}{T} \sum_{t=1}^{T-1} \left(h(x_t | q'_{[0,t-1]}) - h(x_t | q'_{[0,t]}) \right)$$

$$= \lim \sup_{T \to \infty} \frac{1}{T} \sum_{t=1}^{T-1} \left(h(f(x_{t-1}, w_{t-1}) + Bu_{t-1} | q'_{[0,t-1]}) - h(x_t | q'_{[0,t]}) \right)$$

$$= \lim \sup_{T \to \infty} \frac{1}{T} \sum_{t=1}^{T-1} \left(h(f(x_{t-1}, w_{t-1}) | q'_{[0,t-1]}) - h(x_t | q'_{[0,t]}) \right)$$

$$= \lim \sup_{T \to \infty} \frac{1}{T} \sum_{t=1}^{T-1} \left(\sum_{\zeta_{[0,t-1]}} h(f(x_{t-1}, w_{t-1}) | q'_{[0,t-1]} \right.$$

$$(2.11) \qquad \left. = \zeta_{[0,t-1]}) P(q'_{[0,t-1]} = \zeta_{[0,t-1]}) - h(x_t | q'_{[0,t]}) \right)$$

$$\geq \limsup_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T-1} \left(\sum_{\zeta_{[0,t-1]}} h(f(x_{t-1}, w_{t-1})|q'_{[0,t-1]} \right.$$

(2.12)
$$= \zeta_{[0,t-1]}, w_{t-1}) P(q'_{[0,t-1]} = \zeta_{[0,t-1]}) \bigg) - h(x_t|q'_{[0,t]})$$

$$= \limsup_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T-1} \left(\left(\sum_{\zeta_{[0,t-1]}} \int h(f(x_{t-1}, w)|q'_{[0,t-1]} = \zeta_{[0,t-1]}, w_{t-1} = w) \nu(dw) \right. \right.$$

(2.13)
$$\left. \left. \times P(q'_{[0,t-1]} = \zeta_{[0,t-1]}) \right) - h(x_t|q'_{[0,t]}) \right)$$

$$= \limsup_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T-1} \left(\sum_{\zeta_{[0,t-1]}} P(q'_{[0,t-1]} = \zeta_{[0,t-1]}) \right.$$

$$\times \left(\int \nu(dw) \left(\int P(dx_{t-1}|q'_{[0,t-1]} = \zeta_{[0,t-1]}, w_{t-1} = w) \log_2(|J(f(x_{t-1}, w))|) \right. \right.$$

(2.14)
$$\left. \left. \left. + h(x_{t-1}|q'_{[0,t-1]} = \zeta_{[0,t-1]}, w_{t-1} = w) \right) \right) - h(x_t|q'_{[0,t]}) \right)$$

$$= \limsup_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T-1} \left(\sum_{\zeta_{[0,t-1]}} P(q'_{[0,t-1]} = \zeta_{[0,t-1]}) \right.$$

$$\times \left(\int \nu(dw) \left(\int P(dx_{t-1}|q'_{[0,t-1]} = \zeta_{[0,t-1]}) \log_2(|J(f(x_{t-1}, w))|) \right. \right.$$

(2.15)
$$\left. \left. \left. + h(x_{t-1}|q'_{[0,t-1]} = \zeta_{[0,t-1]}) \right) \right) - h(x_t|q'_{[0,t]}) \right)$$

$$= \limsup_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T-1} \left(\left(\sum_{\zeta_{[0,t-1]}} P(q'_{[0,t-1]} = \zeta_{[0,t-1]}) \int P(dx_{t-1}|q'_{[0,t-1]} = \zeta_{[0,t-1]}) \right. \right.$$

(2.16)
$$\left. \left. \times \int \nu(dw) \log_2(|J(f(x_{t-1}, w))|) \right) + h(x_{t-1}|q'_{[0,t-1]}) - h(x_t|q'_{[0,t]}) \right)$$

$$= \limsup_{T\to\infty} \frac{1}{T} \left(\sum_{t=0}^{T-1} \int \pi_t(dx) \left(\int \nu(dw) \log_2(|J(f(x, w))|) \right) - h(x_{T-1}|q'_{[0,T-1]}) \right)$$

(2.17)
$$\geq V - \liminf_{T\to\infty} \frac{1}{T} h(x_{T-1}|q'_{[0,T-1]}).$$

Here,

(2.18)
$$V := \liminf_{T\to\infty} \frac{1}{T} \sum_{t=0}^{T-1} \int \pi_t(dx) \left(\int \nu(dw) \log_2(|J(f(x, w))|) \right).$$

Equations (2.11) and (2.13) follow from the definition of conditional entropy, and (2.12) follows from conditioning on the random variable w_{t-1}. Equations (2.14)–(2.15) follow from the fact that $x_{t-1} \leftrightarrow q'_{[0,t-1]} \leftrightarrow w_{t-1}$ is a Markov chain and the following. For every realization $q'_{[0,t-1]} = \zeta_{[0,t-1]}$,

$$h(f(x_{t-1}, w_{t-1})|q'_{[0,t-1]} = \zeta_{[0,t-1]}, w_{t-1} = w)$$

$$= h(f_w(x_{t-1})|q'_{[0,t-1]} = \zeta_{[0,t-1]}, w_{t-1} = w)$$

$$= \int P(dx_{t-1}|q'_{[0,t-1]} = \zeta_{[0,t-1]}, w_{t-1} = w) \log_2(|J(f_w(x_{t-1}))|)$$

(2.19) $$\quad + h(x_{t-1}|q'_{[0,t-1]} = \zeta_{[0,t-1]}, w_{t-1} = w)$$

$$= \int P(dx_{t-1}|q'_{[0,t-1]} = \zeta_{[0,t-1]}) \log_2(|J(f(x_{t-1}, w))|) + h(x_{t-1}|q'_{[0,t-1]} = \zeta_{[0,t-1]}),$$

where $f_w(x) := f(x, w)$ is an invertible function for every w, and as a result (2.19) follows from the entropy formula for invertible functions of a random variables (see, e.g., p. 167 of [54] and Lemma 4 in [69]) and the last line follows from the condition $x_{t-1} \leftrightarrow q'_{[0,t-1]} \leftrightarrow w_{t-1}$. Equation (2.16) follows from Fubini's theorem by Assumption 2.2.

By the hypothesis, $\liminf_{t \to \infty} \frac{1}{t} h(x_t) \leq 0$, it must be that $\lim_{T \to \infty} R_T \geq V$. Thus, the capacity also needs to satisfy this bound.

In the above derivation, (2.17) follows from the fact that for two sequences a_n, b_n,

(2.20) $$\limsup_{n \to \infty}(a_n + b_n) \geq \limsup_{n \to \infty} a_n + \liminf_{n \to \infty} b_n.$$

(b) If $\limsup_{t \to \infty} h(x_t|q'_{[0,t-1]})/t \leq 0$, (2.17) can be modified through (2.20) with V defined as

$$\limsup_{T \to \infty} \frac{1}{T} E\left[\left(\sum_{t=1}^{T-1} \int P(dx_t|q'_{[0,t-1]})\left(\int \nu(dw) \log_2(|J(f(x_t, w))|)\right)\right)\right]$$

and in (2.17), $\liminf_{T \to \infty} \frac{1}{T} h(x_{T-1}|q'_{[0,T-1]})$ being replaced with $\limsup$ of the same expression. □

Remark 3. We note that if the system had been of a model in (2.5), the expression involving $J(f(x, w))$ would explicitly depend on the control policy, which would in turn depend possibly on the entire past channel outputs making the expression computationally more involved.

2.1. A generalization of Bode's integral formula for nonlinear systems. The proof of Theorem 2.1 reveals an interesting connection with and generalization of Bode's integral formula (and what is known as the *waterbed effect*) [38] to nonlinear systems, which we state formally in the following. The result also suggests that an appropriate generalization for nonlinear systems is through an information theoretic approach that partially recovers Bode's original result for the linear case, as we discuss further below.

THEOREM 2.2. (i) *Let f have the form in* (2.2), (ii), *let Assumptions* 2.1 *and* 2.2 *hold, and* (iii) *let x_0 have finite differential entropy. If there is an admissible coding and control policy with* $\limsup_{t \to \infty} h(x_t)/t \leq 0$ *it must be that*

$$\limsup_{T \to \infty} \frac{1}{T} I(q_{[0,T-1]} \to q'_{[0,T-1]})$$

(2.21) $$\geq \limsup_{T \to \infty} \frac{1}{T} \sum_{t=0}^{T-1} \int \pi_t(dx)\left(\int \nu(dw) \log_2(|J(f(x, w))|)\right).$$

Proof. This follows directly from (2.9), (2.10), and (2.17). □

Remark 4 (reduction to Bode's integral formula for linear systems and Gaussian noise). If the system considered is linear with all open-loop eigenvalues unstable, the channel is an additive noise channel so that $q'_t = q_t + v_t$ for some stationary Gaussian noise, and time-invariant control policies are considered leading to a stable system, then with the more common notation of $y_t = q'_t$, the right-hand side of (2.20) would be the sum of the unstable eigenvalues of the linear system matrix. For a stationary Gaussian process (see [8, p. 274]) the entropy rate can be written as

$$\frac{1}{2}\log(2\pi e) + \int_{-1/2}^{1/2}\frac{1}{2}\log(S(f))df$$

with S denoting the spectral density of the process. Now, (2.8) becomes

$$I(q'_t; q_{[0,t]}|q'_{[0,t-1]}) = h(q'_t|q'_{[0,t-1]}) - h(q'_t|q_{[0,t]}, q'_{[0,t-1]}) = h(q'_t|q'_{[0,t-1]}) - h(v_t|v_{[0,t-1]}),$$

and thus the left-hand side of (2.20) reduces to the difference between the entropy rate of the process q'_t (that is, $\lim_{t\to\infty} h(q'_t|q_{[0,t-1]})$) and that of the stationary noise process v_t (that is, $\lim_{t\to\infty} h(v_t|v_{[0,t-1]})$). Then, the left-hand side of (2.20) equals

$$\int_{-1/2}^{1/2}\frac{1}{2}\log(\frac{S_y(f)}{S_v(f)})df,$$

which then is equal to the integral of the log-sensitivity function (corresponding to the transfer function from the disturbance process v_t to the output process q'_t). This leads to the celebrated Bode's integral formula. In the context of linear systems, earlier extensions of this formula have been studied in [12] with an information theoretic interpretation under the restriction to linear policies (see, e.g., Theorem 4.6 in [12]), in [31] under more general possibly nonlinear stabilizing control policies which lead to a stationary process, and in [36] and [69] for a class of nonlinear noise-free systems.

3. Asymptotic mean stationarity. In the following, we build on but significantly modify the approaches in [33] and [67] to account for nonlinearity of the system.

Consider the system (2.3), under some admissible policy, controlled over a channel.

Assumption 3.1. We assume

$$M := \sup_{x\in\mathbb{R}^N} \log_2|J(f(x))| < \infty,$$
$$L := \inf_{x\in\mathbb{R}^N} \log_2|J(f(x))| > -\infty.$$

THEOREM 3.1. *Consider the system* (2.3) *controlled over a Class* A *type noisy channel with feedback where* $h(x_0) < \infty$ *and Assumptions* 2.1 *and* 3.1 *hold. If* $C < L$, *then under any admissible policy,*

$$\limsup_{T\to\infty} P(|x_T| \le b(T)) \le 1 - \frac{L-C}{M}$$

for all $b(T) > 0$ *such that* $\lim_{T\to\infty}\frac{1}{T}\log_2(b(T)) = 0$.

The proof is in Appendix B. An implication of this result follows.

THEOREM 3.2. *Consider the system* (2.3) *controlled over a Class* A *type noisy channel with feedback where* $h(x_0) < \infty$ *and Assumptions* 2.1 *and* 3.1 *hold. If under some causal encoding and controller policy the state process is AMS, the channel capacity* C *must satisfy* $C \geq L$.

We recover the following result for linear systems in [67] as a special case.

COROLLARY 3.1. *For the linear case with* $f(x) = Ax$ *with eigenvalues* $|\lambda_i| \geq 1$, $C \geq \sum_k \log_2(|\lambda_i|)$ *is a necessary condition for the AMS property under any admissible coding and control policy.*

Proof of Theorem 3.2. If the process is AMS (see Appendix A), then there exists a stationary measure $\bar{P}$ such that

$$\lim_{N \to \infty} \frac{1}{N} \sum_{k=1}^{N} P(T^{-k}D) = \bar{P}(D) \tag{3.1}$$

for all (cylinder) events D. Let for $b_B \in \mathbb{R}_+$, $B \in \mathcal{B}(\mathbb{R}^N)$ be given by $B = \{x : |x| \leq b_B\}$ and $X_n(z) = z_n$ be the coordinate function (see Appendix A) where $z = \{z_0, z_1, z_2, \ldots\}$.

If by Theorem 3.1

$$\limsup_{T \to \infty} P(|x_T| \leq b_B) \leq 1 - \frac{(L - C)}{M} < 1 \tag{3.2}$$

holds for all $b_B \in \mathbb{R}_+$, then $\bar{P}_n(B) < 1 - \frac{(L-C)}{M}$ for all compact B, where $\bar{P}_n$ is the marginal probability on the nth coordinate defined as

$$\bar{P}_n(B) = \bar{P}\left(x : |X_n(x)| \leq b_B \right).$$

But then $\bar{P}_n$, as an individual probability measure, must be tight [3]; therefore, for every $\delta > 0$ there exists $b_B < \infty$ such that $\bar{P}_n(B) \geq 1 - \delta$. But, by (3.1), this would imply that $\limsup_{t \to \infty} P(T^{-t}B) = \limsup_{t \to \infty} P(|x_t| \in B) \geq 1 - \delta$, leading to a contradiction with (3.2) for $\delta < \frac{L-C}{M}$. Hence, the AMS property cannot be achieved. $\square$

We end this section with a remark.

Remark 5. In information theory, a well-established result is that for noiseless coding of information stable sources (this includes all finite state stationary and ergodic sources) over a class of information stable noisy channels (which includes the channels we consider here), an asymptotically noise-free recovery is possible if the channel capacity is greater than the source entropy through the use of noncausal codes; see, e.g., [60] [25]. However, for the problem we consider (i) the source is nonstationary and open-loop unstable, (ii) the encoding is causal, and (iii) the source process space is not finite-alphabet. Nonetheless, we see that the invariance properties of the source process do appear in the rate bounds that we obtain.

4. Stationarity and ergodicity under structured (stationary) policies. In many applications, one uses a state-space formulation for coding and control policies. In the following, we will consider stationary update rules which have the form that

$$q_t = \gamma^e(x_t, m_t),$$
$$u_t = \gamma^d(m_t, q_t'),$$
(4.1) $$\qquad\qquad m_t = \eta(m_{t-1}, q_{t-1}')$$

for functions γ^e, γ^d, and η. In the form above, m is an $\mathbb{S}$-valued memory or *quantizer state* variable. A large class of adaptive encoding policies have this form. This includes delta modulation, differential pulse coded modulation, adaptive differential pulse coded modulation, Goodman–Gersho type adaptive quantizers (see, e.g., [26], [27]), as well as the coding schemes used for stabilization of networked control systems under fixed-rate codes [64]. Even further, jointly optimal source and channel codes for zero-delay coding schemes under infinite horizon optimization criteria also have the form above (where $\mathbb{S}$ is a space of probability measures [30]). We now present a necessary structural result on the encoders.

4.1. A necessary structural result on the encoders. Let m_t, the state of the encoder, take values in $\mathbb{S}$. Consider (2.3). A stabilizing time-invariant encoder/decoder/controller policy given in (4.1), in general, cannot have $|\mathbb{S}| < \infty$.

THEOREM 4.1. *Consider (2.3) with scalar x_t and w_t with a probability measure ν such that it has a density positive everywhere and $E_\nu[\gamma^{-w_0}] < \infty$ for some $\gamma > 1$. Suppose that there exists $K > 0$ so that*

$$\inf_{x>K} \frac{df}{dx}(x) > 1$$

and $\frac{df}{dx}(x)$ is bounded. Then, a finite cardinality for $\mathbb{S}$ under (4.1) leads to a transient system in the sense that

$$P_x(\tau_S < \infty) < 1,$$

where for some $s > 0$, $S = (-\infty, s)$ is an open set containing the origin, $x > s$, and $\tau_S := \inf(t > 0 : x_t \in S)$. A similar result applies for the condition

$$\sup_{x<-K} \frac{df}{dx}(x) < -1$$

with $S = (s, \infty)$ for some $s < 0$ and $x < s$.

Proof. Let $\inf_{x>K} \frac{df}{dx}(x) > \bar{a} > 1$. It follows from $f(x) = f(K) + \int_K^x \frac{df}{dx}(s)ds$ that for some $M < \infty$, $f(x) \geq M + \bar{a}x$ for $x > K$. Since both q_t' and m_t can take finitely many values, there exists U such that $|u_t| \leq U$ for all t. Let, with $\gamma > 1$, a Lyapunov function be picked as $V(x) = \gamma^{-x}$, defined for positive x. Now, it follows that for sufficiently large x, $E[V(x_{t+1})|x_t = x] \leq V(x)$, since $E[\gamma^{-(f(x)+u_t+w_t)}] = E[\gamma^{-f(x)}\gamma^{-u_t}\gamma^{-w_t}] \leq E[\gamma^{-(M-U)}\gamma^{-(\bar{a}x+w_t)}] = \gamma^{-(M-U)}\gamma^{-\bar{a}x}E[\gamma^{-w}]$, for all $x \in \{x : \gamma^{(\bar{a}-1)x} > E[\gamma^{-w}]\gamma^{-M+U}\}$. Due to the additive noise process the source can escape any bounded interval with a nonzero probability. As a result, by Theorem 6.2.8 in [67] (see also Theorem 8.4.1 in [37]), transience follows. □

Transience prohibits the existence of a stationary probability measure. The discussion above is parallel to Theorem 7.3.1 in [67] for linear systems. Related to the discussion above, for linear systems, the unboundedness of second moments in Proposition 5.1 in [40] and the transience of such a controlled state process were established in Theorem 4.2 in [66]. We also note that [47] studied conditions for stabilization when the control actions are uniformly bounded, and the controlled multidimensional system is marginally stable and is driven by noise with unbounded support.

4.2. Stationarity and ergodicity. In this section, instead of asymptotic mean stationarity, we will consider the more stringent condition of (asymptotic) stationarity of the controlled source process. For ease in presentation we will assume that m_t takes values in a countable set, even though the extension to more general spaces is possible.

LEMMA 4.1. *If the channel is memoryless, the process (x_t, m_t) is a Markov chain.*

Proof. For any $t \in \mathbb{N}$,

$$P(dx_t, m_t | x_s, m_s, s \le t-1)$$
$$= \sum P(dx_t, m_t, q'_{t-1} | x_s, m_s, s \le t-1)$$
$$= \sum P(dx_t | x_{t-1}, \gamma^d(m_{t-1}, q'_{t-1})) P(q'_{t-1} | \gamma^e(x_{t-1}, m_{t-1})) P(m_t | q'_{t-1}, m_{t-1})$$
$$(4.2) = \sum P(dx_t, m_t, q'_{t-1} | x_{t-1}, m_{t-1}) = P(dx_t, m_t | x_{t-1}, m_{t-1}),$$

where we use the fact that the channel is of Class A and (1.3) and (4.1). □

In the following, we assume that the channel is memoryless. For the Markov chain (x_t, m_t), let $\pi_t(B) = P(x_t \in B)$ for all Borel B, that is, π_t is the marginal occupation probability for the state process x_t.

THEOREM 4.2. *Suppose that the encoding, control, and memory update laws are given by (4.1). (i) Let f have the form (2.2), (ii) Assumptions 2.1 and 2.2 hold, and (iii) $h(x_0) < \infty$. For the positive Harris recurrence of the process x_t, m_t (which implies the existence of a unique invariant measure π (and thus ergodicity)), it must be that*

$$(4.3) \qquad C \ge \int \pi(dx) \left(\int \nu(dw) \log_2(|J(f(x, w))|) \right),$$

provided that $\limsup_{t \to \infty} \frac{1}{t} h(x_t) \le 0$.

Proof. First note that

$$C \ge I(q_t, q'_t) = H(q'_t) - H(q'_t | q_t)$$
$$\ge H(q'_t | m_t) - H(q'_t | q_t) = H(q'_t | m_t) - H(q'_t | q_t, x_t, m_t)$$
$$\ge H(q'_t | m_t) - H(q'_t | x_t, m_t) = I(q'_t; x_t | m_t).$$

Hence,

$$C \ge \liminf_{T \to \infty} \frac{1}{T} \sum_{t=0}^{T-1} I(q'_t; x_t | m_t)$$

$$= \liminf_{T \to \infty} \frac{1}{T} \sum_{t=0}^{T-1} \left(h(x_t | m_t) - h(x_t | m_t, q'_t) \right)$$

$$= \liminf_{T \to \infty} \left(\frac{1}{T} \sum_{t=1}^{T-1} \left(h\left(f(x_{t-1}, w_{t-1}) + Bu_{t-1} | m_t \right) - h(x_t | m_t, q'_t) \right) + I(q'_0; x_0 | m_0) \right)$$

$$= \liminf_{T \to \infty} \frac{1}{T} \sum_{t=1}^{T-1} h\left(f(x_{t-1}, w_{t-1}) + Bu_{t-1}) | m_t \right) - h(x_t | m_t, q'_t)$$

$$\ge \liminf_{T \to \infty} \frac{1}{T} \sum_{t=1}^{T-1} h\left(f(x_{t-1}, w_{t-1}) + Bu_{t-1} | m_t, m_{t-1}, q'_{t-1} \right) - h(x_t | m_t, q'_t)$$

$$= \liminf_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T-1} h\Big(f(x_{t-1}, w_{t-1}) + Bu_{t-1} | m_{t-1}, q'_{t-1} \Big) - h(x_t | m_t, q'_t)$$

$$\geq \liminf_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T-1} h\Big(f(x_{t-1}, w_{t-1}) + Bu_{t-1} | w_{t-1}, m_{t-1}, q'_{t-1} \Big) - h(x_t | m_t, q'_t)$$

$$= \liminf_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T-1} h\Big(f(x_{t-1}, w_{t-1}) | w_{t-1}, m_{t-1}, q'_{t-1} \Big) - h(x_t | m_t, q'_t)$$

$$= \liminf_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T-1} \Big(\int \nu(dw) \Big(\sum P(m_{t-1} = m, q'_{t-1} = q')$$

$$\times \int P(x_{t-1} \in dx | m_{t-1} = m, q'_{t-1} = q', w_{t-1} = w) \log_2(|J(f(x, w_{t-1}))|)$$

$$+ h(x_{t-1} | m_{t-1} = m, q'_{t-1} = q', w_{t-1} = w) \Big) - h(x_t | m_t, q'_t) \Big)$$

$$= \liminf_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T-1} \Big(\int \nu(dw) \Big(\sum P(m_{t-1} = m, q'_{t-1} = q')$$

$$\times \int P(x_{t-1} \in dx | m_{t-1} = m, q'_{t-1} = q') \log_2(|J(f(x, w))|)$$

$$\text{(4.4)} \qquad + h(x_{t-1} | m_{t-1} = m, q'_{t-1} = q') \Big) - h(x_t | m_t, q'_t) \Big)$$

$$= \liminf_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T-1} \Big(\int \nu(dw) \Big(\int \pi_{t-1}(dx) \log_2(|J(f(x, w))|)$$

$$\text{(4.5)} \qquad + h(x_{t-1} | m_{t-1}, q'_{t-1}) \Big) - h(x_t | m_t, q'_t) \Big)$$

$$= \liminf_{T\to\infty} \frac{1}{T} \Big(\sum_{t=1}^{T-1} \int \pi_{t-1}(dx) \Big(\int \nu(dw) \log_2(|J(f(x, w))|) \Big)$$

$$\text{(4.6)} \qquad - h(x_{T-1} | m_{T-1}, q'_{T-1}) \Big)$$

$$\geq \liminf_{T\to\infty} \frac{1}{T} \Big(\sum_{t=1}^{T-1} \int \pi_{t-1}(dx) \Big(\int \nu(dw) \log_2(|J(f(x, w))|) \Big) - h(x_{T-1}) \Big)$$

$$\geq \liminf_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T-1} \int \pi_{t-1}(dx) \Big(\int \nu(dw) \log_2(|J(f(x, w))|) \Big) - \limsup_{T\to\infty} \frac{1}{T} h(x_{T-1})$$

$$\text{(4.7)} \quad \geq \liminf_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T-1} \int \pi_{t-1}(dx) \Big(\int \nu(dw) \log_2(|J(f(x, w))|) \Big)$$

$$\text{(4.8)} \quad = \liminf_{T\to\infty} \int \pi_0(dx) E_x \Big[\frac{1}{T} \sum_{t=0}^{T-1} \int \pi_{t-1}(dx) \Big(\int \nu(dw) \log_2(|J(f(x, w))|) \Big) \Big]$$

$$(4.9) \quad \geq \int \pi_0(dx) \liminf_{T \to \infty} E_x \left[\frac{1}{T} \sum_{t=0}^{T-1} \int \pi_{t-1}(dx) \left(\int \nu(dw) \log_2(|J(f(x,w))|) \right) \right]$$

$$(4.10) \quad = \int \pi_0(dx) \left(\int \pi(dz) \left(\int \nu(dw) \log_2(|J(f(z,w))|) \right) \right)$$

$$(4.11) \quad = \int \pi(dx) \left(\int \nu(dw) \log_2(|J(f(x,w))|) \right).$$

In the first lines above, we use the fact that conditioning on a random variable reduces the entropy and the update laws (4.1). The equality (4.5) holds since for every w, the map $f(.,w)$ is invertible (here $J(f(x,w_t))$ is the Jacobian for the realized value of w_t) and that w_t is an independent noise process using the laws of total probability. Here (4.4) follows due to the independence of w_t, (4.6) follows from Fubini's theorem since $\log_2(|J(f(x,w))|)$ is bounded, (4.8) follows from Fatou's lemma given the assumption that $\log_2(|J(f(x,w))|)$ is bounded from below, and (4.9) follows from positive Harris recurrence (see [20, Theorem 4.3.1]). $\qquad \square$

Remark 6. If one considers a more general control-affine model such as of the form (2.5) with $x_{t+1} = f(x_t, w_t) + B(u_t)x_t$, the condition would read as

$$C \geq \int \pi(dx, m, q') \left(\int \nu(dw) \log_2 \left(\left| J \left(f(x,w) + B(\gamma^d(m, q'))x \right) \right| \right) \right),$$

where π is invariant for the (enlarged) Markov chain (x_t, m_t, q'_t).

5. Discrete noiseless channels and a stationary and ergodic construction. In this section, we provide achievability results and a stabilizing coding/control policy. As discussed earlier, the study of nonlinear systems has typically considered noise-free controlled systems, e.g., [2], [29], and [46]. As also noted earlier, for noise-free systems, it typically suffices to consider only a sufficiently small *invariant* neighborhood of an equilibrium point to obtain stabilizability conditions which is not necessarily the case when the system is driven by an additive noise process. We consider such an example in the following.

THEOREM 5.1. *Consider a nonlinear system of the form* (2.4), *where* $\{w_t\}$ *is a sequence of zero-mean Gaussian random vectors and there exists a control function* $\kappa(z)$ *such that* $|f(x, \kappa(z))|_\infty \leq |a||x - z|_\infty$ *for all* $x, z \in \mathbb{R}^N$, *with* $\kappa(0) = 0$. *For the stationarity and ergodicity of* $\{x_t\}$ *(and thus with a unique invariant probability measure), it suffices that* $C > N \log_2(|a|) + 1$.

Remark 7. It may be possible in general to reduce the rate requirements by the use of variable-rate encoding schemes; for example, if there exists a compact region outside of which the constant a can be upper bounded by a smaller number, a region-dependent quantization rate can be applied which can reduce the average data rate required for system stability. In this paper, since there is an explicit channel, our focus has been on fixed-rate coding schemes.

Proof. The proof follows essentially from the approach developed in [64] and [21] with extension to nonlinear analysis. Consider the case with $N = 2$. Let $\Delta > 0$ denote the bin size for a uniform quantizer and let for each coordinate $x^i \in \mathbb{R}, i = 1, 2,$

$$(5.1) \quad Q_K^\Delta(x^i) = \begin{cases} (k - \frac{1}{2}(K+1))\Delta & \text{if } x^i \in [(k-1-\frac{1}{2}K)\Delta, (k-\frac{1}{2}K)\Delta), \\ (\frac{1}{2}(K-1))\Delta & \text{if } x^i = \frac{1}{2}K\Delta, \\ 0 & \text{if } x^i \notin [-\frac{1}{2}K\Delta, \frac{1}{2}K\Delta], \end{cases}$$

and define $\mathbf{Q}^{\Delta}(x) = (Q_K^{\Delta}(x^1), Q_K^{\Delta}(x^2))$ if $Q_K^{\Delta}(x^i) \neq 0$ for $i = 1, 2$ and $\mathbf{Q}^{\Delta}(x) = 0$ if $Q_K^{\Delta}(x^i) = 0$ for some i. Thus, the number of symbols in the image of $\mathbf{Q}^{\Delta}$ is $K^2 + 1$ (and not $(K+1)^2$). The quantizer outputs are transmitted through a memoryless erasure channel, after being subjected to a bijective mapping, which is performed by the channel encoder. The channel encoder maps the quantizer output symbols to corresponding channel inputs $q \in \mathcal{M} := \{1, 2 \ldots, K^2 + 1\}$. A channel encoder at time t, denoted here by $\mathcal{E}_t$, maps the quantizer outputs to $\mathcal{M}$ such that $\mathcal{E}_t(Q_t(x_t)) = q_t \in \mathcal{M}$. For $i = 1, 2$, let $R' = \log_2(K)$. For $t \geq 0$ and with $\Delta_0^1 = \Delta_0^2 \in \mathbb{R}$, define

$$h_t^i = \frac{x_t^i}{\Delta_t^i 2^{R'-1}},$$

and with

$$\hat{x}_t = \begin{bmatrix} \hat{x}_t^1 \\ \hat{x}_t^2 \end{bmatrix}$$

consider

$$u_t = -\kappa(\hat{x}_t),$$

$$(5.2) \qquad \begin{bmatrix} \hat{x}_t^1 \\ \hat{x}_t^2 \end{bmatrix} = \begin{bmatrix} Q_{K_1}^{\Delta_t^1}(x_t^1) \\ Q_{K_2}^{\Delta_t^2}(x_t^2) \end{bmatrix} 1_{\{\max_i |h_i| \leq 1\}} + \begin{bmatrix} 0 \\ 0 \end{bmatrix} 1_{\{\max_i |h_i| > 1\}},$$

$$(5.3) \qquad \Delta_{t+1}^1 = \Delta_t^1 \bar{Q}(|h_t^1|, |h_t^2|, \Delta_t^1, \Delta_t^2), \qquad \Delta_{t+1}^2 = \Delta_t^2 \bar{Q}(|h_t^1|, |h_t^2|, \Delta_t^1, \Delta_t^2),$$

with, for $i = 1, 2$, $\delta > 0$ $\alpha \in (0, 1)$, $L > 0$, such that

$$\begin{aligned}
\bar{Q}(x, y, \Delta^1, \Delta^2) &= |a| + \delta \quad \text{if} \quad |x| > 1, \quad \text{or} \quad |y| > 1, \\
\bar{Q}(x, y, \Delta^1, \Delta^2) &= \alpha \quad \text{if} \quad |x| \leq 1, |y| \leq 1; \quad \Delta^1 > L, \Delta^2 > L, \\
\bar{Q}(x, y, \Delta^1, \Delta^2) &= 1 \quad \text{if} \quad |x| \leq 1, |y| \leq 1; \quad \Delta^1 \leq L \quad \text{or} \quad \Delta^2 \leq L.
\end{aligned}$$

Note that the above imply $\Delta_t^i \geq \alpha L$. See Figure 2 for a depiction of the quantizer used at a particular time. To make the state space for the bin size process countable as in [67], [64], we take that $\log_2(\bar{Q}(\cdot))$ take values in integer multiples of s where the integers taken are relatively prime (that is, they share no common divisors except for 1); see [67, Lemma 7.6.2]. We note the following without proof.

LEMMA 5.1. *The process (x_t, Δ_t) is a Markov chain.*

We define a sequence of stopping times as follows:

$$\mathcal{T}_0 = 0, \quad \mathcal{T}_{z+1} = \inf\{k > \mathcal{T}_z : |h_k^i| \leq 1, i \in \{1, 2\}\}, \quad z \in \mathbb{Z}_+.$$

By the strong Markov property and the nature of the stopping times, $(x_{\mathcal{T}_z}, h_{\mathcal{T}_z})$ is also Markov. In the following, we show that there exist $b_0 > 0$, $b_1 < \infty$ such that

$$(5.4) \qquad E[\log(\Delta_{\mathcal{T}_{z+1}}^2)|\Delta_{\mathcal{T}_z}, h_{\mathcal{T}_z}] \leq \log(\Delta_{\mathcal{T}_z}^2) - b_0 + b_1 1_{\{|\Delta_{\mathcal{T}_z}| \leq F\}}.$$

We first bound the probability $P(\mathcal{T}_{z+1} - \mathcal{T}_z \geq k | \Delta_{\mathcal{T}_z}, h_{\mathcal{T}_z})$ from above.

LEMMA 5.2. *The discrete probability measure $P(\mathcal{T}_{z+1} - \mathcal{T}_z = k \mid x_{\mathcal{T}_z}, \Delta_{\mathcal{T}_z})$ has the upper bound*

$$P(\mathcal{T}_{z+1} - \mathcal{T}_z \geq k | x_{\mathcal{T}_z}, \Delta_{\mathcal{T}_z}) \leq M(\Delta_{\mathcal{T}_z}) r^{-k},$$

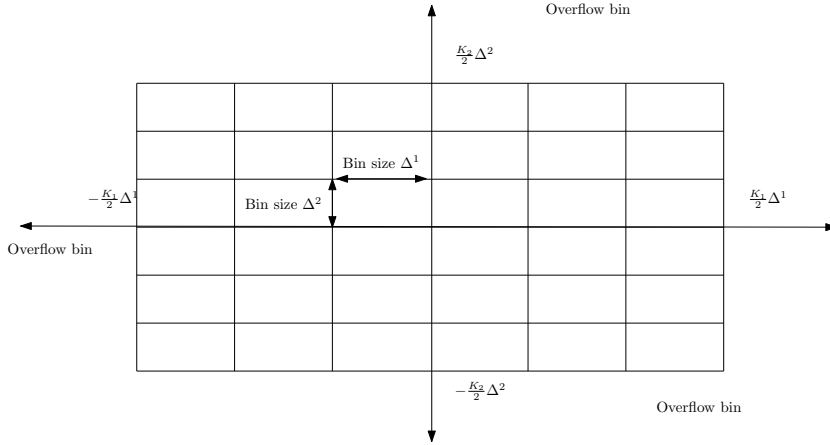*for some $r > 1$ and $\lim_{\Delta \to \infty} M(\Delta) = 0$.*

FIG. 2. *The vector quantizer. There is a single overflow bin.*

Proof. Observe that for $0 < k < \tau_1$, $\hat{x}_k = 0$ and $u_k = \kappa(0) = 0$. Let $|x| = \|x\|_\infty$. Now, for $k \geq 2$

$$P(\mathcal{T}_1 \geq k | x_0, \Delta_0) \leq P_{x_0, \Delta_0}\left(|x_{k-1}| \geq (|a| + \delta)^{k-2} 2^{R'-1} \alpha \Delta_0 \right)$$

$$\leq P_{x_0, \Delta_0}\left(|f(x_{k-2})| + |w_{k-2}| \geq (|a| + \delta)^{k-2} 2^{R'-1} \alpha \Delta_0 \right)$$

$$\leq P_{x_0, \Delta_0}\left(|a(x_{k-2})| + |w_{k-2}| \geq (|a| + \delta)^{k-2} 2^{R'-1} \alpha \Delta_0 \right)$$

$$\leq P_{x_0, \Delta_0}\left(|a||x_{k-2}| + |w_{k-2}| \geq (|a| + \delta)^{k-2} 2^{R'-1} \alpha \Delta_0 \right)$$

$$(5.5) \qquad \leq P_{x_0, \Delta_0}\left(\sum_{i=0}^{k-2} |a|^{-i} |w_i| \geq \frac{(|a| + \delta)^{k-2} 2^{R'-1} \alpha \Delta_0}{|a|^{k-1}} - |x_0 - \hat{x}_0| \right)$$

$$\leq P_{x_0, \Delta_0}\left(\sum_{i=0}^{k-2} |a|^{-i} |w_i| \geq \frac{(|a| + \delta)^{k-2} 2^{R'-1} \alpha \Delta_0}{|a|^{k-1}} - \Delta_0/2 \right)$$

$$(5.6) \qquad = P_{x_0, \Delta_0}\left(\sum_{i=0}^{k-2} |a|^{-i} |w_i| \geq \Delta_0/2 \left((\frac{|a| + \delta}{|a|})^{k-2} \frac{2^{R'} \alpha}{|a|} - 1 \right) \right)$$

$$(5.7) \qquad \leq \frac{E[\sum_{i=0}^{\infty} |a|^{-i} |w_i|]}{\Delta_0/2 \left((\frac{|a| + \delta}{|a|})^{k-2} \frac{2^{R'} \alpha}{|a|} - 1 \right)}$$

$$(5.8) \qquad \leq M(\Delta_0) r^{-k}$$

with

$$M(\Delta_0) = \frac{KE[\sum_{i=1}^{\infty} |a|^{-i} |w_i|]}{(\frac{|a| + \delta}{|a|})^2 \Delta_0 (\frac{2^{R'} \alpha}{|a|} - 1)} < \infty$$

for some $K < \infty$ and $r \in (1, (|a| + \delta)/|a|)$ so that $\lim_{\Delta_0 \to \infty} M(\Delta_0) = 0$. Here, (5.5) follows from an inductive argument, (5.6) follows from the fact that the term

$$\left(2^{R'-1} (\frac{|a| + \delta}{|a|})^{k-2} \frac{\alpha}{|a|} - \frac{1}{2} \right)$$

is positive for $k \geq 2$ provided that $2^{R'} > \frac{|a|}{\alpha}$, (5.7) follows from Markov's inequality, and (5.8) follows from the fact that w_i is Gaussian together with the property $|w| \leq N(1 + |w|^2)$ leading to the finiteness of $E[\sum_{i=1}^{\infty} |a|^{-i} |w_i|]$. $\qquad \square$

We now invoke [68, Theorem 2.1]: Let $\boldsymbol{X}$ be an $\mathbb{X}$-valued Markov chain (where $\mathbb{X}$ is a standard Borel space) and $\mathcal{T}_z, z \geq 0$, be a sequence of stopping times measurable on the filtration generated by the state process with $\mathcal{T}_0 = 0$.

THEOREM 5.2 (see [68, Theorem 2.1]). *Suppose that $\boldsymbol{X}$ is a φ-irreducible and aperiodic Markov chain. Suppose moreover that there are functions $V : \mathbb{X} \to (0, \infty)$, $\delta : \mathbb{X} \to [1, \infty)$, $f : \mathbb{X} \to [1, \infty)$, a small set C on which V is bounded, and a constant $b \in \mathbb{R}$, such that the following hold:*

(5.9)
$$E[V(x_{\mathcal{T}_{z+1}}) \mid \mathcal{F}_{\mathcal{T}_z}] \leq V(x_{\mathcal{T}_z}) - \delta(x_{\mathcal{T}_z}) + b 1_{\{x_{\mathcal{T}_z} \in C\}},$$
$$E\Big[\sum_{k=\mathcal{T}_z}^{\mathcal{T}_{z+1}-1} f(x_k) \mid \mathcal{F}_{\mathcal{T}_z} \Big] \leq \delta(x_{\mathcal{T}_z}), \qquad\qquad z \geq 0.$$

Then the following hold:
  (i) *$\boldsymbol{X}$ is positive Harris recurrent, with unique invariant distribution π.*
  (ii) *$\pi(f) := \int f(x) \, \pi(dx) < \infty$.*
  (iii) *For any function g that is bounded by f, in the sense that $\sup_x |g(x)|/f(x) < \infty$, we have convergence in the mean, and the law of large numbers holds:*

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

By taking $f(x) = 1$ for all $x \in \mathbb{X}$, the following holds.

THEOREM 5.3 (see [68]). *Suppose that $\boldsymbol{X}$ is a φ-irreducible Markov chain with natural filtration $\mathcal{F}_t$. Suppose moreover that there is a function $V : \mathbb{X} \to (0, \infty)$, a small set C on which V is bounded, and a constant $b \in \mathbb{R}$, such that the following hold:*

(5.10)
$$E[V(x_{\mathcal{T}_{z+1}}) \mid \mathcal{F}_{\mathcal{T}_z}] \leq V(x_{\mathcal{T}_z}) - 1 + b 1_{\{x_{\mathcal{T}_z} \in C\}},$$
$$\sup_{z \geq 0} E[\mathcal{T}_{z+1} - \mathcal{T}_z \mid \mathcal{F}_{\mathcal{T}_z}] < \infty.$$

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

Now, with the candidate Lyapunov function $V_0(x_t, \Delta_t) = \log(\Delta_t^2)$, for $\Delta_{\mathcal{T}_z} > L$,

$$E[V_0(x_{\mathcal{T}_{z+1}}, \Delta_{\mathcal{T}_{z+1}}) \mid x_{\mathcal{T}_z}, \Delta_{\mathcal{T}_z}] = P(\mathcal{T}_{z+1} - \mathcal{T}_z = 1)\left(2\log(\alpha) + \log(\Delta_{\mathcal{T}_z}^2) \right)$$
$$+ \sum_{k=2}^{\infty} \log(\Delta_{\mathcal{T}_z+k}^2) P(\mathcal{T}_{z+1} - \mathcal{T}_z = k \mid x_{\mathcal{T}_z}, \Delta_{\mathcal{T}_z})$$

$$= P(\mathcal{T}_{z+1} - \mathcal{T}_z = 1)\left(2\log(\alpha) + \log(\Delta^2_{\mathcal{T}_z})\right)$$

$$+ \sum_{k=2}^{\infty} 2(\log_2(\alpha) + (k-1)(|a| + \delta)M(\Delta)r^{-k}.$$

Now, by (5.8), $\lim_{\Delta_0 \to \infty} P(\mathcal{T}_{z+1} - \mathcal{T}_z = 1|\Delta_0, x_0) = 1$ uniformly in $|x_0| \le 2^{R'-1}\Delta_0$. As a result, the drift condition of Theorem 5.3 holds. We need to ensure, however, the small/petite set [37] property of compact sets to establish positive Harris recurrence. A sufficiently small compact set for this chain is petite due to the countability of the values that Δ_t takes and the *uniform countable additivity* property of the Markov chain due to the presence of the additive Gaussian noise, as on p. 206 of [67] and the continuity of f in x. This argument applies for N-dimensional systems as well with $N > 2$. This completes the proof of Theorem 5.1. □

Remark 8. The approach adopted in the proof of Theorem 5.1 applies for more general channels (such as erasure channels or discrete memoryless channels) subject to more tedious error bounds.

6. Discussion and conclusion. In this paper, conditions on information channels leading to the stochastic stability of nonlinear systems controlled over noisy channels have been investigated. Stochastic stability notions considered were asymptotic mean stationarity, ergodicity, and stationarity. Results for linear systems are recovered as a special case.

In the following we present some future directions and a comparison with the results involving topological entropy.

6.1. Comparison with invariance entropy and deterministic nonlinear systems controlled over noiseless channels. As noted earlier, noise-free systems and noiseless discrete channels have been studied in the literature in the context of topological entropy and invariance entropy. Here, we establish some connections.

One related result in this literature is with regard to stabilization to a point: Under the assumptions that (i) f has the form in (2.4) (without noise) with continuous partial derivatives, (ii) there exists a fixed point (equilibrium) x^* so that $x^* = f(x^*, u^*)$, (iii) a local strong invariability condition is satisfied which relates the size of an invariant set and the size of a control action set in the sense that for any $\epsilon > 0$, there exist $\rho > 0$ so that for all $\epsilon' \in (0, \rho]$, the set $\{x : |x - x^*| \le \epsilon'\}$ is strongly invariant with the control action set $U = \{u : |u - u^*| \le \epsilon\}$, and (iv) the pair (A, B) is controllable where A, B are the Jacobians of f with respect to state and control at x^*, u^*, [41] has reported that for convergence to the equilibrium an average rate $R > \sum_{|\lambda_i|>1} \log_2(|\lambda_i|)$ is sufficient, where λ_i are the eigenvalues of the Jacobian at the equilibrium point.

A further related result in the spirit of our paper is from a case where there exists an invariant set with a nonempy interior: For continuous-time systems of the form $\frac{dx}{dt} = f(x, u), u \in U$, Colonius and Kawan [6] establish a lower bound on *invariance entropy* as

$$(6.1) \qquad\qquad \max\left(0, \min_{x,u \in Q \times U} \sum_i \frac{\partial f_i}{\partial x_i}(x, u)\right),$$

where Q is a weakly invariant set and f_i is the ith coordinate function of f. More refined bounds are present if further structural properties are imposed: in [53], under

a uniform hyperbolicity assumption (see [53, Definition 4.4]), Theorem 4.8 states a similar lower bound by considering the unstable components in an invariant set.

These results can be viewed to be related to Theorems 3.2 and 4.2, as well as Theorem 5.1, in that the average entropy growth as measured by the eigenvalues of the Jacobian matrix under the invariant probability measure is lower bounded by a minimum over the elements in the support set, or is upper bounded by a maximizing element in the support set. In the stabilization to the point example of [41], the invariant measure is a delta measure on a single point. In the invariant set example leading to (6.1), the set Q can be viewed to be the support set of some invariant measure under the system dynamics if such a measure were to exist. Likewise, [29] and [28] have obtained conditions for noise-free systems controlled over noiseless channels. Due to the absence of noise, one could identify an invariant compact set and consider a bound on the Lipschitz growth parameter for the system over this invariant set to obtain sufficiency conditions. When the system is (Lebesgue) irreducible, however, due to the effect of noise, local properties are not descriptive and the invariant probability measure reflects the rate conditions and entropy growth in the system. In this case, the local growth integrated under an invariant measure gives a proper bound.

Differential entropy is a useful measure for how much a stochastic system generates uncertainty; however, our analysis does not distinguish between the stable and unstable modes of a controlled system and is only able to resemble the classical results in ergodic theory (Pesin's formula [62]) for *expanding* systems and thus with only positive Lyapunov exponents. In the linear case, the arguments follow by restricting the state space to those corresponding to the unstable modes. For a general nonlinear system, however, a careful geometric study needs to be done. On the other hand, for deterministic systems, under a topological entropy formulation, the rate of growth can be measured by local Jacobian matrices, but such a topological discussion requires further geometric analysis with regard to the use of appropriate metrics, as studied extensively in [24]. Thus, the connection between the differential entropy method and geometric approaches requires further study.

We note also that recently a metric entropy generalization of some of the results in [24] have been developed [5].

6.2. Some open directions on stationary coding and control policies and information theory. It would be interesting to show, for a class of systems, that stationary coding and control policies can be used to arrive at stability with a stationary closed-loop process provided that the capacity of the channel satisfies the entropy growth bound and the channel satisfies certain ergodicity conditions. However, except for linear Gaussian systems controlled over Gaussian channels and erasure channels (see [67] for a detailed discussion for both setups), this question has not been answered even for linear systems controlled over general discrete memoryless channels (that is, nonstationary coding schemes have been used for more general discrete memoryless channels). Furthermore, the tightness of the converse results is another direction. A further direction is the causal coding problem for nonergodic sources: In the information theory literature, through noncausal codes, a class of source coding theorems for nonergodic sources exists (see, e.g., [17]); however, the extensions of these for even control-free nonlinear systems under causal coding require further research.

Appendix A. Stationary, ergodic, and AMS processes. In this appendix, we review ergodic theory, in the context of information theory (that is, with the transformations being specific to the shift operation). A comprehensive discussion is available in Shields [50], Gray [15], Gray and Kieffer [19], and Appendix C in [67].

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

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

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

DEFINITION A.1. *Let μ be the measure on a process. This random process is ergodic if $A = T^{-1}A$ implies that $\mu(A) \in \{0, 1\}$.*

That is, the events that are unchanged with a shift operation are trivial events.

Mixing is a sufficient condition for ergodicity. Thus, a source is ergodic if $\lim_{n \to \infty} P(A \cap T^{-n} B) = P(A) P(B)$, since the process forgets its initial condition. For the special case of Markov sources, we have the following: A positive Harris recurrent Markov chain is ergodic, since such a process is mixing and stationary.

DEFINITION A.2. *A process on a probability space $(\Omega, \mathcal{F}, \mathbf{P})$ with process measure P is asymptotically mean stationary (AMS) if there exists a probability measure $\bar{P}$ such that*

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

for all events $F \in \mathcal{B}(\Sigma)$. Here $\bar{P}$ is called the stationary mean of P and is a stationary measure.

Note that $\bar{P}$ is stationary since by definition $\bar{P}(F) = \bar{P}(T^{-1} F)$. For the importance of the AMS property, its relations with Birkhoff's ergodic theorem, some applications and sufficient conditions, see [15] or [19].

Appendix B. Proof of Theorem 3.1. Define the event for $K > 0$ so that $P(|x_0| < K) > 0$ as

$$\mathcal{S}_\eta^K = \{\omega : |x_0| \leq K, w = \eta, i.e., w_k = \eta_k, \eta_k \in \mathbb{R}^p, k \geq 0\}$$

such that the noise realizations are fixed and deterministic. In the following, we will drop the subscript and superscripts and let $P_{\mathcal{S}}$ or $P(\cdot | \mathcal{S})$ denote the conditional probabilities given the event $\mathcal{S}_\eta^K$. We recall here that $\{w_t, t \geq 0\}$ and x_0 are assumed to be independent. By Definition 1.1, first note that the capacity expression satisfies

$$C = \lim_{T \to \infty} \max_{\{P(q_t | q_{[0,t-1]}, q'_{[0,t-1]}), \quad 0 \leq t \leq T-1\}} \frac{1}{T} I(q_{[0,T-1]} \to q'_{[0,T-1]})$$

(B.1) $$= \lim_{T \to \infty} \max_{\{P(q_t | q_{[0,t-1]}, q'_{[0,t-1]}), \quad 0 \leq t \leq T-1\}} \frac{1}{T} I(q_{[0,T-1]} \to q'_{[0,T-1]} | \mathcal{S}),$$

where the conditional directed information is given by

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

Here, (B.1) is a result of the following: Consider an encoder policy given by

$$P^* = \{P^*(q_0), P^*(q_1|q_0, q_0'), \cdots, P^*(q_t|q_{[0,t-1]}, q_{[0,t-1]}'), \cdots\}.$$

For any $t \in \mathbb{N}$, almost surely the following holds:

$$P(q_t'|q_{[0,t-1]}', \mathcal{S})$$

$$= \sum_{q_{[0,t]}} P(q_t', q_{[0,t]}|q_{[0,t-1]}', \mathcal{S})$$

$$= \sum_{q_{[0,t]}} P(q_t'|q_{[0,t]}, q_{[0,t-1]}', \mathcal{S}) P(q_{[0,t]}|q_{[0,t-1]}', \mathcal{S})$$

(B.2) $\quad = \sum_{q_{[0,t]}} P(q_t'|q_{[0,t]}, q_{[0,t-1]}') P(q_{[0,t]}|q_{[0,t-1]}', \mathcal{S})$

$$= \sum_{q_{[0,t]}} P(q_t'|q_{[0,t]}, q_{[0,t-1]}') P^*(q_t|q_{[0,t-1]}', q_{[0,t-1]}, \mathcal{S}) P(q_{[0,t-1]}|q_{[0,t-1]}', \mathcal{S})$$

(B.3) $\quad = \sum_{q_{[0,t]}} P(q_t'|q_{[0,t]}, q_{[0,t-1]}') P^*(q_t|q_{[0,t-1]}', q_{[0,t-1]}) P(q_{[0,t-1]}|q_{[0,t-1]}', \mathcal{S})$

(B.4) $\quad = \sum_{q_{[0,t]}} P(q_t'|q_{[0,t]}, q_{[0,t-1]}') P^*(q_t|q_{[0,t-1]}', q_{[0,t-1]}) P(q_{[0,t-1]}|q_{[0,t-1]}')$

(B.5) $\quad = P(q_t'|q_{[0,t-1]}'),$

where (B.2) follows from Definition 1.1, (B.3) from the structure of a coding policy, and (B.4) from the following inductive argument. Note that $P(q_0, q_0'|\mathcal{S}) = P(q_0, q_0')$. If $P(q_{[0,t-1]}, q_{[0,t-1]}'|\mathcal{S}) = P(q_{[0,t-1]}, q_{[0,t-1]}')$, it follows that

$$P(q_{[0,t]}, q_{[0,t]}'|\mathcal{S})$$

$$= P(q_t'|q_{[0,t]}, q_{[0,t-1]}', \mathcal{S}) P^*(q_t|q_{[0,t-1]}', q_{[0,t-1]}, \mathcal{S}) P(q_{[0,t-1]}, q_{[0,t-1]}'|\mathcal{S})$$

$$= P(q_t'|q_{[0,t]}, q_{[0,t-1]}') P^*(q_t|q_{[0,t-1]}', q_{[0,t-1]}) P(q_{[0,t-1]}, q_{[0,t-1]}'|\mathcal{S})$$

$$= P(q_t'|q_{[0,t]}, q_{[0,t-1]}') P^*(q_t|q_{[0,t-1]}', q_{[0,t-1]}) P(q_{[0,t-1]}, q_{[0,t-1]}')$$

(B.6) $\quad = P(q_{[0,t]}, q_{[0,t]}').$

As a result, (B.4) simplifies to (B.5) by eliminating the conditioning on $\mathcal{S}$ and (B.1) holds.

We now use a similar argument as in (2.17) but need to modify the steps due to the conditioning on $\mathcal{S}$:

$$\lim_{T\to\infty} R_T \geq \limsup_{T\to\infty} \frac{1}{T} \left(\sum_{t=1}^{T-1} I(x_t; q_t'|q_{[0,t-1]}', \mathcal{S}) + I(x_0; q_0'|\mathcal{S}) \right)$$

$$= \limsup_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T-1} I(x_t; q_t'|q_{[0,t-1]}', \mathcal{S})$$

$$= \limsup_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T-1} h_{\mathcal{S}}(x_t|q_{[0,t-1]}') - h_{\mathcal{S}}(x_t|q_{[0,t]}')$$

$$= \limsup_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T-1} h_\mathcal{S}(f(x_{t-1}) + Bu_{t-1} + w_{t-1}|q'_{[0,t-1]}) - h_\mathcal{S}(x_t|q'_{[0,t]})$$

$$= \limsup_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T-1} h_\mathcal{S}(f(x_{t-1}) + Bu_{t-1}|q'_{[0,t-1]}) - h_\mathcal{S}(x_t|q'_{[0,t]})$$

$$= \limsup_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T-1} h_\mathcal{S}(f(x_{t-1})|q'_{[0,t-1]}) - h_\mathcal{S}(x_t|q'_{[0,t]})$$

$$= \limsup_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T-1} \left(E\left[P_\mathcal{S}(dx_{t-1}|q'_{[0,t-1]}) \log_2(|J(f(x_{t-1}))|) \right] \right.$$

$$\text{(B.7)} \qquad\qquad\qquad \left. + h_\mathcal{S}(x_{t-1}|q'_{[0,t-1]}) - h_\mathcal{S}(x_t|q'_{[0,t]}) \right)$$

$$\text{(B.8)} \quad \geq V_\mathcal{S} - \liminf_{T\to\infty} \left(\frac{1}{T} h_\mathcal{S}(x_{T-1}|q'_{[0,T-1]}) \right)$$

Here,

$$\text{(B.9)} \qquad V_\mathcal{S} = \liminf_{T\to\infty} E\left[\frac{1}{T} \left(\sum_{t=1}^{T-1} P_\mathcal{S}(dx_{t-1}|q'_{[0,t-1]}) \log_2(|J(f(x_{t-1}))|) \right) \right],$$

and (B.7) follows from the fact that

$$h_\mathcal{S}(f(x_{t-1})|q'_{[0,t-1]}) = E\left[\int P_\mathcal{S}(dx_{t-1}|q'_{[0,t-1]}) \log_2(|J(f(x_{t-1}))|) \right] + h_\mathcal{S}(x_{t-1}|q'_{[0,t-1]}),$$

where the expectation is over the realizations of $q'_{[0,t-1]}$. Finally, we use the bound-edness of $h(x_0)$ (and thus $h(x_0|q'_0)$) in (B.8). Thus, with $V_\mathcal{S} \geq L$, it follows that

$$\text{(B.10)} \qquad\qquad \liminf_{T\to\infty} \left(\frac{1}{T} h_\mathcal{S}(x_{T-1}|q'_{[0,T-1]}) \right) \geq L - C.$$

We now seek to obtain an upper bound on $h_\mathcal{S}(x_{T-1}|q'_{[0,T-1]})$. As in [33], note that

$$h_\mathcal{S}(x_T|q'_{[0,T]}) \leq h_\mathcal{S}(x_T, \mathcal{Y}|q'_{[0,T]}),$$

where $\mathcal{Y}$ is a binary random variable which is 1 if $|x_T| \leq b(T)$ and 0 otherwise. Let

$$P_\mathcal{S}(\mathcal{Y} = 1) = P_\mathcal{S}(|x_T| \leq b(T)) =: p_T^\mathcal{S}.$$

Then,

$$h_\mathcal{S}(x_T, \mathcal{Y}|q'_{[0,T]}) = h_\mathcal{S}(x_T|q'_{[0,T]}, \mathcal{Y}) + H_\mathcal{S}(\mathcal{Y}|q'_{[0,T]})$$
$$\leq h_\mathcal{S}(x_T|q'_{[0,T]}, \mathcal{Y}) + 1,$$

since $\mathcal{Y}$ is binary. We have that

$$h_\mathcal{S}(x_T|q'_{[0,T]}, \mathcal{Y}) \leq p_T^\mathcal{S} \frac{n}{2} \log_2(2\pi e b^2(T))$$

$$+ (1 - p_T^\mathcal{S}) h_\mathcal{S}\left(x_T \middle| q'_{[0,T]}, |x_T| \geq b(T) \right)$$

and

$$h_{\mathcal{S}}\left(x_T \Big| q'_{[0,T]}, |x_T| > b(T)\right)$$

$$= h_{\mathcal{S}}\left(f(x_{T-1}) + Bu_{T-1} + w_{T-1}\Big| q'_{[0,T]}, |x_T| > b(T)\right)$$

$$(\text{B.11}) \quad \leq h_{\mathcal{S}}\left(f(x_{T-1}) + Bu_{T-1} + w_{T-1}\Big| q'_{[0,T-1]}, |x_T| > b(T)\right)$$

$$= h_{\mathcal{S}}\left(f(x_{T-1}) + w_{T-1}\Big| q'_{[0,T-1]}, |x_T| > b(T)\right)$$

$$(\text{B.12}) \quad = h_{\mathcal{S}}\left(f(x_{T-1})\Big| q'_{[0,T-1]}, |x_T| > b(T)\right)$$

$$= E\left[\int P_{\mathcal{S}}\left(dx_{T-1}\Big| q'_{[0,T-1]}, |x_T| > b(T)\right) \log_2\left(\Big| J(f(x_{T-1}))\Big|\right)\right]$$

$$(\text{B.13}) \qquad\qquad + h_{\mathcal{S}}\left(x_{T-1}\Big| q'_{[0,T-1]}, |x_T| > b(T)\right)$$

$$\leq M + h_{\mathcal{S}}\left(x_{T-1}\Big| q'_{[0,T-1]}, |x_T| > b(T)\right)$$

$$\vdots$$

$$(\text{B.14}) \quad \leq MT + h_{\mathcal{S}}\left(x_0\Big| |x_T| > b(T)\right).$$

Here (B.11) follows from that conditioning on a random variable reduces the differential entropy, and (B.12) follows due to the fact that $\mathcal{S}$ determines the noise realizations. We note that nonlinearity of f adds further technical issues when compared with the linear setup.[1] Here, M is the supremum of $\log_2(|J(f(x))|)$. In the above derivation in (B.13), we use the fact that f is invertible.

Thus, by (B.10)–(B.11) and (B.14) we have

$$\liminf_{T\to\infty} \frac{1}{T}\left(1 + (1 - p_T^{\mathcal{S}})\left(MT + h_{\mathcal{S}}(x_0\big||x_T| > b(T))\right) + p_T^{\mathcal{S}}\frac{n}{2}\log_2(2\pi e b^2(T))\right)$$

$$(\text{B.15}) \qquad\qquad\qquad \geq L - C,$$

Because given $\mathcal{S}, x_0$ is bounded by K, it follows that $h_{\mathcal{S}}(x_0\||x_T| > b(T)) \leq (n/2)$ $\log_2(2\pi e K^2)$, and thus for all K and η

$$\limsup_{T\to\infty} P_{\mathcal{S}_\eta^K}(|x_T| \leq b(T)) \leq \frac{M - (L - C)}{M}$$

for all $b(T)$ such that $\lim_{T\to\infty} \log_2(b(T))/T = 0$. But now

---

[1] Two technical intricacies here are as follows: For differential entropy (unlike discrete entropy) the relationship $h(x + y) \leq h(x) + h(y)$ does not in general hold for random variables x, y; this is why first a conditioning on $\mathcal{S}$ is taken in the proof. Furthermore, we cannot obtain an upper bound by taking out the conditioning on the event $|x_T| > b(T)$, since conditioning on a single event may decrease or increase entropy; note that conditioning on a random variable, however, does not increase the entropy.

$$\limsup_{T\to\infty} P(|x_T| \le b(T))$$

$$\le \limsup_{T\to\infty} P(|x_T| \le b(T), |x_0| \le K) + \limsup_{T\to\infty} P(|x_T| \le b(T), |x_0| \ge K)$$

$$\le \limsup_{T\to\infty} P(|x_T| \le b(T), |x_0| \le K) + P(|x_0| \ge K)$$

$$= \limsup_{T\to\infty} \int P(d\eta) P_{\mathcal{S}_\eta^K}(|x_T| \le b(T)) + P(|x_0| \ge K)$$

$$(\text{B.16}) \quad \le \int P(d\eta) \limsup_{T\to\infty} P_{\mathcal{S}_\eta^K}(|x_T| \le b(T)) + P(|x_0| \ge K)$$

$$\le \int P(d\eta) \frac{M - (L - C)}{M} + P(|x_0| \ge K)$$

$$= \frac{M - (L - C)}{M} + P(|x_0| \ge K),$$

where we use Fatou's lemma in (B.16) and the fact that (B.15) holds for every restriction of the noise realizations η and K values. Since an individual probability measure is tight, $\lim_{K\to\infty} P(|x_0| \ge K) = 0$, the right-hand side can be made arbitrarily close to $\frac{M-(L-C)}{M}$ and the result follows.

REFERENCES

[1] J. Baillieul, *Feedback designs for controlling device arrays with communication channel bandwidth constraints*, in Proceedings of the 4th ARO Workshop on Smart Structures, State College, PA, 1999.

[2] J. Baillieul, *Data-rate requirements for nonlinear feedback control*, in Proceedings of the 6th IFAC Symposium on Nonlinear Control Systems, Stuttgart, Germany, 2004, pp. 1277–1282.

[3] P. Billingsley, *Convergence of Probability Measures*, Wiley, New York, 1968.

[4] R. Brockett and D. Liberzon, *Quantized feedback stabilization of linear systems*, IEEE Trans. Automat. Control, 45 (2000), pp. 1279–1289.

[5] F. Colonius, *Metric invariance entropy and conditionally invariant measures*, Ergodic Theory Dynam. Systems, to appear.

[6] F. Colonius and C. Kawan, *Invariance entropy for control systems*, SIAM J. Control Optim., 48 (2009), pp. 1701–1721.

[7] F. Colonius and C. Kawan, *Invariance entropy for outputs*, Math. Control Signals Systems, 22 (2011), pp. 203–227.

[8] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, New York, Wiley, 1991.

[9] L. Coviello, P. Minero, and M. Franceschetti, *Stabilization over Markov feedback channels: The general case*, in Proceedings of the IEEE Conference on Decision and Control, 2011, pp. 3776–3782.

[10] N. Şen, F. Alajaji, and S. Yüksel, *Feedback capacity of a class of symmetric finite-state Markov channels*, IEEE Trans. Inform. Theory, 56 (2011), pp. 4110–4122.

[11] R. Dabora and A. Goldsmith, *On the capacity of indecomposable finite-state channels with feedback*, in Proceedings of the Allerton Confernce on Communication, Control, and Computing, 2008, pp. 1045–1052.

[12] N. Elia, *When Bode meets Shannon: control-oriented feedback communication schemes*, IEEE Trans. on Automat. Control, 49 (2004), pp. 1477–1488.

[13] M. Franceschetti and P. Minero, *Elements of information theory for networked control systems*, in Information and Control in Networks, Springer, New York, 2014, pp. 3–37.

[14] A. E. Gamal and Y. H. Kim, *Network Information Theory*, Cambridge University Press, Cambridge, UK, 2012.

[15] R. M. Gray, *Probability, Random Processes, and Ergodic Properties*, Springer-Verlag, New York, 1988.

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

[17] R. M. GRAY AND L. D. DAVISSON, *Source coding theorems without the ergodic assumption*, IEEE Trans. Inform. Theory, 20 (1974), pp. 502–516.

[18] R. M. GRAY AND T. HASHIMOTO, *A note on rate-distortion functions for nonstationary Gaussian autoregressive processes*, IEEE Trans. Inform. Theory, 54 (2008), pp. 1319–1322.

[19] R. M. GRAY AND J. C. KIEFFER, *Asymptotically mean stationary measures*, Ann. Probab., 8 (1980), pp. 962–973.

[20] O. HERNANDEZ-LERMA AND J. B. LASSERRE, *Markov Chains and Invariant Probabilities*, Birkhäuser, Basel, 2003.

[21] A. JOHNSTON AND S. YÜKSEL, *Stochastic stabilization of partially observed and multi-sensor systems driven by unbounded noise under fixed-rate information constraints*, IEEE Trans. Automat. Control, 59 (2014), pp. 792–798.

[22] E. A. JONCKHEERE, A. A. HAMMAD, AND B. F. WU, *Chaotic disturbance rejection a Kolmogorov-Sinai entropy approach*, in Proceedings of the IEEE Conference Decision and Control, San Antonio, Texas, 1993.

[23] A. KATOK, *Fifty years of entropy in dynamics:* 1958–2007, J. Mod. Dyn, 1 (2007), pp. 545–596.

[24] C. KAWAN, *Invariance Entropy for Deterministic Control Systems*, Springer, New York, 2013.

[25] J. C. KIEFFER, *Zero-error stationary coding over stationary channels*, Z. Wahrscheinlichkeitstheorie Verwandte Gebiete, 56 (1981), pp. 113–126.

[26] J. C. KIEFFER, *Stochastic stability for feedback quantization schemes*, IEEE Trans. Inform. Theory, 28 (1982), pp. 248–254.

[27] J. C. KIEFFER AND J. G. DUNHAM, *On a type of stochastic stability for a class of encoding schemes*, IEEE Trans. Inform. Theory, 29 (1983), pp. 793–797.

[28] D. LIBERZON, *Stabilization by quantized state or output feedback: A hybrid control approach*, in Proceedings of the IFAC 15th Triennial World Congress, 2002.

[29] D. LIBERZON AND J. P. HESPANHA, *Stabilization of nonlinear systems with limited information feedback*, IEEE Trans. Automat. Control, 50 (2005), pp. 910–915.

[30] T. LINDER AND S. YÜKSEL, *On optimal zero-delay quantization of vector Markov sources*, IEEE Trans. Inform. Theory, 60 (2014), pp. 2975–5991.

[31] N. C. MARTINS AND M. A. DAHLEH, *Feedback control in the presence of noisy channels: Bode-like fundamental limitations of performance*, IEEE Trans. Automt. Control, 53 (2008), pp. 1604–1615.

[32] N. C. MARTINS, M. A. DAHLEH, AND N. ELIA, *Feedback stabilization of uncertain systems in the presence of a direct link*, IEEE Trans. Automat. Control, 51 (2006), pp. 438–447.

[33] A. S. MATVEEV, *State estimation via limited capacity noisy communication channels*, Math. Control Signals Systems, 20 (2008), pp. 1– 35.

[34] A. S. MATVEEV AND A. SAVKIN, *An analogue of Shannon information theory for detection and stabilization via noisy discrete communication channels*, SIAM J. Control Optim. 46 (2007), pp. 1323–1367.

[35] A. S. MATVEEV AND A. V. SAVKIN, *Estimation and Control over Communication Networks*, Birkhäuser, Boston, 2008.

[36] P. G. MEHTA, U. VAIDYA, AND A. BANASZUK, *Markov chains, entropy, and fundamental limitations in nonlinear stabilization*, IEEE Trans. Automat. Control, 53 (2008), pp. 784–791.

[37] S. P. MEYN AND R. TWEEDIE, *Markov Chains and Stochastic Stability*, Springer-Verlag, London, 1993.

[38] R. H. MIDDLETON, *Trade-offs in linear control system design*, Automatica, 27 (1991), pp. 281–292.

[39] P. MINERO, M. FRANCESCHETTI, S. DEY, AND G. N. NAIR, *Data rate theorem for stabilization over time-varying feedback channels*, IEEE Trans. Automat. Control, 54 (2009), pp. 243–255.

[40] G. N. NAIR AND R. J. EVANS, *Stabilizability of stochastic linear systems with finite feedback data rates*, SIAM J. Control Optimi. 43 (2004), pp. 413–436.

[41] G. N. NAIR, R. J. EVANS, I. M. Y. MAREELS, AND W. MORAN, *Topological feedback entropy and nonlinear stabilization*, IEEE Trans. Automat. Control, 49 (2004), pp. 1585–1597.

[42] G. N. NAIR, F. FAGNANI, S. ZAMPIERI, AND J. R. EVANS, *Feedback control under data constraints: An overview*, Proc. IEEE, (2007), pp. 108–137.

[43] K. OKANO, S. HARA, AND H. ISHII, *Characterization of a complementary sensitivity property in feedback control: An information theoretic approach*, Automatica, 45 (2009), pp. 504–509.

[44] D. ORNSTEIN, *An application of ergodic theory to probability theory*, Ann. Probab. 1 (1973), pp. 43–58.

[45] H. H. PERMUTER, T. WEISSMAN, AND A. J. GOLDSMITH, *Finite state channels with time-invariant deterministic feedback*, IEEE Trans. Inform. Theory, 55 (2009), pp. 644–662.

[46] C. D. PERSIS AND A. ISIDORI, *Stabilizability by state feedback implies stabilizability by encoded state feedback*, Systems Control Lett. 53 (2004), pp. 249–258.

[47] F. RAMPONI, A. M.-A. D. CHATTERJEE, P. HOKAYEM, AND J. LYGEROS, *Attaining mean square boundedness of a marginally stable stochastic linear system with a bounded control input*, IEEE Trans. Automat. Control, 55 (2010), pp. 2414–2418.

[48] A. SAHAI, *Anytime Information Theory*, Ph.D. dissertation, Massachusetts Institute of Technology, Cambridge, MA, 2001.

[49] 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. Inform. Theory, 52 (2006), pp. 3369–3395.

[50] P. C. SHIELDS, *The Ergodic Theory of Discrete Sample Paths*, AMS, Providence, RI, 1996.

[51] P. C. SHIELDS, *The interactions between ergodic theory and information theory*, IEEE Trans. Inform. Theory, 44 (1998), pp. 2079 – 2093.

[52] A. D. SILVA, *Invariance entropy for random control systems*, Math. Control Signals Systems, 25 (2013), pp. 491–516.

[53] A. D. SILVA AND C. KAWAN, *Invariance entropy of hyperbolic control sets*, Discrete Contin. Dyn. Syst., 36 (2016), pp. 97–136.

[54] H. STARK AND J. W. WOODS, *Probability, Statistics, and Random Processes for Engineers*, Pearson, Upper Saddle River, NJ, 2012.

[55] S. TATIKONDA, *Control Under Communication Constraints*, Ph.D. dissertation, Massachusetts Institute of Technology, Cambridge, MA, 2000.

[56] S. TATIKONDA AND S. MITTER, *Control under communication constraints*, IEEE Trans. Automat. Control, 49 (2004), pp. 1056–1068.

[57] S. TATIKONDA AND S. MITTER, *The capacity of channels with feedback*, IEEE Trans. Inform. Theory, 55 (2009), pp. 323–349.

[58] U. VAIDYA AND N. ELIA, *Stabilization of nonlinear systems over packet-drop links: Scalar case*, Systems Control Lett. 61 (2012), pp. 959–966.

[59] U. VAIDYA AND N. ELIA, *Limitations for nonlinear stabilization over uncertain channels*, preprint, arXiv:1605.04432, 2016.

[60] S. VEMBU, S. VERDÚ, AND Y. STEINBERG, *The source-channel separation theorem revisited*, IEEE Trans. Inform. Theory, 41 (1995), pp. 44–54.

[61] W. S. WONG AND R. W. BROCKETT, *Systems with finite communication bandwidth constraints—part* II*: Stabilization with limited information feedback*, IEEE Trans. Automat. Control, 42 (1997), pp. 1294–1299.

[62] L.-S. YOUNG, *Entropy in dynamical systems*, in Entropy, A. Greven, G. Keller, and G. Warnecke, eds. Princeton University Press, Princeton, NJ, 2003, pp. 313–327.

[63] S. YU AND P. G. MEHTA, *Bode-like fundamental performance limitations in control of nonlinear systems*, IEEE Trans. Automat. Control, 55 (2010), pp. 1390–1405.

[64] S. YÜKSEL, *Stochastic stabilization of noisy linear systems with fixed-rate limited feedback*, IEEE Trans. Automat. Control, 55 (2010), pp. 2847–2853.

[65] S. YÜKSEL, *Characterization of information channels for asymptotic mean stationarity and stochastic stability of non-stationary/unstable linear systems*, IEEE Trans. Inform. Theory, 58 (2012), pp. 6332–6354.

[66] S. YÜKSEL AND T. BAŞAR, *Control over noisy forward and reverse channels*, IEEE Trans. Automat. Control, 56 (2011), pp. 1014–1029.

[67] S. YÜKSEL AND T. BAŞAR, *Stochastic Networked Control Systems: Stabilization and Optimization Under Information Constraints*, Birkhäuser, New York, 2013.

[68] S. YÜKSEL AND S. P. MEYN, *Random-time, state-dependent stochastic drift for Markov chains and application to stochastic stabilization over erasure channels*, IEEE Trans. Automat. Control, 58 (2013), pp. 47–59.

[69] H. ZANG AND P. A. IGLESIAS, *Nonlinear extension of Bode's integral based on an information-theoretic interpretation*, Systems Control Lett. 50 (2003), pp. 11–19.