Ergodicity of Controlled Stochastic Nonlinear Systems under Information Constraints: Refined Bounds via Splitting

Nicolas Garcia[1], Christoph Kawan[2], and Serdar Yüksel[3]

Abstract— This paper considers the problem of stabilizing a discrete-time non-linear stochastic system over a finite capacity noiseless channel. Our focus is on systems which decompose into a stable and unstable component, and the stability notion considered is asymptotic ergodicity of the $\mathbb{R}^N$-valued state process. We establish a necessary lower bound on channel capacity for the existence of a coding and control policy which renders the closed-loop system stochastically stable. In the literature, it has been established that under technical assumptions, the channel capacity must not be smaller than the logarithm of the determinant of the system linearization, averaged over the noise and ergodic state measures. In this paper, we establish that for systems with a stable component, it suffices to consider only the unstable dimensions, providing a refinement on the general channel capacity bound for a large class of systems. The result is established using the notion of stabilization entropy, a notion adapted from invariance entropy, used in the study of noise-free systems under information constraints.

I. INTRODUCTION

In the field of control under communication constraints, a commonly studied problem is to characterize the minimum amount of information required by a controller in order to achieve a given control task. In this paper we consider the above problem for discrete-time non-linear stochastic systems with additive control. The control objective considered is to render the state process stochastically stable for the stability criterion of asymptotic ergodicity. More precisely, we consider the system

$$x_{t+1} = f(x_t, w_t) + Bu_t \tag{1}$$

where x_t, w_t and u_t are the state, noise, and control at time t respectively and B is an appropriately sized matrix. Additionally, we impose that the state information travel through a finite capacity noiseless channel at each time step before reaching the controller, as depicted in Figure 1. We formalize the notion of a coding and control policy as follows. First, let $\mathcal{M} := \{1, 2, .., M\}$ denote the alphabet of the channel, thus its capacity in bits is given by $C := \log_2 M$. At time t, the coder (also known as the encoder) generates a channel input q_t from past state realizations $x_0, .., x_t$. The channel input $q_t \in \mathcal{M}$ is therefore determined by a map

$\gamma_t^e : (\mathbb{R}^N)^{t+1} \to \{1, 2, \ldots, M\}$. The symbol q_t is transmitted over the channel, reaching the controller. The controller generates u_t based on channel outputs $q_0, .., q_t$ according to a map $\gamma_t^c : \mathcal{M}^{t+1} \to \mathbb{R}^N$. A coding and control policy is therefore a pair of maps $(\gamma_t^e)_{t \in \mathbb{N}}$ and $(\gamma_t^c)_{t \in \mathbb{N}}$. Once we fix a coding and control policy, $(x_t)_{t \in \mathbb{N}}$ is a well defined autonomous stochastic process, with randomness coming from the possibly random initial state x_0, and the noise process $(w_t)_{t \in \mathbb{N}}$.
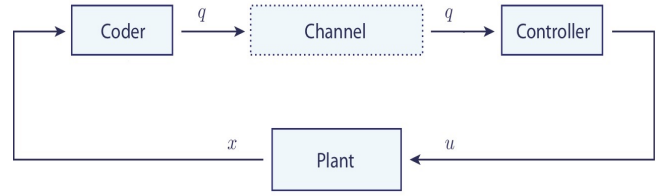

Fig. 1. System Controlled Over Communication Channel

In this paper, we establish a necessary lower bound on channel capacity required for the existence of coding and control policies which result in the state process $(x_t)_{t \in \mathbb{N}}$ being asymptotically ergodic. The lower bound is related to the asymptotic mean (defined in the next section), i.i.d. noise law, and system dynamics function. We now discuss the stability notion in detail, but first provide some notational conventions.

A. Notation

Throughout this paper, $\mathbb{Z}$ denotes the integers, $\mathbb{R}$ the real numbers, and $\mathbb{N}$ the non-negative integers. The Lebesgue measure is denoted by m, where the dimension will be clear from context. A discrete interval in the integers will be denoted by $[a; b]$ (i.e., $[a; b] = \{a, a+1, \ldots, b-1, b\}$ for $a \leq b$ in $\mathbb{Z}$). Given a topological space $\mathcal{X}$, $\mathcal{B}(\mathcal{X})$ denotes its Borel σ-algebra. For a function $f : \mathbb{R}^n \to \mathbb{R}^m$, we denote the Jacobian (matrix of partial derivatives) by Df. We use $\sqcup$ to emphasize that a union in question is disjoint. When applied to a set, $|\cdot|$ denotes cardinality. Given a sequence $x := (x_n)_{n \in \mathbb{N}}$ taking values in a given space, θ denotes the left shift map, so that $(\theta x)_n = x_{n+1}$ for every $n \in \mathbb{N}$. Given a topological space $\mathcal{X}$, we let $\mathcal{X}^{\mathbb{N}}$ denote the set of sequences indexed by $\mathbb{N}$, taking values in $\mathcal{X}$. We endow $\mathcal{X}^{\mathbb{N}}$ with the product topology.

B. Asymptotic Ergodicity

In this section we discuss the stochastic stability notion considered in this paper. Let $(\Omega, \mathcal{F}, P)$ denote the common probability space on which all random variables are defined. Note that fixing a coding and control policy and specifying an initial state distribution for the system (1) results in an autonomous state process $(x_t)_{t \in \mathbb{N}}$ which induces a process measure on $\mathcal{B}((\mathbb{R}^N)^{\mathbb{N}})$. Let us denote this measure by μ.

Definition 1: Consider the process $(x_t)_{t \in \mathbb{N}}$ with process measure μ as above. We say it is:

- *stationary* iff $\mu(\theta^{-1}(A)) = \mu(A)$ for all $A \in \mathcal{B}((\mathbb{R}^N)^{\mathbb{N}})$.
- *asymptotically mean stationary (AMS)* iff there exists a probability measure Q (called the asymptotic mean of the process) on $\mathcal{B}((\mathbb{R}^N)^{\mathbb{N}})$ such that

$$\lim_{T \to \infty} \frac{1}{T} \sum_{k=0}^{T-1} \mu(\theta^{-k}(A)) = Q(A) \quad \text{for every } A \in \mathcal{B}((\mathbb{R}^N)^{\mathbb{N}}).$$

- *ergodic* iff it is stationary, and for $A \in \mathcal{B}((\mathbb{R}^N)^{\mathbb{N}})$ we have that $A = \theta^{-1}(A) \implies \mu(A) \in \{0,1\}$.
- *AMS ergodic* iff it is AMS, and the asymptotic mean is ergodic.

Note that if a process is *AMS*, then the asymptotic mean is a stationary measure on the sequence space. Note also that a stationary measure on $\mathcal{B}((\mathbb{R}^N)^{\mathbb{N}})$ can unambiguously be projected to a measure on $\mathcal{B}(\mathbb{R}^N)$. By slight abuse of notation, we do not distinguish between a stationary measure on the sequence space and its projected coordinate measure, as the measure in question will be clear from context. In this paper, the stability notion that we will consider is *AMS ergodicity* (or informally, *asymptotic ergodicity*). Ergodicity allows us to take advantage of the following pointwise ergodic theorem:

Theorem 1: (Pointwise Ergodic Theorem) Let $(\Omega, \mathcal{F}, P, T)$ be an ergodic system. Then for any $f \in L^1(\Omega, \mathcal{F}, P)$ we have

$$\frac{1}{N} \sum_{k=0}^{N-1} f \circ T^k \xrightarrow[N \to \infty]{a.s} \int f \, dP.$$

For a proof, see [1, Theorem 1.14].

Suppose that the stochastic process in Definition 1 is *AMS ergodic* with process measure μ and asymptotic mean Q. The application of the pointwise ergodic theorem to the L^1 map $\mathbb{1}_{x_0 \in B} : (\mathbb{R}^N)^{\mathbb{N}} \to \{0,1\}$ for $B \in \mathcal{B}(\mathbb{R}^N)$ yields

$$\mu\left(\left\{x \in (\mathbb{R}^N)^{\mathbb{N}} : \lim_{T \to \infty} \frac{1}{T} \sum_{k=0}^{T-1} \mathbb{1}_B(x_k) = Q(B)\right\}\right) = 1, \quad (2)$$

In principle, the pointwise ergodic theorem tells us that the above set has full measure with respect to Q. From [2, Lem. 7.5 and Eq. (7.22)] it turns out that μ and Q agree on all Q-trivial sets, thus allowing one to write (2). The above equation which provides almost sure guarantees on asymptotic sample path behavior will be crucial in the results of this paper, for which we now provide some motivating examples.

C. Problem Motivation

Suppose that system (1) is controlled over a noiseless channel of finite capacity C, and made to be asymptotically ergodic with AMS mean Q. Then under slightly different technical assumptions, [3] and [4] establish that

$$\int \int \log_2 |\det D f_w(x)| dQ(x) d\nu(w) \leq C \quad (3)$$

where ν is the distribution of the i.i.d. noise and f_w denotes the map $x \mapsto f(x,w)$ for some fixed $w \in \mathcal{W}$. This bound however is in general not tight, as the following two examples illustrate.

Example 1: Consider the two dimensional linear system given by

$$\begin{bmatrix} x_{t+1} \\ y_{t+1} \end{bmatrix} = \begin{bmatrix} 2 & 0 \\ 0 & 1/2 \end{bmatrix} \begin{bmatrix} x_t \\ y_t \end{bmatrix} + w_t + u_t \quad (4)$$

where u_t and w_t take values in $\mathbb{R}^2$ and the noise is i.i.d with zero mean. The LHS of (3) is easily seen to equal zero for this system, thus providing a vacuous bound on channel capacity. It is well known from the literature however that a tight bound on linear systems is the log-sum of the unstable eigenvalues. Note that by replacing $1/2$ in the above matrix with any number no smaller than one, the bound in (3) recovers the tight linear bound. As we will see, the refinement of the channel capacity bound in this paper will recover the tight bound in the general linear case (thus, also with stable eigenvalues). Consider now a second example:

Example 2: Consider the system $(x_t, y_t)_{t \in \mathbb{N}}$ in $\mathbb{R}^2$ evolving with scalar-valued i.i.d. noise according to

$$\begin{bmatrix} x_{t+1} \\ y_{t+1} \end{bmatrix} = \begin{bmatrix} (x_t^3 + x_t)(1 + y_t^2) \\ \frac{1}{2} y_t + w_t \end{bmatrix} + u_t \quad (5)$$

with x_0 and y_0 independent and admitting bounded densities. We note that the y-component of the above system is stochastically stable. Moreover, the presence of the y-term in the dynamics of the x-component cannot be modeled as noise, as the i.i.d. assumption required in data-rate theorems of nonlinear systems is not satisfied. Suppose the above system is made asymptotically ergodic via a coding and control policy with AMS mean Q. We compute

$$D f_w(x,y) = \begin{bmatrix} (3x^2 + 1)(1 + y^2) & (x^3 + x)2y \\ 0 & 1/2 \end{bmatrix} \quad (6)$$

and apply (3) to obtain

$$\int \log_2 \left| \frac{(1 + y^2)}{2} (3x^2 + 1) \right| dQ(x,y) \leq C. \quad (7)$$

Note that there is a factor of $1/2$ coming from the stable second component in the integrand. It seems sensible that the bound should hold without this factor, as the coding and control policy need not be concerned with the stochastically stable component. Indeed, the result in this paper establishes that the above bounds holds when removing the factor of $1/2$ and is therefore a strict refinement for certain systems. We now move on to a literature review, followed by our main result and its proof.

II. Literature Review and Information Requirements for Stochastic Stability

The presence of real-world control problems where perfect and instantaneous state information is not necessarily available to a controller has motivated the field of control under communication constraints. In this field, one wishes to study if and how it is possible to accomplish a control task under varying degrees of imperfect information. A ubiquitous problem in the field is to characterize minimum data rates required to stabilize a dynamical system. This problem has been considered extensively for linear deterministic and stochastic systems, for which one can usually characterize the minimum data rate required for closed-loop stability as the log-sum of the unstable open-loop eigenvalues.

Some related earlier papers considering the linear case include [5], [6], [7], [8], [9], [10] and [11]. More recent contributions include [12], [13], [14], [15], [16], [17], and [18], [19], [20], [21], [22] where this latter group of papers presented necessary and sufficient conditions for stability criteria such as existence of invariant measures, positive Harris recurrence and (asymptotic) ergodicity. There has been a separate line of work for the special Gaussian channel setup, which we do not review in this paper.

For non-linear systems, however, the majority of papers have focused on deterministic systems. Some early works include [23], where it was established that global asymptotic stabilization of a non-linear continuous time system is feasible provided that data rates exceed a quantity related to system dimension and a Lipschitz constant, and [24] where non-linear feed-forward systems were considered. In [25] the authors presented the first systematic approach for determining minimal data rates for stabilization and introduced the notion of topological feedback entropy, a notion inspired by the classical open cover definition of topological entropy in dynamical systems due to Adler et al. [26]. It was established in [25] that a necessary and sufficient condition for stabilization to a compact set is the condition that the data rate in the control loop exceeds the topological feedback entropy. For the same stabilization problem, invariance entropy was introduced in [27]. This notion serves as a way to quantify the difficulty of a control task through the minimum number of open loop control sequences required to achieve it. The monograph [28] provides a detailed account of the applications of invariance entropy in determining minimum data rates, particularly for continuous time (non-linear) systems. In [29], it was further established that under a strong invariance condition, the notions of topological feedback entropy and invariance entropy coincide in the discrete time case. A recent related development was the introduction of metric invariance entropy in [30]. Many more interesting results have been obtained under a wealth of setting, and we refer the reader to [31] and [32] for a more detailed overview of the literature.

To the best of our knowledge, the first converse result on channel capacity for non-linear stochastic systems was established in [4] using information theoretic methods. The paper provided lower bounds on channel capacity necessary for stochastic stabilization of discrete time non-linear systems over both noisy and noiseless channels for stability notions of ergodicity and entropy growth conditions. With a fundamentally different approach via stochastic growth properties, for the ergodic case a similar result was established in [3], which relied instead on stabilization entropy. This notion, introduced in [33], was a modification of invariance entropy for the stochastic case and was first used to obtain lower bounds on channel capacity required for AMS stability. The paper at hand builds on the techniques involving stabilization entropy, and provides a refinement for the lower bound in [3, Theorem 4.1] and [4, Theorem 4.2] for the stability notion of (asymptotic) ergodicity. In particular, the statement of the result in this paper resembles [3, Theorem 4.1] and a similar approach is used. Our main result is a strict refinement, and the proof requires a modification of stabilization entropy.

III. Result

Consider a subset $p \subseteq \{1, .., N\}$ of indices listed in increasing order as $p_1 < p_2 < \cdots < p_{|p|}$. Let $z_1 < \cdots < z_{N-|p|}$ denote the elements in $\{1, .., N\} \setminus p$. We define the permutation $\psi_p : \mathbb{R}^N \to \mathbb{R}^N$ by

$$\psi_p(x_1, ..., x_N)_i = \begin{cases} x_{p_i} & i \le |p| \\ x_{z_{i-|p|}} & i > |p| \end{cases}$$

for $i \in \{1, .., N\}$. Also, let $\pi_p : \mathbb{R}^N \to \mathbb{R}^{|p|}$ denote the natural projection of coordinates $p_1, .., p_{|p|}$. For a map $f : \mathbb{R}^N \to \mathbb{R}^N$, a set p as above, and a fixed vector $(y_1, .., y_{N-|p|})$ we define the map $f^p(\cdot, y_1, .., y_{N-|p|}) : \mathbb{R}^{|p|} \to \mathbb{R}^{|p|}$ by

$$f^p(x, y_1, .., y_{N-|p|}) := \pi_p(f(\psi_p^{-1}(x, y_1, .., y_{N-|p|}))) \quad (8)$$

where $x \in \mathbb{R}^{|p|}$. As an example, consider $N = 4, p = \{2, 4\}$, a fixed vector (y_1, y_2), and a function $f : \mathbb{R}^4 \to \mathbb{R}^4$ written as $f = (f_1, f_2, f_3, f_4)$ for maps $f_i : \mathbb{R}^4 \to R$. Then

$$f^p(x_1, x_2, y_1, y_2) = (f_2(y_1, x_1, y_2, x_2), f_4(y_1, x_1, y_2, x_2)).$$

This notation allows us to precisely state our main result. Consider the system

$$x_{t+1} = f(x_t, w_t) + Bu_t \quad (9)$$

where x_t is $\mathbb{R}^N$-valued for some $N \in \mathbb{N}$, $B \in \mathbb{R}^{N \times N'}$, u_t is $\mathbb{R}^{N'}$-valued, and w_t takes values in a standard probability space $\mathcal{W}$. For a fixed $w \in \mathcal{W}$, let us denote the map $x \mapsto f(x, w)$ by f_w. Suppose that the following holds:

(i) The state evolution map f is Borel measurable.
(ii) The noise process $(w_t)_{t \in \mathbb{N}}$ is i.i.d. By abuse of notation, ν denotes both the i.i.d. measure on $\mathcal{B}(\mathcal{W})$ and the noise process measure on $\mathcal{B}(\mathcal{W}^{\mathbb{N}})$.
(iii) The map $f_w(\cdot) : \mathbb{R}^N \to \mathbb{R}^N$ is C^1 and injective for any $w \in \mathcal{W}$.
(iv) The initial state $x_0 \in \mathbb{R}^N$ is random and independent of the noise process, and its law π_0 admits a bounded density.

(v) The set $\Gamma = \Big\{ p \subseteq \{1,..,N\} : \exists c_p > 0$ such that $|\det D f_w^p(x_{p_1},..,x_{p_{|p|}},x_{z_1},..,x_{z_{N-|p|}})| > c_p, \forall x \in \mathbb{R}^N, w \in \mathcal{W} \Big\}$ is non-empty.

(vi) Let $\{t_1,..,t_s\} \in \Gamma$ be arbitrary and write the random initial state as $x_0 = (x_0^1,..,x_0^N)$. Then there exists a set S consisting of possible realizations of the initial state components not indexed by Γ such that the event that these non-indexed initial states take a realization in S has non-zero probability, and the law of $(x_0^{t_1},...,x_0^{t_s})$ admits a bounded density when conditioned on the event that the non-indexed initial states take on a realization $\hat{x}$, for any $\hat{x} \in S$.

Theorem 2: Consider system (9) satisfying assumptions (i)–(vi), controlled over a noiseless channel with finite alphabet $\mathcal{M}$ and capacity $C := \log_2 |\mathcal{M}|$. if there exists a coding and control policy which renders the state process $(x_t)_{t \in \mathbb{N}}$ AMS ergodic (asymptotically ergodic) with asymptotic mean Q, then we must have that

$$\max_{p \in \Gamma} \int \int \log |\det D f_w^p(x_{p_1},..,x_{p_{|p|}},x_{z_1},..,x_{z_{N-|p|}})| \quad (10)$$

$$dQ(x_1,..,x_N)d\nu(w) \le C.$$

where the Jacobian above is the $|p| \times |p|$ matrix of partial derivatives of $f_w^p(\cdot,x_{z_1},..,x_{z_{N-|p|}})$ evaluated at $(x_{p_1},..,x_{p_{|p|}}) \in \mathbb{R}^{|p|}$.

Remark 1: Observe that by taking $p = \{1,..,M\}$ (if $\{1,..,M\} \in \Gamma$), we recover the bound (3) established previously in [3] and [4]. For a large class of systems however, it is clear that Theorem 2 is a strict refinement, as can be seen by noting that in Example 2, taking $p = \{1\}$ recovers the sharper bound

$$\int \log_2 |(1+y^2)(3x^2+1)|dQ(x,y) \le C. \quad (11)$$

It is clear that for linear system, the new bound recovers the tight linear bound such as in Example 1.

Remark 2: Note also that the technical assumption (vi) is satisfied if the initial state has independent components each admitting a bounded density. Suppose now that for a given system, the assumptions (i)-(v) of Theorem 2 are satisfied, but (vi) only hold for certain subsets of Γ. Then the theorem will still hold, however the max in (10) should be taken only over subsets of Γ for which the assumption (vi) holds. This last observation will become clear from the proof.

Remark 3: Noting that the Jacobian determinant is invariant under a linear change of coordinates, we note that the bound in (2) is invariant under a linear change of coordinates. Note however that assumption (v) is not coordinate independent; it is not hard to see that for certain systems, the choice of coordinates may result in a different (or empty) set Γ. Under a non-linear coordinate change, it is not clear if the above bound is invariant (or if control even remains additive), thus a possible future research direction is to consider the problem of optimizing the coordinate system chosen in order to maximize the bound.

IV. Proofs

We first fix an integer $m \le M$ and view the map f_w as a function of two vectors, i.e. we decompose the state into a pair (x,y) where $x \in \mathbb{R}^m$ and $y \in \mathbb{R}^{N-m}$. Consider the control system in Theorem 2 and note that for a fixed sequence of controls $u := (u_t)_{t \in \mathbb{N}}$, a fixed sequence of noise symbols $w := (w_t)_{t \in \mathbb{N}}$, and a fixed initial state $x_0 \in \mathbb{R}^N$, the state process $(x_t)_{t \in \mathbb{N}}$ is deterministic. Let us introduce the notation $\varphi(t,x_0,u,w) := x_t$ for every $t \in \mathbb{N}$. Letting π_m and π_{N-m} denote the natural projection of $\mathbb{R}^N$ on to the first m and last $N-m$ coordinates respectively, we further define $\varphi^m(t,x_0,u,w) := \pi_m(x_t)$ and $\varphi^{N-m}(t,x_0,u,w) := \pi_{N-m}(x_t)$ so that

$$\varphi(t,x_0,u,w) = (\varphi^m(t,x_0,u,w),\varphi^{N-m}(t,x_0,u,w)) \text{ for every } t \in \mathbb{N}.$$

We now provide a sketch of the proof, which relies on the notion of stabilization entropy and an associated lemma relating it to channel capacity. Compared to [3], we consider a version of stabilization entropy with an additional collection of sets since we are decomposing the state space into two components. The definition follows:

A. Stabilization Entropy

Definition 2: Let $(D_j)_{j=1}^{j=d} \subseteq \mathcal{B}(\mathbb{R}^m)$, $(E_k)_{k=1}^{k=e} \in \mathcal{B}(\mathbb{R}^{N-m})$ and $(F_l)_{l=1}^{f} \in \mathcal{B}(\mathcal{W})$ be finite disjoint unions of Borel sets and define

$$D := \bigsqcup_{j=1}^d D_j \qquad E := \bigsqcup_{k=1}^e E_k \qquad F := \bigsqcup_{l=1}^f F_l.$$

Let also R denote a collection of numbers $r_{j,k,l} \in [0,1]$ for $j \in \{1,\ldots,d\}$, $k \in \{1,\ldots,e\}$ and $l \in \{1,\ldots,f\}$ satisfying

$$1-r := \sum_{j=1}^d \sum_{k=1}^e \sum_{l=1}^f (1-r_{j,k,l}) \in [0,1]$$

and fix $T \in \mathbb{N}$ and $\rho \in (0,1)$. A set $S \subseteq (\mathbb{R}^N)^T$ of control sequences of length T is called (T,D,E,F,ρ,R)-spanning iff there exists $\tilde{\Omega} \in \mathcal{F}$ such that the following conditions:

- $P(\tilde{\Omega}) \ge 1-\rho$.
- For each $\omega \in \tilde{\Omega}$, there exists a control sequence $u \in S$ such that

$$\frac{1}{T}|\{t \in [0;T-1] : (\varphi^m(t,x_0(\omega),u,w(\omega)),\varphi^{N-m}(t,x_0(\omega),u,$$

$$w(\omega)),w_t(\omega)) \in D_j \times E_k \times F_l\}| \ge 1-r_{j,k,l}$$

for all j, k and l.

both hold.

We slightly abuse notation writing (T,D,E,F,ρ,R)-spanning instead of $(T,(D_j)_{j=1}^d,(E_k)_{k=1}^e,(F_l)_{l=1}^f,\rho,R)$-spanning. Whenever we do this however, the specific sequences of sets making up the disjoint unions will be clear from context. We will use the size of spanning sets to quantify the difficulty of a control task. This leads to:

Definition 3: (Stabilization Entropy) For the system (9), and sequences of sets as in Definition 2, we define the (D, E, F, ρ, R)-stabilization entropy by

$$h(D, E, F, \rho, R) := \limsup_{T \to \infty} \frac{1}{T} \log s(T, D, E, F, \rho, R),$$

where $s(T, D, E, F, \rho, R)$ denotes the smallest cardinality of a (T, D, E, F, ρ, R)-spanning set. We define this quantity to be ∞ if no finite spanning set exists.

Finite (T, D, E, F, ρ, R)-spanning sets need not exist in general but as we will shortly see, they exist in desired scenarios. The following lemma relates stabilization entropy with channel capacity.

Lemma 1: Consider system (9) with the assumptions of Theorem 2 (i.e., a coding and control policy exists over a noiseless channel of capacity $C = \log_2 |\mathcal{M}|$ which makes the state process AMS ergodic with asymptotic mean Q). Let D, E and F be as in Definition 2 and let $\rho \in (0, 1)$ be arbitrary. Let $\epsilon > 0$ and define the collection of numbers $R_\epsilon := (r_{j,k,l})_{1 \le j \le d, 1 \le k \le e, 1 \le l \le f}$, where

$$r_{j,k,l} := \begin{cases} (1+\epsilon)(1 - Q(D_j \times E_k)\nu(F_l)) & \text{if } \kappa_{j,k,l} \in (0,1) \\ 1 & \text{if } \kappa_{j,k,l} = 0 \\ \epsilon & \text{if } \kappa_{j,k,l} = 1 \end{cases}$$

where we use the shorthand $\kappa_{j,k,l} := Q(D_j \times E_k)\nu(F_l)$. Although the $r_{j,k,l}$'s are ϵ-dependent, we suppress this from the notation. The claim of the lemma is that for all sufficiently small $\epsilon > 0$, the stabilization entropy is well defined and satisfies

$$h(D, E, F, \rho, R_\epsilon) \le C. \tag{12}$$

Proof: We note that for $\epsilon > 0$ sufficiently enough, the conditions

(i) $1 - r := \sum_{j,k,l}(1 - r_{j,k,l}) \in [0, 1]$,
(ii) $1 - (1+\epsilon)(1 - Q(D_j, E_k)\nu(F_l)) \in (0, 1)$ for all j, k, l with $Q(D_j, E_k)\nu(F_l) \in (0, 1)$,

are both satisfied, thus ensuring that for such a small ϵ the stabilization entropy $h(D, E, F, \rho, R_\epsilon)$ is well defined. Consider system (9) evolving according to the fixed coding and control policy which renders the state process $(x_t)_{t \in \mathbb{N}}$ AMS ergodic with AMS mean Q. To prove that inequality we consider three cases:

Case 1: We first consider the case where $Q(D_j \times E_k)\nu(F_l) \in (0, 1)$ for all j, k, l. Let $\epsilon > 0$ be small enough such that $\epsilon < \rho$ as well as conditions (i) and (ii) are satisfied. We will show that for any such ϵ the claim holds. Let us denote the process measure by μ, which is AMS by assumption. Now, for any $V \in \mathcal{B}(\mathcal{W})$, it is clear by the i.i.d. property that

$$P\left(\left\{\omega \in \Omega : \lim_{T \to \infty} \frac{1}{T} \sum_{t=0}^{T-1} \mathbb{1}_V(w_t(\omega)) = \nu(V)\right\}\right) = 1.$$

Noting that x_t and w_t are independent at each time step t, $(w_t)_{t \in N}$ is i.i.d., and recalling equation (2), it follows that

$P(\hat{\Omega}) = 1$ where

$$\hat{\Omega} := \left\{\omega \in \Omega : \lim_{T \to \infty} \frac{1}{T} \sum_{t=0}^{T-1} \mathbb{1}_{D_j}(\pi_m(x_t(\omega))) \cdot \right. \tag{13}$$

$$\left. \mathbb{1}_{E_k}(\pi_{N-m}(x_t(\omega))) \cdot \mathbb{1}_{F_l}(w_t(\omega)) = Q(D_j \times E_k)\nu(F_l), \ \forall j, k, l\right\}$$

where we note that the above set can be written as the intersection of a finite number of full measure sets. We continue by defining the events

$$E_I^J := \left\{\omega \in \Omega : \left|\frac{1}{T} \sum_{t=0}^{T-1} \mathbb{1}_{D_j}(\pi_m(x_t(\omega))) \mathbb{1}_{E_k}(\pi_{N-m}(x_t(\omega)))\right. \right.$$

$$\left. \left. \mathbb{1}_{F_l}(w_t(\omega)) - Q(D_j \times E_k)\nu(F_l)\right| < \frac{1}{I} \forall j, k, l \text{ whenever } T \ge J\right\}$$

and note that for any $I \in \mathbb{N}$, it is clear that $\hat{\Omega} \subseteq \bigcup_{J=1}^{\infty} E_I^J$ therefore $P\left(\bigcup_{J=1}^{\infty} E_I^J\right) = 1$. Let now I_0 be large enough such that

$$\frac{1}{I_0} \le \epsilon(1 - Q(D_j \times E_k)\nu(F_l)) \quad \text{for all } j, k, l$$

and observe that $E_{I_0}^1 \subseteq E_{I_0}^2 \subseteq E_{I_0}^3 \subseteq \cdots$. By continuity of probability, we have

$$\lim_{J \to \infty} P(E_{I_0}^J) = P\left(\bigcup_{J=1}^{\infty} E_{I_0}^J\right) = 1,$$

and thus there exists $J_0 \in \mathbb{N}$ such that $P(E_{I_0}^J) \ge 1 - \epsilon$ for all $J \ge J_0$. For an arbitrary $T \ge J_0$, we define the set of control sequences

$$S_T := \{u_{[0;T-1]}(\omega) : \omega \in E_{I_0}^T\}.$$

We claim that this set is $(T, D, E, F, \rho, R_\epsilon)$-spanning. We use the set $\tilde{\Omega}_T := E_{I_0}^T \in \mathcal{F}$ to show this, where we note that $P(\tilde{\Omega}_T) \ge 1 - \epsilon > 1 - \rho$, satisfying the first requirement of the spanning set definition (Definition 2). To check the second condition, observe that for every $\omega \in \tilde{\Omega}_T$ and every triple j, k, l, the control sequence $u_{[0;T-1]}(\omega) \in S_T$ results in the joint state-noise process satisfying

$$\left|\frac{1}{T} \sum_{t=0}^{T-1} \mathbb{1}_{D_j}(\pi_m(x_t(\omega))) \mathbb{1}_{E_k}((\pi_{N-m}(x_t(\omega))) \mathbb{1}_{F_l}(w_t(\omega)) - \right.$$

$$\left. Q(D_j \times E_k)\nu(F_l)\right| < \frac{1}{I_0} \le \epsilon(1 - Q(D_j \times E_k)\nu(F_l)).$$

which implies that

$$\frac{1}{T}|\{t \in [0;T-1] : (\varphi(t, x_0(\omega), u_{[0;T-1]}(\omega), w(\omega)), w_t(\omega)) \in$$

$$D_j \times E_k \times F_l\}| \ge 1 - (1+\epsilon)(1 - Q(D_j, E_k)\nu(F_l)) = 1 - r_{j,k,l}$$

which establishes the second condition, since the triple j, k, l was arbitrary. We have thus established that S_T is $(T, D, E, F, \rho, R_\epsilon)$-spanning. Since the fixed causal coding and control policy can generate at most $|\mathcal{M}|^T$ distinct control sequences by time T, it follows that $|S_T| \le |\mathcal{M}|^T$, therefore $s(T, D, E, F, \rho, R_\epsilon) \le |\mathcal{M}|^T$. Recalling that $T \ge J_0$ was arbitrary, we find that

$$\log s(T, B, D, \rho, R_\epsilon) \le T \log_2 |\mathcal{M}| = TC \quad \text{for all } T \ge J_0,$$

and therefore dividing by T and letting $T \to \infty$ yields the desired capacity bound (12), completing the proof for Case 1.

Case 2: We now consider the case where every triple of sets (D_j, E_k, F_l) satisfies $Q(D_j \times E_k)\nu(F_l) \in [0,1)$. Suppose that j,k,l, is such that $Q(D_j \times E_k)\nu(F_l) = 0$. Then $1 - r_{j,k,l} = 0$ and the second condition in Definition 2 is vacuously satisfied. Combining this with Case 1, the result follows.

Case 3: Finally, we consider the case where for some indices j,k,l, $Q(D_j \times E_k)\nu(F_l) = 1$. Because each collection of sets is disjoint, $Q(D_{j'} \times E_{k'})\nu(F_{l'}) = 0$ whenever $(j,k,l) \neq (j',k',l')$. The analysis reduces to establishing the second condition in Definition 2 for the single set $D_j \times E_k \times F_l$ with $Q(D_j \times E_k)\nu(F_l) = 1$. Using an almost identical argument as in Case 1, the result follows. Alternatively, the analysis of a single set can be found in [33], where AMS was considered instead of AMS ergodicity as the control objective. Since AMS ergodicity implies AMS, and $h(D,E,F,\rho,R)$ reduces to the stabilization entropy notion used in [33] in case of a single set, the desired inequality follows. ∎

B. Proof of Theorem 2

To prove Theorem 2, we will approximate the integral in equation (10) form below using simple functions. We will prove that each of these approximations is upper bounded by the stabilization entropy, which in turn is no larger than the channel capacity. Taking a limit will yield the result. First, a few simplifications are in order.

Proof:

Recalling that we fixed an integer $m \le N$, we define $p := \{1,..,m\}$. WLOG, it suffices to establish

$$\int \int \log_2 |\det Df_w^p(x_{p_1},...,x_{p_m},x_{z_1},...,x_{z_{N-m}})|$$
$$dQ(x_1,...,x_N)d\nu(w) \le C \quad (14)$$

since by a relabeling of coordinates, any other set $p' \in \Gamma$ can be written in the form $\{1,2,3,..,|p'|\}$.

Recall now that by assumption (vi) in Theorem 2, there exists a set $S \subseteq \mathbb{R}^{N-m}$ with positive probability in the sense that

$$P(\{\omega \in \Omega : (x_0^{m+1}(\omega), x_0^{m+2}(\omega),...,x_0^N(\omega)) \in S\}) > 0.$$

Moreover, the set S has the property that for any $\hat{x} \in S$, conditioning on the event $\{\omega \in \Omega : (x_0^{m+1},..,x_0^{N-m})(\omega) = \hat{x}\}$ results in the law of the random vector $(x_0^1,..,x_0^m)$ admitting a bounded density. Let π_0' denote this conditional law. We now establish inequality (14) under slightly different assumptions than those of Theorem 2. More specifically, we impose that

- The last $N - m$ components of the initial state are deterministic, taking the value $\hat{x}$ for some arbitrary $\hat{x} \in S$.
- The initial m components of x_0 are distributed according to the law π_0'.

We claim that if we can establish (14) under these modified assumptions, the inequality will also hold under the assumptions of Theorem 2. To see this, suppose otherwise. Then a coding and control policy exists which stabilizes the system

in Theorem 2 over a channel of capacity strictly less than the LHS of (14). Since the stabilizing scheme works almost surely, this coding and control policy would also stabilize the system under the modified assumptions above for at least one $\hat{x} \in S$ (since S has non-zero measure), resulting in a contradiction since we are assuming that (14) holds for the modified system and for every $\hat{x} \in S$. We now proceed under the modified assumptions, and redefine $\pi_0 := \pi_0'$ to refer to the conditional law of $\pi_m(x_0)$.

Let $c \in (0,1)$ be such that $c < |\det Df_w^p(x,y)|$ for all $(x,y) \in \mathbb{R}^m \times \mathbb{R}^{N-m}$ and $w \in \mathcal{W}$. Let also $\delta > 0$ (think of this as small) and $\rho \in (0,1)$ (think of this as close to 1) be arbitrary. Next, fix Borel sets $D \subset \mathbb{R}^m$ and $E \subset \mathbb{R}^{N-m}$ satisfying that $D \times E$ have finite N-dimensional Lebesgue measure and that

$$Q(D \times E) > 1 - \frac{\delta}{2|\log c|}$$

holds (such sets can easily be found due to continuity of probability). Put also $F = \mathcal{W}$ and let $(D_j)_{j=1}^d$, $(E_k)_{k=1}^e$ and $(F_l)_{l=1}^f$ be (disjoint) partitions of D, E and F respectively. Let now $\epsilon > 0$ be small enough so that Lemma 1 holds, resulting in

$$h(D,E,F,\rho,R_\epsilon) \le C,$$

where R_ϵ is the associated collection of $r_{j,k,l}$'s as defined in Lemma 1. Let also $1 - r := \sum(1 - r_{j,k,l})$. Expanding out, it is easy to see (recalling that $\nu(F) = 1$) that $r = 1 - (1 + \epsilon)Q(D \times E) + def\epsilon$ (or $r = \epsilon$ if one of the $D_j \times E_k \times F_l$'s has full $Q \times \nu$-measure) thus we see that for every sufficiently small ϵ,

$$2r < \frac{\delta}{|\log c|}. \quad (15)$$

Now fix a sufficiently large $T \in \mathbb{N}$ and let S_T be a finite $(T,D,E,F,\rho,R_\epsilon)$-spanning set (whose existence is guaranteed by the proof of Lemma 1) with $\tilde{\Omega}_T \in \mathcal{F}$, $P(\tilde{\Omega}) \ge 1 - \rho$, the associated subset of Ω. Letting x_0^m denote the vector consisting of the first m components of x_0, we proceed by defining

$$A := \{(w(\omega), x_0^m(\omega)) : \omega \in \tilde{\Omega}\},$$
$$A(u) := \{(w,x) \in \mathcal{W}^{\mathbb{N}} \times \mathbb{R}^m : \forall j,k,l,$$
$$\frac{1}{T}\sum_{t=0}^{T-1} \mathbb{1}_{D_j \times E_k \times F_l}(\varphi(t,(x,\hat{x}),u,w),w_t) \ge 1 - r_{j,k,l}\}.$$
$$A(u,w) := \{x \in \mathbb{R}^m : (w,x) \in A(u)\}.$$

Letting m denote the m-dimensional Lebesgue measure, we see that

$$A \subseteq \bigcup_{u \in S_T} A(u), \quad (\nu \times m)(A(u)) = \int m(A(u,w))d\nu(w),$$

$$(16)$$

where the equality follows from the Fubini-Tonelli theorem (and the containment by definition of the sets). Letting $M > 0$ be an upper bound for the density of π_0, we have that

$$1 - \rho \le (\nu \times \pi_0)(A) \le M \cdot (\nu \times m)(A). \quad (17)$$

Combining (16) and (17), we obtain the key inequality:

$$\frac{1}{M}(1-\rho) \le (\nu \times m)(A) \le |S_T| \max_{u \in S_T} (\nu \times m)(A(u))$$

$$= |S_T| \max_{u \in S_T} \int m(A(u,w)) d\nu(w). \quad (18)$$

The next step in the proof is to obtain upper bounds for the volume $m(A(u,w))$. We proceed by defining a set consisting of disjoint collections of subsets of $\{0,\ldots,T-1\}$:

$$\mathbb{A} := \{\Lambda = \{\Lambda_{j,k,l}\}_{j,k,l} : \bigsqcup_{j=1}^{d}\bigsqcup_{k=1}^{e}\bigsqcup_{l=1}^{f} \Lambda_{j,k,l} \subseteq \{0,\ldots,T-1\},$$

$$|\Lambda_{j,k,l}| \ge (1-r_{j,k,l})T, \forall j=1,..,d, \ k=1,..,e, \ l=1,..,f\}$$

and note that as a consequence of the definition, $|\bigsqcup_{j=1}^{d}\bigsqcup_{k=1}^{e}\bigsqcup_{l=1}^{f}\Lambda_{j,k,l}| \ge (1-r)T$ for all $\Lambda \in \mathbb{A}$. We note that such sets can only be found for T sufficiently large, however as we will be taking a limit as $T \to \infty$, this is not a problem. For $\Lambda \in \mathbb{A}$, define the set

$$A(u,w,\Lambda) := \{x \in \mathbb{R}^m : (\varphi(t,(x,\hat{x}),u,w),w_t) \in D_j \times E_k \times F_l \Leftrightarrow$$
$$t \in \Lambda_{j,k,l} \text{ for all } j,k,l\}.$$

It is not hard to see that $A(u,w) = \bigsqcup_{\Lambda \in \mathbb{A}} A(u,w,\Lambda)$ is a disjoint union, thus (18) becomes

$$\frac{1}{M}(1-\rho) \le |S_T| \max_{u \in S_T} \int \sum_{\Lambda \in \mathbb{A}} m(A(u,w,\Lambda)) d\nu(w). \quad (19)$$

Our next step is to bound the volumes of the form $m(A(u,w,\Lambda))$. Writing $\varphi_{t,u,w}(\cdot) := \varphi(t,(\cdot,\hat{x}),u,w)$ we define

$$A_t(u,w,\Lambda) := \varphi_{t,u,w}(A(u,w,\Lambda)), \quad t=0,1,\ldots,T-1,$$

and observe that

$$A_t(u,w,\Lambda) \subseteq D_j \quad \text{whenever } t \in \Lambda_{j,k,l} \ \forall \ j,k,l.$$

Next, we define the following numbers:

$$c_{j,k,l} := \inf_{(x,y,w) \in D_j \times E_k \times F_l} |\det Df_w^p(x,y)|.$$

Recalling that by assumption $f_w^p(\cdot,y)$ is injective and C^1, it follows that for all (j,k,l) we have that

$$m(A_{t+1}(u,w,\Lambda)) \ge c_{j,k,l} \cdot m(A_t(u,w,\Lambda)) \text{ whenever } t \in \Lambda_{j,k,l},$$
$$m(A_{t+1}(u,w,\Lambda)) \ge c \cdot m(A_t(u,w,\Lambda)) \text{ whenever } t \notin \bigsqcup \Lambda_{j,k,l}.$$

Letting $t^*(\Lambda_{j,k,l}) := \max \Lambda_{j,k,l}$, $t^*(\Lambda) := \max_{j,k,l} t^*(\Lambda_{j,k,l})$, applying the above inequalities repeatedly, and recalling that $c \le c_{j,k,l}$, it is not hard to see that

$$m(A(u,w,\Lambda))\Big(\prod_{j=1}^{d}\prod_{k=1}^{e}\prod_{l=1}^{f} c_{j,k,l}^{|\Lambda_{j,k,l}|-1}\Big)c^{rT+def}$$
$$\le m(A_{t^*(\Lambda)}(u,w,\Lambda)).$$

where in principle, all the exponents of the $c_{j,k,l}$'s should be $|\Lambda_{j,k,l}|$, except for possibly one which should be $|\Lambda_{j,k,l}|-1$.

We do not know which one though, so we write the weaker inequality as above. Combining this with (19), we obtain

$$\frac{1}{M}(1-\rho) \le |S| \max_{u \in S_T} \sum_{\Lambda \in \mathbb{A}} \int m(A_{t^*(\Lambda)}(u,w,\Lambda))c^{-(rT+def)}$$
$$\prod_{j=1}^{d}\prod_{k=1}^{e}\prod_{l=1}^{f} c_{j,k,l}^{-(|\Lambda_{j,k,l}|-1)} d\nu(w),$$

and note that the right hand side of the above can be written as

$$= |S| \cdot c^{-(rT+def)} \max_{u \in S_T} \sum_{t_{1,1,1}=(1-r_{1,1,1})T}^{T} \cdots \sum_{t_{d,e,f}=(1-r_{d,e,f})T}^{T}$$
$$\int \sum_{\Lambda \in \mathbb{A}: \ t^*(\Lambda_{j,k,l})=t_{j,k,l}\forall j,k,l} m(A_{t^*(\Lambda)}(u,w,\Lambda))$$
$$\prod_{j=1}^{d}\prod_{k=1}^{e}\prod_{l=1}^{f} c_{j,k,l}^{-(|\Lambda_{j,k,l}|-1)} d\nu(w)$$

$$\le |S| \cdot c^{-(2rT+def)} \max_{u \in S_T} \sum_{t_{1,1,1}=(1-r_{1,1,1})T}^{T} \cdots \sum_{t_{d,e,f}=(1-r_{d,e,f})T}^{T}$$
$$\int \sum_{\Lambda \in \mathbb{A}: \ t^*(\Lambda_{j,k,l})=t_{j,k,l}\forall j,k,l} m(A_{t^*(\Lambda)}(u,w,\Lambda))$$
$$\prod_{j=1}^{d}\prod_{k=1}^{e}\prod_{l=1}^{f} c_{j,k,l}^{-((1-r_{j,k,l})T-1)} d\nu(w).$$

where the last inequality follows by noting that

$$c^{rT+def} \prod_{j,k,l} c_{j,k,l}^{|\Lambda_{j,k,l}|-1} = c^{rT+\sum_{j,k,l}|\Lambda_{j,k,l}|} \prod_{j,k,l}\Big(\frac{c_{j,k,l}}{c}\Big)^{|\Lambda_{j,k,l}|-1}$$
$$\ge c^{rT+\sum_{j,k,l}|\Lambda_{j,k,l}|} \prod_{j,k,l}\Big(\frac{c_{j,k,l}}{c}\Big)^{(1-r_{j,k,l})T-1}$$
$$= c^{rT+\sum_{j,k,l}|\Lambda_{j,k,l}|-(1-r)T+def} \prod_{j,k,l} c_{j,k,l}^{(1-r_{j,k,l})T-1}$$
$$\ge c^{2rT+def} \prod_{j,k,l} c_{j,k,l}^{(1-r_{j,k,l})T-1}.$$

Observe that the sets $A_{t^*(\Lambda)}(u,w,\Lambda)$ with $\Lambda \in \mathbb{A}$, $t^*(\Lambda)$ fixed, are pairwise disjoint, since they are the images of the corresponding sets $A(u,w,\Lambda)$ under the injective map $\varphi_{t^*(\Lambda),u,w}$. Moreover, all of these sets are contained in D, hence

$$\sum_{\Lambda \in \mathbb{A}: t^*(\Lambda_{j,k,l})=t_{j,k,l}\forall j,k,l} m(A_{t^*(\Lambda)}(u,w,\Lambda)) \le m(D).$$

Together with the above chain of inequalities, this implies

$$\frac{1}{M}(1-\rho) \le |S_T| \cdot m(D) \cdot c^{-(2rT+def)}.$$
$$\prod_{j=1}^{d}\prod_{k=1}^{e}\prod_{l=1}^{f} c_{j,k,l}^{-((1-r_{j,k,l})T-1)} \prod_{j=1}^{d}\prod_{k=1}^{e}\prod_{l=1}^{f}(r_{j,k,l}T+1).$$

Since this inequality holds for every T sufficiently large, we can take logarithms on both sides, divide by T and let $T \to \infty$.

This results in

$$0 \le h(D,E,F,\rho,R_\epsilon) - 2r\log c - \sum_{j=1}^{d}\sum_{k=1}^{e}\sum_{l=1}^{f}(1-r_{j,k,l})\log c_{j,k,l}.$$

Recalling the definition of $r_{j,k,l}$, the fact that ϵ can be chosen arbitrarily small and (15), this leads to the estimate

$$C+\delta \ge \sum_{j=1}^{d}\sum_{k=1}^{e}\sum_{l=1}^{f} Q(D_j \times E_k)\nu(F_l)$$
$$\inf_{(x,y,w)\in D_j\times E_k\times F_l}\log|\det Df_w^p(x,y)|.$$

Considering the supremum of the right-hand side over all finite measurable partitions of D, E and $F = \mathcal{W}$ leads to

$$C+\delta \ge \int\int \mathbb{1}_{D\times E}(x_1,..,x_N)\log|\det Df_w^p(x_1,..,x_N)|$$
$$dQ(x_1,..,x_N)d\nu(w),$$

where we use that the integrand is uniformly bounded below by $\log c$ (and hence, we can assume that it is non-negative). Considering now an increasing sequence of sets $D_k \times E_k \subset \mathbb{R}^N$ whose union is $\mathbb{R}^N$, we can invoke the theorem of monotone convergence to obtain the desired estimate, observing that δ can be made arbitrarily small as $D_k \times E_k$ becomes arbitrarily large. This completes the proof. ∎

V. CONCLUSIONS

In conclusion we have - for a certain class of non-linear systems - established a sharper bound on channel capacity required for ergodic stabilization. The techniques involved in the proof are stabilization entropy, a volume growth argument, and the property that almost surely, system sample paths visit regions of the state space at a frequency given by an ergodic measure. There are three possible avenues of further investigation. First, it would be interesting to enlarge the class of noise processes for which the bounds in this paper hold. We have considered only i.i.d. noise, however it is possible that an ergodic-like property (i.e. that equation (13) holds) for the joint state-noise process will hold for less restrictive classes of noise. Secondly, it seems possible to attempt the generalization of this result for the noisy channel case. Using stabilization entropy techniques, [3] established the bound (3) for scalar systems controlled over Discrete Memoryless Channels. Given that sharper bounds can be established for multi-dimensional systems, it seems worthwhile to attempt to generalize the one dimensional noisy-channel result to many dimensions, and combine it with the arguments in this paper to sharpen the bound. At a first glance, there appear to be no significant technical challenges to overcome. Finally, we note that using the current method, it is not possible to establish the bound for the most general class of systems of the form $x_{t+1} = f(x_t, w_t, u_t)$. The reason for this is that the proof in this paper relies heavily on the fact that with additive control, volume growth of the map $x \mapsto f(x,w) + u$ does not depend on the choice of u. A future direction of investigation is to consider the most

general class of systems, and impose restrictions on the type of causal coding and control policies in such a way so as to ensure that the control process has ergodic properties. We conclude by noting that we were unable to obtain the sharper bound established in this paper using information theoretic methods, with the main impediment being the fact that when splitting the state and conditioning on a past state realization, the unstable and stable state components may not be independent random variables, which is required for the information theoretic methods to apply.

REFERENCES

[1] P. Walters, *An introduction to ergodic theory*. Springer Science & Business Media, 2000, vol. 79.

[2] R. M. Gray, *Probability, random processes, and ergodic properties*. Springer, 2009.

[3] N. Garcia, C. Kawan, and S. Yüksel, "Ergodicity conditions for controlled stochastic nonlinear systems under information constraints: A volume growth approach," *SIAM Journal on Control and Optimization*, vol. 59, no. 1, pp. 534–560, 2021.

[4] S. Yüksel, "Stationary and ergodic properties of stochastic nonlinear systems controlled over communication channels," *SIAM Journal on Control and Optimization*, vol. 54, no. 5, pp. 2844–2871, 2016.

[5] J. Hespanha, A. Ortega, and L. Vasudevan, "Towards the control of linear systems with minimum bit-rate," in *Proc. of the Int. Symp. on the Mathematical Theory of Networks and Syst*. Citeseer, 2002, p. 1.

[6] D. F. Delchamps, "Stabilizing a linear system with quantized state feedback," *IEEE transactions on automatic control*, vol. 35, no. 8, pp. 916–924, 1990.

[7] N. Elia and S. K. Mitter, "Stabilization of linear systems with limited information," *IEEE Transactions on Automatic Control*, vol. 46, no. 9, pp. 1384–1400, 2001.

[8] G. N. Nair and R. J. Evans, "Stabilizability of stochastic linear systems with finite feedback data rates," *SIAM Journal on Control and Optimization*, vol. 43, no. 2, pp. 413–436, 2004.

[9] S. Tatikonda and S. Mitter, "Control under communication constraints," *IEEE Transactions on Automatic Control*, vol. 49, no. 7, pp. 1056–1068, 2004.

[10] W. S. Wong and R. W. Brockett, "Systems with finite communication bandwidth constraints. ii. stabilization with limited information feedback," *IEEE Transactions on Automatic Control*, vol. 44, no. 5, pp. 1049–1053, 1999.

[11] R. Bansal and T. Başar, "Simultaneous design of measurement and control strategies in stochastic systems with feedback," *Automatica*, vol. 45, pp. 679–694, September 1989.

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

[13] 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 Transactions on Information Theory*, vol. 52, no. 8, pp. 3369–3395, 2006.

[14] P. Minero, L. Coviello, and M. Franceschetti, "Stabilization over markov feedback channels: the general case," *IEEE Transactions on Automatic Control*, vol. 58, no. 2, pp. 349–362, 2012.

[15] V. Kostina, Y. Peres, G. Ranade, and M. Sellke, "Exact minimum number of bits to stabilize a linear system," in *2018 IEEE Conference on Decision and Control (CDC)*. IEEE, 2018, pp. 453–458.

[16] N. C. Martins and M. A. Dahleh, "Feedback control in the presence of noisy channels: 'Bode-like' fundamental limitations of performance," *IEEE Transactions on Automatic Control*, vol. 53, pp. 1604–1615, August 2008.

[17] G. N. Nair, F. Fagnani, S. Zampieri, and J. R. Evans, "Feedback control under data constraints: an overview," *Proceedings of the IEEE*, pp. 108–137, 2007.

[18] S. Yüksel, "Stochastic stabilization of noisy linear systems with fixed-rate limited feedback," *IEEE Transactions on Automatic Control*, vol. 55, pp. 2847–2853, December 2010.

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

[20] A. P. Johnston and S. Yüksel, "Stochastic stabilization of partially observed and multi-sensor systems driven by unbounded noise under fixed-rate information constraints," *IEEE Transactions on Automatic Control*, vol. 59, no. 3, pp. 792–798, 2013.

[21] S. Yüksel, "Characterization of information channels for asymptotic mean stationarity and stochastic stability of non-stationary/unstable linear systems," *IEEE Transactions on Information Theory*, vol. 58, pp. 6332–6354, October 2012.

[22] S. Yüksel and T. Başar, *Stochastic Networked Control Systems: Stabilization and Optimization under Information Constraints*. New York: Springer, 2013.

[23] D. Liberzon and J. P. Hespanha, "Stabilization of nonlinear systems with limited information feedback," *IEEE Transactions on Automatic Control*, vol. 50, no. 6, pp. 910–915, 2005.

[24] C. De Persis, "n-bit stabilization of n-dimensional nonlinear systems in feedforward form," *IEEE Transactions on Automatic Control*, vol. 50, no. 3, pp. 299–311, 2005.

[25] G. N. Nair, R. J. Evans, I. M. Mareels, and W. Moran, "Topological feedback entropy and nonlinear stabilization," *IEEE Transactions on Automatic Control*, vol. 49, no. 9, pp. 1585–1597, 2004.

[26] R. L. Adler, A. G. Konheim, and M. H. McAndrew, "Topological entropy," *Transactions of the American Mathematical Society*, vol. 114, no. 2, pp. 309–319, 1965.

[27] F. Colonius and C. Kawan, "Invariance entropy for control systems," *SIAM Journal on Control and Optimization*, vol. 48, no. 3, pp. 1701–1721, 2009.

[28] C. Kawan, "Invariance entropy for deterministic control systems," *Lecture notes in mathematics*, vol. 2089, 2013.

[29] F. Colonius, C. Kawan, and G. Nair, "A note on topological feedback entropy and invariance entropy," *Systems & Control Letters*, vol. 62, no. 5, pp. 377–381, 2013.

[30] F. Colonius, "Metric invariance entropy and conditionally invariant measures," *Ergodic Theory and Dynamical Systems*, vol. 38, no. 3, pp. 921–939, 2018.

[31] G. N. Nair, F. Fagnani, S. Zampieri, and R. J. Evans, "Feedback control under data rate constraints: An overview," *Proceedings of the IEEE*, vol. 95, no. 1, pp. 108–137, 2007.

[32] A. S. Matveev and A. V. Savkin, *Estimation and control over communication networks*. Springer Science & Business Media, 2009.

[33] C. Kawan and S. Yüksel, "Invariance properties of nonlinear stochastic dynamical systems under information constraints," *IEEE Transactions on Automatic Control, to appear (arXiv: 1901.02825)*, 2020.