Error Correcting Codes for Distributed Control

Ravi Teja Sukhavasi and Babak Hassibi

Abstract

The problem of stabilizing an unstable plant over a noisy communication link is an increasingly important one that arises in applications of networked control systems. Although the work of Schulman and Sahai over the past two decades, and their development of the notions of "tree codes" and "anytime capacity", provides the theoretical framework for studying such problems, there has been scant practical progress in this area because explicit constructions of tree codes with efficient encoding and decoding did not exist. To stabilize an unstable plant driven by bounded noise over a noisy channel one needs real-time encoding and real-time decoding and a reliability which increases exponentially with decoding delay, which is what tree codes guarantee. We prove that linear tree codes occur with high probability and, for erasure channels, give an explicit construction with an expected decoding complexity that is constant per time instant. We give novel sufficient conditions on the rate and reliability required of the tree codes to stabilize vector plants and argue that they are asymptotically tight. This work takes an important step towards controlling plants over noisy channels, and we demonstrate the efficacy of the method through several examples.

I. INTRODUCTION

Control theory deals with regulating the behavior of dynamical systems using real-time output feedback. Most traditional control systems are characterized by the measurement and control subsystems being co-located. Hence, there were no loss of measurement and control signals in the feedback loop. There is a very mature theory for this setup and there are concrete theoretical tools to analyze the overall system performance and its robustness to modeling errors [1]. There are

Ravi Teja Sukhavasi is a graduate student with the department of Electrical Engineering, California Institute of Technology, Pasadena, USA `teja@caltech.edu`

Babak Hassibi is a faculty with the department of Electrical Engineering, California Institute of Technology, Pasadena, USA `hassibi@caltech.edu`

increasingly many applications of networked control systems, however, where the measurement and control signals are communicated over noisy channels. Some examples include the smart grid, distributed computation, intelligent highways, etc (e.g., see [2]).

Applications of networked control systems represent different levels of decentralization in their structure. At a high level, the measurement unit and the controller are not co-located but each is individually centralized. In addition, the measurement and control subsystems are themselves comprised of arrays of sensors and actuators that in turn communicate with each other over a network. Our focus is on the former. We consider the setup where the measurement and control subsystems are individually centralized but are separated by communicated channels.

Several aspects of this problem have been studied in the literature [3]–[7]. When the communication links are modeled as rate-limited noiseless channels, significant progress has been made (see e.g., [8]–[10]) in understanding the bandwidth requirements for stabilizing open loop unstable systems. [11] considered robust feedback stabilization over communication channels that are modeled as variable rate digital links where the encoder has causal knowledge of the number of bits transmitted error free. Under a packet erasure model, [12] studied the problem of LQG (Linear Quadratic Gaussian) control in the presence of measurement erasures and showed that closed loop mean squared stability is not possible if the erasure probability is higher than a certain threshold. So, clearly the measurement and control signals need to be encoded to compensate for the channel errors.

There are two key differences between the communication paradigm for distributed control and that traditionally studied in information theory. Shannon's information theory, in large part, is concerned with reliable one-way communication while communication for control is fundamentally interactive: the plant measurements to be encoded are determined by the control inputs, which in turn are determined by how the controller decodes the corrupted plant measurements. Furthermore, conventional channel codes achieve reliability at the expense of delay which, if present in the feedback loop of a control system, can adversely affect its performance.

In this context, [13] provides a necessary and sufficient condition on the communication reliability needed over channels that are in the feedback loop of unstable scalar linear processes, and proposes the notion of anytime capacity as the appropriate figure of merit for such channels. In essence, the encoder is causal and the probability of error in decoding a source symbol that was transmitted d time instants ago should decay exponentially in the decoding delay d.

Although the connection between communication reliability and control is clear, very little is known about error-correcting codes that can achieve such reliabilities. Prior to the work of [13], and in the context of distributed computation, [14] proved the existence of codes which under maximum likelihood decoding achieve such reliabilities and referred to them as tree codes. Note that any real-time error correcting code is causal and since it encodes the entire trajectory of a process, it has a natural tree structure to it. [14] proves the existence of nonlinear tree codes and gives no explicit constructions and/or efficient decoding algorithms. [15] and [14] also propose sequential decoding algorithms whose expected complexity per time instant is fixed but the probability that the decoder complexity exceeds C decays with a heavy tail as $C^{-\gamma}$. Much more recently [16] proposed efficient error correcting codes for unstable systems where the state grows only polynomially large with time. When the state of an unstable scalar linear process is available at the encoder and when there is noiseless feedback of channel outputs, [17] and [18] develop encoding-decoding schemes that can stabilize such a process over the binary symmetric channel and the binary erasure channel respectively. But when the state is available only through noisy measurements or when there is no channel feedback, little is known in the way of stabilizing an unstable scalar linear process over a stochastic communication channel.

The subject of error correcting codes for control is in its relative infancy, much as the subject of block coding was after Shannon's seminal work in [19]. So, a first step towards realizing practical encoder-decoder pairs with anytime reliabilities is to explore linear encoding schemes. We consider rate $R = \frac{k}{n}$ causal linear codes which map a sequence of k-dimensional binary vectors $\{b_\tau\}_{\tau=0}^\infty$ to a sequence of $n-$dimensional binary vectors $\{c_\tau\}_{\tau=0}^\infty$ where c_t is only a function of $\{b_\tau\}_{\tau=0}^t$. Such a code is anytime reliable if at all times t and delays $d \geq d_o$, $P(\hat{b}_{t-d|t} \neq b_{t-d}) \leq \eta 2^{-\beta n d}$ for some $\beta > 0$. We show that linear tree codes exist and further, that they exist with a high probability. For the binary erasure channel, we propose a maximum likelihood decoder whose average complexity of decoding is constant per each time iteration and for which the probability that the complexity at a given time t exceeds KC^3 decays exponentially in C. This allows one to stabilize a partially observed unstable scalar linear process over a binary erasure channel and to the best of the authors' knowledge, this has not been done before.

In Section II, we present some background and motivate the need for anytime reliability with a simple example. In Section IV, we come up with a sufficient condition for anytime reliability in terms of the weight distribution of the code. In Section V, we introduce the ensemble of

time invariant codes and use the results from Section IV to prove that time invariant codes with anytime reliability exist with a high probability. In Section VI, we invoke some standard results from the literature on coding theory to improve the results obtained in Section V. In Section VII, we present a simple decoding algorithm for the erasure channel.

II. BACKGROUND

Owing to the duality between estimation and control, the essential complexity of stabilizing an unstable process over a noisy communication channel can be captured by studying the open loop estimation of the same process. We will motivate the kind of communication reliability needed for control through a simple example.

A toy example: Consider tracking the following random walk, $x_{t+1} = \lambda x_t + w_t$, where w_t is Bernoulli$\left(\frac{1}{2}\right)$, i.e., 0 or 1 with equal probability, $x_0 = 0$ and $|\lambda| > 1$. Suppose an observer observes x_t and communicates over a noisy communication channel to an estimator. Also assume that the estimator knows the system model and the initial state $x_0 = 0$. The observer clearly needs to communicate whether w_t is 0 or 1. Note that the observer only has causal access to $\{w_i\}$, i.e., at any time t, the observer has access to $\{w_0, \ldots, w_{t-1}\}$. Let the encoding function of the observer at time t be $f_t : \mathbb{GF}_2^t \mapsto \mathcal{X}^n$, where $\mathcal{X}$ is the channel input alphabet and n is the number of channel uses available for each step of the system evolution. One can visualize such a causal encoding process over a binary tree as in Fig. 1. While the information bits determine the path in the tree, the label on each branch denotes the symbol transmitted by the observe/encoder. The codeword associated to a given path in the tree is given by the concatenation of the branch symbols along that path. Upon receiving the channel outputs until time t, the estimator generates estimates $\{\hat{w}_{0|t}, \hat{w}_{1|t}, \ldots, \hat{w}_{t-1|t}\}$ of the noise sequence $\{w_0, w_1, \ldots, w_{t-1}\}$. Then, the estimator's estimate of the state, $\hat{x}_{t+1|t}$, is given by

$$\hat{x}_{t+1|t} = \sum_{j=0}^{t} \lambda_{t-j} \hat{w}_{j|t} \tag{1}$$

Suppose $P_{d,t}^e = P\left(\operatorname{argmin}_j(\hat{w}_{j|t} \neq \hat{w}_j) = t - d + 1\right)$, i.e., $P_{d,t}^e$ is the probability that the position of the earliest erroneous $\hat{w}_{j|t}$ is at time $j = t - d + 1$. The probability here is over the randomness
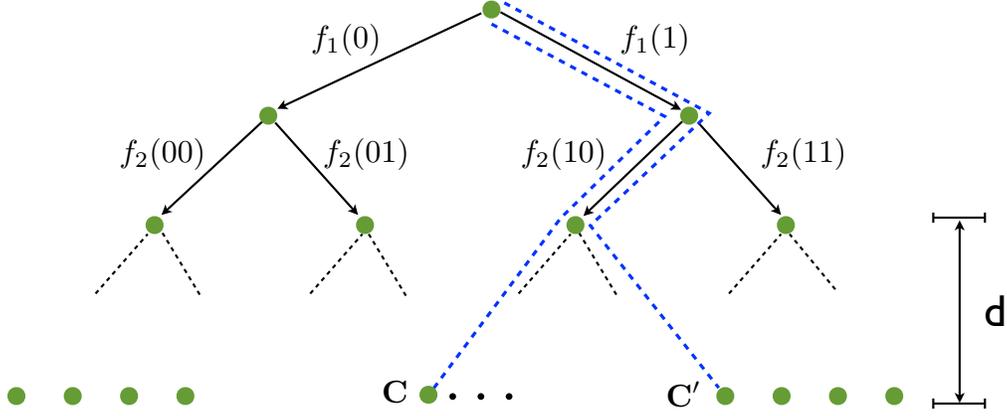
Fig. 1. One can visualize any causal code on a tree. The distance property is: $\|\mathbf{C} - \mathbf{C}'\|_{\mathcal{H}} \propto d$. This must be true for any two paths with a common root and of equal length in the tree

of the channel. From (1), we can bound $\mathbb{E}\left|x_{t+1} - \hat{x}_{t+1|t}\right|^2$ from above as

$$\sum_{w_{0:t}, \hat{w}_{0:t|t}} P\left(w_{0:t}, \hat{w}_{0:t|t}\right) \left|\sum_{j=1}^{n} \lambda^{t-j}(w_j - \hat{w}_{j|t})\right|^2$$

$$\leq \sum_{d \leq t} P_{d,t}^e \left|\sum_{j=t-d+1}^{t} \lambda^{t-j}(w_j - \hat{w}_{j|t})\right|^2$$

$$\leq \frac{1}{(|\lambda| - 1)^2} \sum_{d \leq t} P_{d,t}^e |\lambda|^{2d}$$

Clearly, a sufficient condition for $\limsup_t \mathbb{E}\left|x_{t+1} - \hat{x}_{t+1|t}\right|^2$ to be finite is as follows

$$P_{d,t}^e \leq |\lambda|^{-(2+\delta)d} \ \forall \ d \geq d_o, \ t > t_o \text{ and } \delta > 0 \tag{2}$$

where d_o and t_o are constants that do no depend on t, d.

In the context of control, it was first observed in [13] that exponential reliability of the form (2) is required to stabilize unstable plants over noisy communication channels. For a given channel, encoder-decoder pairs that achieve (2) are said to be anytime reliable. This definition will be made more precise in Section III. In the context of distributed computation, it was observed in [14] that a causal code under maximum likelihood decoding over a discrete memoryless channel is anytime reliable provided that the code has a certain distance property which is illustrated in Fig. 1. Avoiding mathematical clutter, one can describe the distance property as follows. For any two paths with a common root and of equal length in the tree whose least common ancestor
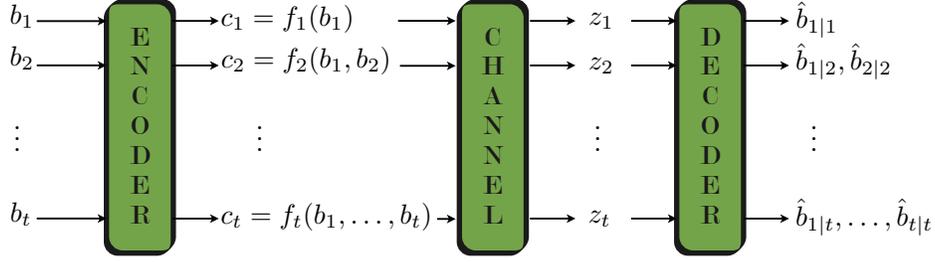
Fig. 2. Causal encoding and decoding

TABLE I

$H(.)$	The binary entropy function
$H^{-1}(y)$	The smaller root of the equation $H(x) = y$
For a matrix F, $\overline{F}$	abs(F), i.e., $\overline{F}_{i,j} = |F_{i,j}|.\forall\ i,j$
$\rho(F)$	Spectral radius of F
For a vector x, $x^{(i)}$	The ith component of x
$\mathbf{1}_m$	$[1, \ldots, 1]^T$, i.e., a column with m 1's
For $w, v \in \mathbb{R}^m$, $w \gtreqqless v$	Component-wise inequality
$\log(.)$	Logarithm in base 2
For $0 \le x, y \le 1$, $KL\left(x\|y\right)$	$x \log \dfrac{x}{y} + (1 - x) \log \dfrac{1 - x}{1 - y}$, i.e., Kullbeck-Leibler divergence between Bernoulli(x) and Bernoulli(y)

is at a height d from the bottom, the Hamming distance between their codewords should be proportional to d. [14] referred to codes with this distance property as *tree codes* and showed that they exist. There has recently been increased interest (e.g., [20]–[22]) in studying tree codes for interactive communication problems. But the tree codes are, in general, non-linear and the existence was not with high probability.

We will prove the existence, with high probability, of *linear tree codes* and exploit the linearity to develop an efficiently decodable anytime reliable code for the erasure channel.

III. PROBLEM SETUP

The notation to be used in the rest of the paper is summarized in Table I. Consider the following m_x−dimensional unstable linear system with m_y−dimensional measurements. Assume

that (F, H) is observable and (F, G) is controllable.

$$x_{t+1} = Fx_t + Gu_t + w_t, \quad y_t = Hx_t + v_t \tag{3}$$

where $\rho(F) > 1$, u_t is the m_u-dimensional control input and, w_t and v_t are bounded process and measurement noise variables, i.e., $\|w_t\|_\infty < \frac{W}{2}$ and $\|v_t\|_\infty < \frac{V}{2}$ for all t. The measurements $\{y_t\}$ are made by an observer while the control inputs $\{u_t\}$ are applied by a remote controller that is connected to the observer by a noisy communication channel. We assume that the control input is available to the plant losslessly. We do not assume that the observer has access to either the channel outputs or the control inputs. As is shown to be possible, e.g., in [9], [13], we do not use the control actions to communicate the channel outputs back to the observer through the plant because this could have a detrimental affect on the performance of the controller.

Before proceeding further, a word is in order about the boundedness assumption on the noise. If the process and/or measurement noise have unbounded support, it is not clear how one can stabilize the system without additional assumptions on the channel. For example, [17] assumes feedback of channel outputs to the observer in order to stabilize an unstable process perturbed by Gaussian noise over an erasure channel while [23] proposes a forward side channel between the observer and the controller that has a positive zero error capacity. We avoid this difficulty by assuming that the noise has bounded support which may be a reasonable assumption to make in practice.

The measurements $y_{0:t-1}$ will need to be quantized and encoded by the observer to provide protection from the noisy channel while the controller will need to decode the channel outputs to estimate the state x_t and apply a suitable control input u_t. This can be accomplished by employing a channel encoder at the observer and a decoder at the controller. For simplicity, we will assume that the channel input alphabet is binary. Suppose one time step of system evolution in (3) corresponds to n channel uses[1], i.e., n bits can be transmitted for each measurement of the system. Then, at each instant of time t, the operations performed by the observer, the channel encoder, the channel decoder and the controller can be described as follows. The observer generates a $k-$bit message, $b_t \in \mathbb{GF}^k$, that is a causal function of the measurements, i.e., it

---

[1]In practice, the system evolution in (3) is obtained by discretizing a continuous time differential equation. So, the interval of discretization could be adjusted to correspond to an integer number of channel uses, provided the channel use instances are close enough.

depends only on $y_{0:t}$. Then the channel encoder causally encodes $b_{0:t} \in \mathbb{GF}^{kt}$ to generate the n channel inputs $c_t \in \mathbb{GF}^n$. Note that the rate of the channel encoder is $R = k/n$. Denote the n channel outputs corresponding to c_t by $z_t \in \mathcal{Z}^n$, where $\mathcal{Z}$ denotes the channel output alphabet. Using the channel outputs received so far, i.e., $z_{0:t} \in \mathcal{Z}^{nt}$, the channel decoder generates estimates $\{\hat{b}_{\tau|t}\}_{\tau \leq t}$ of $\{b_\tau\}_{\tau \leq t}$, which, in turn, the controller uses to generate the control input u_{t+1}. This is illustrated in Fig. 2. Now, define

$$P_{t,d}^e = P\left(\min\{\tau : \hat{b}_{\tau|t} \neq b_\tau\} = t - d + 1\right)$$

Thus, $P_{t,d}^e$ is the probability that the earliest error is d steps in the past.

Definition 1 (Anytime reliability): Given a channel, we say that an encoder-decoder pair is $(R, \beta, d_o)-$anytime reliable over that channel if

$$P_{t,d}^e \leq \eta 2^{-n\beta d}, \quad \forall \ t, d \geq d_o \tag{4}$$

In some cases, we write that a code is $(R, \beta)-$anytime reliable. This means that there exists a fixed $d_o > 0$ such that the code is $(R, \beta, d_o)-$anytime reliable. □

We will show in Sections VIII and IX that $(R, \beta)-$anytime reliability with an appropriately large rate, R, and exponent, β, is a sufficient condition to stabilize (3) in the mean squared sense[2]. In what follows, we will demonstrate causal linear codes which under maximum likelihood (ML) decoding achieve such exponential reliabilities.

IV. LINEAR ANYTIME CODES

As discussed earlier, a first step towards developing practical encoding and decoding schemes for automatic control is to study the existence of linear codes with anytime reliability. We will begin by defining a causal linear code.

Definition 2 (Causal Linear Code): A causal linear code is a sequence of linear maps $f_\tau : \mathbb{GF}_2^{k\tau} \mapsto \mathbb{GF}_2^n$ and hence can be represented as

$$f_\tau(b_{1:\tau}) = G_{\tau 1}b_1 + G_{\tau 2}b_2 + \ldots + G_{\tau\tau}b_\tau \tag{5}$$

where $G_{ij} \in \mathbb{GF}_2^{n \times k}$ □

---

[2]can be easily extended to any other norm

TABLE II

$\mathbb{H}_{n,R}^t$	$\overline{n}t \times nt$ leading principal minor of $\mathbb{H}_{n,R}$
$\mathcal{C}_t$	$\{c \in \{0,1\}^{nt} : \mathbb{H}_{n,R}^t c = 0\}$
$\mathcal{C}_{t,d}$	$\{c \in \mathcal{C}_t : c_{\tau < t-d+1} = 0, \ c_{t-d+1} \neq 0\}$
$\|c\|$	Hamming weight of c
$N_{w,d}^t$	$|\{c \in \mathcal{C}_{t,d} : \|c\| = w\}|$
$w_{\min,d}^t$	$\mathrm{argmin}_w(N_{w,d}^t \neq 0)$
$P_{t,d}^e$	$P\left(\min\{\tau : \hat{b}_{\tau|t} \neq b_\tau\} = t - d + 1\right)$

We denote $c_\tau \triangleq f_\tau(b_{1:\tau})$. Note that a tree code is a more general construction where f_τ need not be linear. Also note that the associated code rate is $R = k/n$. The above encoding is equivalent to using a semi-infinite block lower triangular generator matrix $\mathbb{G}_{n,R}$ given by

$$\mathbb{G}_{n,R} = \begin{bmatrix} G_{11} & 0 & \ldots & \ldots & \ldots \\ G_{21} & G_{22} & 0 & \ldots & \ldots \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ G_{\tau 1} & G_{\tau 2} & \ldots & G_{\tau\tau} & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix}$$

One can equivalently represent the code with a parity check matrix $\mathbb{H}_{n,R}$, where $\mathbb{G}_{n,R}\mathbb{H}_{n,R} = 0$. The parity check matrix is in general not unique but it is easy to see that one can choose $\mathbb{H}_{n,R}$ to be block lower triangular too.

$$\mathbb{H}_{n,R} = \begin{bmatrix} H_{11} & 0 & \ldots & \ldots & \ldots \\ H_{21} & H_{22} & 0 & \ldots & \ldots \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ H_{\tau 1} & H_{\tau 2} & \ldots & H_{\tau\tau} & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix} \tag{6}$$

where $H_{ij} \in \{0,1\}^{\overline{n} \times n}$ and $\overline{n} = n(1-R)$. In fact, we present all our results in terms of the parity check matrix. Before proceeding further, some of the notation specific to coding is summarized in Table II.

The objective is to study the existence of causal linear codes which are $(R, \beta)-$anytime reliable under maximum likelihood (ML) decoding. With reference to Fig. 1, this amounts to choosing the branch labels, $f_\tau(b_{1:\tau})$, in such a way that they satisfy the distance property, and also are linear functions of the input, $b_{1:\tau}$. Further, we are interested in characterizing the thresholds on the rate, R, and exponent, β, for which such codes exist. In the interest of clarity, we will begin with a self-contained discussion of a weak sufficient condition on the distance distribution, $\{N^t_{w,d}, w^t_{\min,d}\}$, of a causal linear code so that it is anytime reliable under ML decoding. This sufficient condition is an adaptation of the distance property illustrated in Fig. 1 to the case of causal linear codes. In section V, we will demonstrate the existence of causal linear codes that satisfy this sufficient condition. The thresholds thus obtained will be significantly tightened in section VI by invoking some standard results from random coding literature, e.g., [24], [25].

A. A Sufficient Condition

Suppose the decoding instant is t and without loss of generality, assume that the all zero codeword is transmitted, i.e., $c_\tau = 0$ for $\tau \leq t$. We are interested in the error event where the earliest error in estimating b_τ happens at $\tau = t - d + 1$, i.e., $\hat{b}_{\tau|t} = 0$ for all $\tau < t - d + 1$ and $\hat{b}_{t-d+1|t} \neq 0$. Note that this is equivalent to the ML codeword, $\hat{c}$, satisfying $\hat{c}_{\tau < t-d+1} = 0$ and $\hat{c}_{t-d+1} \neq 0$, and $\mathbb{H}^t_{n,R}$ having full rank so that $\hat{c}$ can be uniquely mapped to a transmitted sequence $\hat{b}$. Then, using a union bound, we have

$$P^e_{t,d} = P\left[\bigcup_{c \in \mathcal{C}_{t,d}} (0 \text{ is decoded as } c)\right] \leq \sum_{c \in \mathcal{C}_{t,d}} P(0 \text{ is decoded as } c) \tag{7}$$

Consider a *memoryless binary-input output-symmetric* (MBIOS) channel. Let $\mathcal{X}$ and $\mathcal{Z}$ denote the input and output alphabet respectively. The Bhattacharya parameter, ζ, for such a channel is defined as

$$\zeta = \begin{cases} \int_{z \in \mathcal{Z}} \sqrt{p(z|X=1)p(z|X=0)}dz & \text{if } \mathcal{Z} \text{ is continuous} \\ \sum_{z \in \mathcal{Z}} \sqrt{p(z|X=1)p(z|X=0)} & \text{if } \mathcal{Z} \text{ is discrete valued} \end{cases}$$

Now, it is well known (e.g., see [26]) that, under ML decoding

$$P(0 \text{ is decoded as } c) \leq \zeta^{\|c\|}$$

From (7), it follows that $P_{t,d}^e \leq \sum_{w_{\min,d}^t \leq w \leq nd} N_{w,d}^t \zeta^w$. If $w_{\min,d}^t \geq \alpha nd$ and $N_{w,d}^t \leq 2^{\theta w}$ for some $\theta < \log_2(1/\zeta)$, then

$$P_{t,d}^e \leq \eta 2^{-\alpha nd(\log_2(1/\zeta)-\theta)} \tag{8}$$

where $\eta = (1 - 2^{\log_2(1/\zeta)-\theta})^{-1}$. So, an obvious sufficient condition for $\mathbb{H}_{n,R}$ can be described in terms of $w_{\min,d}^t$ and $N_{w,d}^t$ as follows. For some $\theta < \log_2(1/\zeta)$, we need

$$w_{\min,d}^t \geq \alpha nd \ \ \forall \ t, \ d \geq d_o \tag{9a}$$

$$N_{w,d}^t \leq 2^{\theta w} \ \ \forall \ t, \ d \geq d_o \tag{9b}$$

where d_o is a constant that is independent of d, t. This brings us to the following definition

Definition 3 (Anytime distance and Anytime reliability): We say that a code $\mathbb{H}_{n,R}$ has $(\alpha, \theta, d_o)-$anytime distance, if the following hold

1) $\mathbb{H}_{n,R}^t$ is full rank for all $t > 0$

2) $w_{\min,d}^t \geq \alpha nd$, $N_{w,d}^t \leq 2^{\theta w}$ for all $t > 0$ and $d \geq d_o$. $\qquad\square$

We require that $\mathbb{H}_{n,R}^t$ have full rank so that the mapping from the source bits $b_{1:t}$ to coded bits $c_{1:t}$ is invertible. We summarize the preceeding discussion as the following Lemma.

Lemma 4.1: If a code $\mathbb{H}_{n,R}$ has $(\alpha, \theta, d_o)-$anytime distance, then it is $(R, \beta, d_o)-$anytime reliable under ML decoding over a channel with Bhattacharya parameter ζ where

$$\beta = \alpha \left(\log(1/\zeta) - \theta\right) \qquad\square$$

V. LINEAR ANYTIME CODES - EXISTENCE

Consider causal linear codes with the following Toeplitz structure

$$\mathbb{H}_{n,R}^{TZ} = \begin{bmatrix} H_1 & 0 & \cdots & \cdots & \cdots \\ H_2 & H_1 & 0 & \cdots & \cdots \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ H_\tau & H_{\tau-1} & \cdots & H_1 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix}$$

The superscript TZ in $\mathbb{H}_{n,R}^{TZ}$ denotes 'Toeplitz'. $\mathbb{H}_{n,R}^{TZ}$ is obtained from $\mathbb{H}_{n,R}$ in (6) by setting $H_{ij} = H_{i-j+1}$ for $i \geq j$. Due to the Toeplitz structure, we have the following invariance,

$w^t_{\min,d} = w^{t'}_{\min,d}$ and $N^t_{w,d} = N^{t'}_{w,d}$ for all $d \leq \min(t, t')$. The code $\mathbb{H}^{TZ}_{n,R}$ will be referred to as a time-invariant code. The notion of time invariance is analogous to the convolutional structure used to show the existence of infinite tree codes in [14]. This time invariance allows one to prove that such codes which are anytime reliable are abundant.

Definition 4 (The ensemble $\mathbb{TZ}_p$*):* The ensemble $\mathbb{TZ}_p$ of time-invariant codes, $\mathbb{H}^{TZ}_{n,R}$, is obtained as follows, H_1 is any fixed full rank binary matrix and for $\tau \geq 2$, the entries of H_τ are chosen i.i.d according to Bernoulli(p), i.e., each entry is 1 with probability p and 0 otherwise. $\qquad\square$

For the ensemble $\mathbb{TZ}_p$, we have the following result

Theorem 5.1 (Abundance of time-invariant codes): Let $\bar{p} = \min\{p, 1 - p\}$. Then, for each $R > 0$ and

$$\alpha < H^{-1}(1 - R\log(1/(1 - \bar{p}))), \;\; \theta > -\log\left[(1 - \bar{p})^{-(1-R)} - 1\right], \text{ we have}$$

$$P\left(\mathbb{H}^{TZ}_{n,R} \text{ has } (\alpha, \theta, d_o) - \text{anytime distance}\right) \geq 1 - 2^{-\Omega(nd_o)}$$

Proof: See Appendix A $\qquad\blacksquare$

We can now use this result to demonstrate an achievable region of rate-exponent pairs for a given channel, i.e., the set of rates R and exponents β such that one can guarantee (R, β) anytime reliability using linear codes. Note that the thresholds in Theorem 5.1 are optimal when $p = 1/2$. So, for the rest of the analysis we fix $p = 1/2$. To determine the values of R that will satisfy (8), note that we need

$$\log(1/(2^{1-R} - 1)) < \log(1/\zeta) \implies R < 1 - \log(1 + \zeta)$$

With this observation, we have the following Corollary.

Corollary 5.2: For any rate R and exponent β such that

$$R < 1 - \log(1 + \zeta), \quad \text{and}$$

$$\beta < H^{-1}(1 - R)\left(\log\left(\frac{1}{\zeta}\right) + \log\left(2^{1-R} - 1\right)\right)$$

if $\mathbb{H}^{TZ}_{n,R}$ is chosen from $\mathbb{TZ}_{\frac{1}{2}}$, then

$$P\left(\mathbb{H}^{TZ}_{n,R} \text{ is } (R, \beta, d_o) - \text{anytime reliable}\right) \geq 1 - 2^{-\Omega(nd_o)}$$

□

Note that for BEC(ϵ), $\zeta = \epsilon$ and for BSC(ϵ), $\zeta = 2\sqrt{\epsilon(1-\epsilon)}$. The constant in the exponent $\Omega(nd_o)$ in Corollary 5.2 can be computed explicitly and it decreases to zero if either the rate or the exponent approach their respective thresholds. Further note that almost every code in the ensemble is (R, β)-anytime reliable after a large enough initial delay d_o.

The thresholds in Corollary 5.2 have been obtained by using a simple union bound for bounding the error probability in (7). As one would expect, these thresholds can be improved by doing a more careful analysis. It turns out that the ensemble of random causal linear codes bears close resemblance to random linear block codes. This allows one to borrow results from the random coding literature to tighten the thresholds.

VI. IMPROVING THE THRESHOLDS

We will examine the Toeplitz ensemble more closely and show that its delay dependent distance distribution is bounded above by that of the random binary linear code ensemble, which we will define shortly. This will enable us to significantly improve the rate, exponent thresholds of Section V that were obtained using a simple union bound.

A. A Brief Recap of Random Coding

For an arbitrary discrete memoryless channel, recall the following familiar definition of the random coding exponent, $E_r(R)$, from [24][3]

$$E_r(R) = \max_{0 \leq \rho \leq 1} \max_{\mathbf{Q}} \left[E_o\left(\rho, \mathbf{Q}\right) - \rho R \right], \quad \text{where} \tag{11a}$$

$$E_o\left(\rho, \mathbf{Q}\right) = -\log_2 \sum_{z \in \mathcal{Z}} \left[\sum_{x \in \mathcal{X}} Q(x) p(z|X = x)^{\frac{1}{1+\rho}} \right]^{1+\rho} \tag{11b}$$

In (11b), $Q(.)$ denotes a distribution on the channel input alphabet. The ensemble of random binary linear codes with block length N and rate $R = \frac{K}{N}$ is obtained by choosing an $(N-K) \times N$ binary parity check matrix H, i.e., $H \in GF_2^{(N-K) \times N}$, each of whose entries is chosen i.i.d Bernoulli$\left(\frac{1}{2}\right)$. For such an ensemble, any non-zero binary word $c \in GF_2^N$ is a codeword with probability $2^{-N(1-R)}$. For a given block code, let $w_{\min}$ denote the minimum distance and N_w

---

[3]We use base-2 instead of the natural logarithm

the number of codewords with Hamming weight w. A quick calculation shows that $\mathbb{E}N_w = \binom{N}{w}2^{-N(1-R)}$ and that $w_{\min}$ grows like $H^{-1}(1-R)N$ with a high probability. A *typical* code in this ensemble is defined to be one that has $w_{\min} \approx H^{-1}(1-R)N$ and $N_w \approx \binom{N}{w}2^{-N(1-R)}$. A simple Markov inequality shows that the probability that a code from this ensemble is *atypical* is at most $2^{-\Omega(N)}$. For the typical code over $\mathrm{BSC}(\epsilon)$, the block error probability decays as $2^{-NE_{BSC}(R)}$ where the exponent E_{BSC} has been characterized in [25]. As has been noted in [25], these calculations can be easily extended to a wider class of channels. In particular, the class of MBIOS channels admits a particularly clean characterization. We present the following generalization of the result in [25] without proof.

Lemma 6.1: Consider a linear code with block length N, rate R and distance distribution $\{N_w\}_{w=1}^N$ such that

1) $N_w = 0$ if $w \le H^{-1}(1-R-\delta)$
2) $N_w \le 2^{-N(1-R-\delta+o(1))}\binom{m}{w}$

for some $\delta > 0$. Let the channel be a MBIOS channel with Bhattacharya parameter ζ. Then the block error probability, P_e, under ML decoding is bounded as

$$P_e \le 2^{-N\left(E_\zeta(R)-\delta'\right)} \tag{12}$$

where

$$E_\zeta(R) = \begin{cases} H^{-1}(1-R)\log\frac{1}{\zeta} & , \ 0 \le R \le 1-H\left(\frac{\zeta}{1+\zeta}\right) \\ E_r(R) & , \ 1-H\left(\frac{\zeta}{1+\zeta}\right) \le R \le C \end{cases} \tag{13}$$

and $\delta' \to 0$ as $\delta \to 0$.

Proof: The proof is a straightforward generalization of the result in [25]. ∎

B. The Toeplitz Ensemble

In the causal case, fix an arbitrary decoding instant t and consider the event that the earliest error happens at a delay d. As seen before, the associated error probability depends on the relevant codebook $\mathcal{C}_{t,d}$ and its distance distribution $\{N_{w,d}^t\}_{w=1}^{nd}$. Recall from Table II that

$$\mathcal{C}_{t,d} \triangleq \{c \in \mathcal{C}_t : c_{\tau < t-d+1} = 0, \ c_{t-d+1} \ne 0\}$$

Due to the Toeplitz structure, we have $\mathcal{C}_{t,d} = \mathcal{C}_{d,d}$. So, we drop the subscript t in $N_{w,d}^t$ and write it as $N_{w,d}$. Note that $\mathcal{C}_{d,d}$ is determined by the matrix $\mathbb{H}_{n,R}^d$. Let c be a given nd-dimensional binary word, i.e., $c \in GF_2^{nd}$, and write $c = \left[c_1^T, c_2^T, \ldots, c_d^T\right]^T$, where $c_\tau \in GF_2^n$ notionally corresponds to the n encoder output bits during the τth time slot. Suppose $c_1 \neq 0$, then it is easy to see that

$$P\left(\mathbb{H}_{n,R}^d c = 0\right) = 2^{-\bar{n}d}$$

Recall that $\bar{n} = n(1 - R)$.

Now observe that $\mathbb{E}N_{w,d} \leq \binom{nd}{w} 2^{-\bar{n}d}$. This is same as the average weight distribution of the random binary linear code with a block length nd and rate R. So, applying Lemma 6.1, we get the following result.

Theorem 6.2: For each rate $R < C$ and exponent $\beta < E_\zeta(R)$, if $\mathbb{H}_{n,R}^{TZ}$ is chosen from $\mathbb{TZ}_{\frac{1}{2}}$, then

$$P\left(\mathbb{H}_{n,R}^{TZ} \text{ is } (R, \beta, d_o) - \text{anytime reliable}\right) \geq 1 - 2^{-\Omega(nd_o)}$$

where C is the Shannon capacity of the channel and

$$E_\zeta(R) = \begin{cases} H^{-1}(1 - R)\log\frac{1}{\zeta} & , \quad 0 \leq R \leq 1 - H\left(\frac{\zeta}{1+\zeta}\right) \\ E_r(R) & , \quad 1 - H\left(\frac{\zeta}{1+\zeta}\right) \leq R \leq C \end{cases} \tag{14}$$

$\square$

The problem of stabilizing unstable scalar linear systems over noisy channels in the absence of feedback has been considered in [13]. [13] showed the existence of $(R, \beta)-$anytime reliable codes for $R < C$ and $\beta < E_r(R)$. The code is not linear in general and the existence was not with high probability. Theorem 6.2 proves linear anytime reliable codes for exponent, β, up to $E_\zeta(R)$. When $R < 1 - H\left(\frac{\zeta}{1+\zeta}\right)$, $E_\zeta(R) > E_r(R)$. So, Theorem 6.2 marks a significant improvement in the known thresholds for stabilizing unstable processes over noisy channels, as is demonstrated in Figures 3 and 4.

VII. DECODING OVER THE BINARY ERASURE CHANNEL

Owing to the simplicity of the erasure channel, it is possible to come up with an efficient way to perform maximum likelihood decoding at each time step. Consider an arbitrary decoding instant t, let $c = [c_1^T, \ldots, c_t^T]^T$ be the transmitted codeword and let $z = [z_1^T, \ldots, z_t^T]^T$ denote the

(a) Binary Erasure Channel, $\epsilon = 0.15$

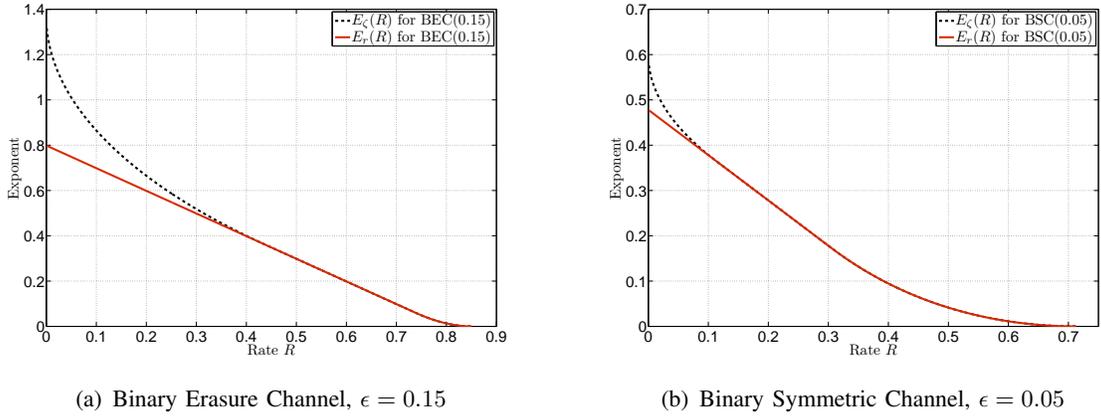(b) Binary Symmetric Channel, $\epsilon = 0.05$

Fig. 3. Comparing the thresholds obtained from Theorem 6.2 and Theorem 5.2 in [13]

corresponding channel outputs. Recall that $\mathbb{H}^t_{n,R}$ denotes the $\overline{n}t \times nt$ leading principal minor of $\mathbb{H}_{n,R}$. Let z_e denote the erasures in z and let H_e denote the columns of $\mathbb{H}^t_{n,R}$ that correspond to the positions of the erasures. Also, let $\tilde{z}_e$ denote the unerased entries of z and let $\tilde{H}_e$ denote the columns of $\mathbb{H}^t_{n,R}$ excluding H_e. So, we have the following parity check condition on z_e, $H_e z_e = \tilde{H}_e \tilde{z}_e$. Since $\tilde{z}_e$ is known at the decoder, $s \triangleq \tilde{H}_e \tilde{z}_e$ is known. Maximum likelihood decoding boils down to solving the linear equation $H_e z_e = s$. Due to the lower triangular nature of H_e, unlike in the case of traditional block coding, this equation will typically not have a unique solution, since H_e will typically not have full column rank. This is alright as we are not interested in decoding the entire z_e correctly, we only care about decoding the earlier entries accurately. If $z_e = [z_{e,1}^T, \ z_{e,2}^T]^T$, then $z_{e,1}$ corresponds to the earlier time instants while $z_{e,2}$ corresponds to the latter time instants. The desired reliability requires one to recover $z_{e,1}$ with an exponentially smaller error probability than $z_{e,2}$. Since H_e is lower triangular, we can write $H_e z_e = s$ as

$$
\begin{bmatrix} H_{e,11} & 0 \\ H_{e,21} & H_{e,22} \end{bmatrix} \begin{bmatrix} z_{e,1} \\ z_{e,2} \end{bmatrix} = \begin{bmatrix} s_1 \\ s_2 \end{bmatrix} \tag{15}
$$

Let $H_{e,22}^{\perp}$ denote the orthogonal complement of $H_{e,22}$, ie., $H_{e,22}^{\perp} H_{e,22} = 0$. Then multiplying both sides of (15) with $\mathrm{diag}(I, H_{e,22})$, we get

$$
\begin{bmatrix} H_{e,11} \\ H_{e,22}^{\perp} H_{e,21} \end{bmatrix} z_{e,1} = \begin{bmatrix} s_1 \\ H_{e,22}^{\perp} s_2 \end{bmatrix} \tag{16}
$$

If $[H_{e,11}^T \ (H_{e,22}^\perp H_{e,21})^T]^T$ has full column rank, then $z_{e,1}$ can be recovered exactly. The decoding algorithm now suggests itself, i.e., find the smallest possible $H_{e,22}$ such that $[H_{e,11}^T \ (H_{e,22}^\perp H_{e,21})^T]^T$ has full rank and it is outlined in Algorithm 1. Note that one can equivalently describe the

---

Algorithm 1 Decoder for the BEC

---

1) Suppose, at time t, the earliest uncorrected error is at a delay d. Identify z_e and H_e as defined above.
2) Starting with $d' = 1, 2, \ldots, d$, partition

$$z_e = [z_{e,1}^T \ z_{e,2}^T]^T \text{ and } H_e = \begin{bmatrix} H_{e,11} & 0 \\ H_{e,21} & H_{e,22} \end{bmatrix}$$

where $z_{e,2}$ correspond to the erased positions up to delay d'.
3) Check whether the matrix $\begin{bmatrix} H_{e,11} \\ H_{e,22}^\perp H_{e,21} \end{bmatrix}$ has full column rank.
4) If so, solve for $z_{e,1}$ in the system of equations

$$\begin{bmatrix} H_{e,11} \\ H_{e,22}^\perp H_{e,21} \end{bmatrix} z_{e,1} = \begin{bmatrix} s_1 \\ H_{e,22}^\perp s_2 \end{bmatrix}$$

5) Increment $t = t + 1$ and continue.

---

decoding algorithm in terms of the generator matrix and it will be very similar to Alg 1.

A. Encoding and Decoding Complexity

Consider the decoding instant t and suppose that the earliest uncorrected erasure is at time $t - d + 1$. Then steps 2) and 3) in Algorithm 1 can be accomplished by just reducing H_e into the appropriate row echelon form, which has complexity $O(d^3)$. The earliest entry in z_e is at time $t - d + 1$ implies that it was not corrected at time $t - 1$, the probability of which is $P_{d-1,t-1}^e \leq \eta 2^{-n\beta(d-1)}$. Hence, if nothing more had to be done, the average decoding complexity would have been at most $K \sum_{d>0} d^3 2^{-n\beta d}$ which is bounded and is independent of t. In particular, the probability of the decoding complexity being Kd^3 would have been at most $\eta 2^{-n\beta d}$. But, inorder to actually solve for $z_{e,1}$ in step 4), one needs to compute the syndromes s_1 and s_2. It is easy to see that the complexity of this operation increases linearly in time t. This is to be expected since the code has infinite memory. A similar computational complexity also plagues the encoder, for, the encoding operation at time t is described by $c_t = G_t b_1 + \ldots + G_1 b_t$ where $\{b_i\}$ denote the source bits and hence becomes progressively hard with t.

We propose the following scheme to circumvent this problem in practice. We allow the decoder to periodically, say at $t = \ell(2T)$ ($\ell = 1, 2 \ldots$) for appropriately chosen T, provide feedback to the encoder on the position of the earliest uncorrected erasure which is, say at time $t - d$. The encoder can use this information to stop encoding the source bits received prior to $t - d$, i.e., $\{b_i\}$ for $i \leq t - d - 1$ starting from time $t + T$. In other words, for $\tau > t + T$, $c_\tau = G_{\tau - t + d + 2} b_{t-d-1} + \ldots + G_1 b_\tau$. The decoder accordingly uses the new generator matrix starting from $t + T$. In practice, this translates to an arrangement where the decoder sends feedback at time t and can be sure that the encoder receives it by time $t + T$. Such feedback, in the form of acknowledgements from the receiver to the transmitter, is common to most packet-based modern communication and networked systems for reasonable values of T. Note that this form of feedback finds a middle ground between one extreme of having no feedback at all and another extreme where every channel output is fed back to the transmitter, the latter being impractical in most cases. The decoder proposed in Alg. 1 is easy to implement and its performance is simulated in Section XI.

B. Extension to Packet Erasures

The encoding and decoding algorithms presented so far have been developed for the case of bit erasures. But it is not difficult to see that the techniques generalize to the case of packet erasures. For example, for a packet length L, what was one bit earlier will now be a block of L bits. Each binary entry in the encoding/parity check matrix will now be an $L \times L$ binary matrix. The rate will remain the same. So, at each time, k packets each of length L will be encoded to n packets each of the same length L. Recall that the *anytime performance* of the code is determined by the delay dependent codebook $\mathcal{C}_{t,d}$ and its distance distribution $\{N_{w,d}^t\}_{w=1}^{nd}$. In the case of packet erasures, one can obtain analogous results by defining the Hamming distance of a codeword slightly differently. By viewing a codeword as a collection of packets, define its Hamming distance to be the number of non zero packets. The definition of the delay dependent distance distribution $\{N_{w,d}^t\}$ will change accordingly. With this modification, one can easily apply the results developed in Sections IV, V and the decoding algrithm in Section VII above to the case of packet erasures.

VIII. Sufficient Conditions for Stabilizability - Scalar Measurements

Recall that we do not assume any feedback about the channel outputs or the control inputs at the observer/encoder. This is the setup we imply whenever we say that no feedback is assumed. In this context [13] derives a sufficient condition for stabilizing scalar linear systems over noisy channels without feedback while [27] considers stabilizing vector valued processes in the presence of feedback. So, to the best of our knowledge, there are no results on stabilizing unstable vector valued processes over a noisy channel when the observer does not have access to either the control inputs or the channel outputs.

We will develop two sufficient conditions for stabilizing vector valued processes over noisy channels without feedback. The two sufficient conditions are based on two different estimation algorithms employed by the controller and neither is stronger than the other. We will then show in Section X-A that both sufficient conditions are asymptotically tight. For ease of presentation, we will treat the case of scalar and vector measurements separately. We will present the sufficient conditions for the case of scalar measurements here while vector measurements will be treated in Section IX

Consider the unstable m_x-dimensional linear state space model in (3) with scalar measurements, i.e., $\rho(F) > 1$, and $m_y = 1$. Suppose that the characteristic polynomial of F is given by

$$f(z) \triangleq z^{m_x} + a_1 z^{m_x-1} + \ldots + a_{m_x}$$

Without loss of generality we assume that (F, H) are in the following canonical form.

$$F = \begin{bmatrix} -a_1 & 1 & 0 & \ldots \\ -a_2 & 0 & 1 & 0 \\ \vdots & \vdots & & \ddots \\ -a_{m-1} & \ldots & \ldots & 0 & 1 \\ -a_m & 0 & \ldots & \ldots & 0 \end{bmatrix}, \quad H = [1, 0, \ldots, 0]$$

Owing to the duality between estimation and control, we can focus on the problem of tracking (3) over a noisy communication channel. For, if (3) can be tracked with an asymptotically finite mean squared error and if (F, G) is stabilizable, then it is a simple exercise to see that there exists a control law $\{u_t\}$ that will stabilize the plant in the mean squared sense, i.e.,

$\limsup_t \mathbb{E}\|x_t\|^2 < \infty$. In particular, if the control gain K is chosen such that $F + GK$ is stable, then $u_t = K\hat{x}_{t|t}$ will stabilize the plant, where $\hat{x}_{t|t}$ is the estimate of x_t using channel outputs up to time t. In control parlance, this amounts to verifying that the control input does not have a *dual effect* [28]. Hence, in the rest of the analysis, we will focus on tracking (3). The control input u_t therefore is assumed to be absent, i.e., $u_t = 0$.

A. Hypercuboidal Filter

We bound the set of all possible states that are consistent with the estimates of the quantized measurements using a hypercuboid, i.e., a region of the form $\{x \in \mathbb{R}^{m_x} | \mathbf{a} \le x \le \mathbf{b}\}$, where $\mathbf{a}, \mathbf{b} \in \mathbb{R}^{m_x}$ and the inequalities are component-wise.

Since we assume that the initial state x_0 has bounded support, we can write $x_{min,0|-1} \le x_0 \le x_{max,0|-1}$ and suppose using the channel ouputs received till time $t-1$, we have $x_{min,t|t-1} \le x_t \le x_{max,t|t-1}$. Since $H = [1, 0, \ldots, 0]$, the measurement update provides information of the form $x_{min,t|t}^{(1)} \le x_t^{(1)} \le x_{max,t|t}^{(1)}$ while there will be no additional information on other components of x_t. Note that an estimate of the state is given by the mid point of this region, i.e., $\hat{x}_{t|t} = 0.5(x_{min,t|t} + x_{max,t|t})$. If we define $\Delta_{t|t} = x_{max,t|t} - x_{min,t|t}$, then the estimation error is asymptotically bounded if every component of $\Delta_{t|t}$ is asymptotically bounded. Using such a filter, we can stabilize the system in the mean squared sense over a noisy channel provided that the rate R and exponent β of the $(R, \beta)-$anytime reliable code used to encode the measurements satisfy the following sufficient condition

Theorem 8.1: It is possible to stabilize (3) in the mean squared sense with an $(R, \beta)-$anytime code provided

$$R > R_n = \frac{1}{n} \log_2 \sum_{i=1}^{m_x} |a_i|, \quad \beta > \beta_n = \frac{2}{n} \log_2 \rho(\overline{F}) \tag{17}$$

Proof: See Appendix B $\blacksquare$

Before proceeding further, we will provide a brief sketch of the proof. Note that $\Delta_{t|t} = x_{max,t|t} - x_{min,t|t}$ is a measure of the uncertainty in the state estimate. From Lemma A.2, $\Delta_{t+1|t} = \overline{F}\Delta_{t|t} + W\mathbf{1}_{m_x}$. The anytime exponent is determined by the growth of Δ_t in the absence of measurements, hence the bound $\beta_n = 2\log_2 \rho(\overline{F})$. The bound on the rate is determined by how fine the quantization needs to be for Δ_t to be bounded asymptotically. It will be shown in Section

E that $\rho\left(\overline{F}\right)$ is always larger than $\rho\left(F\right)$. By using an alternate filtering algorithm, which we call the Ellipsoidal filter, one can improve this requirement on the exponent from $\beta_n > 2\log_2\rho(\overline{F})$ to $\beta_n > 2\log_2\rho(F)$. But this will come at the price of a larger rate.

B. Ellipsoidal Filter

One can alternately bound the set of all possible states that are consistent with the estimates of the quantized measurements using an ellipsoid

$$\mathcal{E}(P,c) \triangleq \left\{x \in \mathbb{R}^{m_x} | \langle x - c, P^{-1}(x-c)\rangle \leq 1\right\}$$

This can be seen as an extension of the technique proposed in [29] to filtering using quantized measurements. If $m_x = 1$, $\rho(\overline{F}) = \rho(F)$. So, let $m_x \geq 2$.

Let $x_0 \in \mathcal{E}(P_0, 0)$ and suppose using the channel outputs received till time $t - 1$, we have $x_t \in \mathcal{E}(P_{t|t-1}, \hat{x}_{t|t-1})$. Since $H = [1, 0, \ldots, 0]$, the measurement update provides information of the form $x^{(1)}_{min,t|t} \leq x^{(1)}_t \leq x^{(1)}_{max,t|t}$, which one may call a slab. $\mathcal{E}(P_{t|t}, \hat{x}_{t|t})$ would then be an ellipsoid that contains the intersection of the above slab with $\mathcal{E}(P_{t|t-1}, \hat{x}_{t|t-1})$, in particular one can set it to be the minimum volume ellipsoid covering this intersection. Lemma A.4 gives a formula for the minimum volume ellipsoid covering the intersection of an ellipsoid and a slab. For the time update, it is easy to see that for any $\epsilon' > 0$ and $P_{t+1} = (1+\epsilon')FP_{t|t}F^T + \frac{W^2}{4\epsilon'}\mathbf{1}_{m_x}\mathbf{1}_{m_x}^T$, $\mathcal{E}(P_{t+1}, F\hat{x}_{t|t})$ contains the state x_{t+1} whenever $\mathcal{E}(P_{t|t}, \hat{x}_{t|t})$ contains x_t. This leads to the following Lemma, the proof of which is contained in the discussion above. For convenience, we write P_t for $P_{t|t-1}$.

Lemma 8.2 (The Ellipsoidal Filter): Whenever $\mathcal{E}(P_0, 0)$ contains x_0, for each $\epsilon' > 0$, the following filtering equations give a sequence of ellipsoids $\left\{\mathcal{E}(P_{t|t}, \hat{x}_{t|t})\right\}$ that, at each time t, contain x_t.

$$P_{t+1} = (1+\epsilon')FP_{t|t}F^T + \frac{W^2}{4\epsilon'}\mathbf{1}_{m_x}, \ \hat{x}_{t+1} = F\hat{x}_{t|t} \tag{18a}$$

$$P_{t|t} = b_t P_t - (b_t - a_t)\frac{P_t e_1 e_1^T P_t}{e_1^T P_t e_1}, \ \hat{x}_{t|t} = \xi_t \frac{P_t e_1}{\sqrt{e_1^T P_t e_1}} \tag{18b}$$

where a_t, b_t and ξ_t can be calculated in closed form using Lemma A.4, and e_1 is the m_x-dimensional unit vector $e_1 = [1, 0, \ldots, 0]^T$.

Using this approach, we get the following sufficient condition.

Theorem 8.3: It is possible to stabilize (3) for $m_x \geq 2$ in the mean squared sense with an $(R, \beta)-$anytime code provided

$$R > R_{e,n} = \frac{1}{n} \log_2 \left[\sqrt{m_x} \sum_{i=1}^{m_x} |a_i| \theta^{i-1} \right] \tag{19a}$$

$$\beta > \beta_{e,n} = \frac{2}{n} \log_2 \rho(F) \tag{19b}$$

where $\theta = \sqrt{\frac{m_x}{m_x - 1}}$

Proof: See Appendix D ∎

IX. SUFFICIENT CONDITIONS FOR STABILIZABILITY - VECTOR MEASUREMENTS

Like in the scalar case, we will assume without loss of generality that (F, H) are in a canonical form (is obtained from a simple transformation of *Scheme I* in Sec 6.4.6 of [30]) with the following structure. F is a $q \times q$ block lower triangular matrix with $F^{i,j}$ denoting the $(i, j)^{th}$ block. So, $F^{i,j} = 0$ if $j > i$. $F^{i,j}$ is an $\ell_i \times \ell_j$ matrix and $\sum_{i=1}^{q} \ell_i = m_x$. The diagonal blocks $F^{i,i}$ have the following structure.

$$F^{i,i} = \begin{bmatrix} -a_{i,1} & 1 & 0 & \dots \\ -a_{i,2} & 0 & 1 & 0 \\ \vdots & \vdots & & \ddots \\ -a_{i,\ell_i-1} & \dots & \dots & 0 & 1 \\ -a_{i,\ell_i} & 0 & \dots & \dots & 0 \end{bmatrix}$$

while the off-diagonal blocks do not have any specific structure. The measurement matrix H is of the form $H = \begin{bmatrix} H_1^T, & H_2^T \end{bmatrix}^T$ where H_1 is a $q \times m_x$ matrix of the following form

$$H_1 = \text{block diag} \left\{ [1 \ 0 \ \cdots \ 0], \ 1 \times \ell_i, \ i = 1, \dots, q \right\} \tag{20}$$

H_2 does not have any particular structure and is not relevant. Note that the characteristic polynomial of F, is given by $f(z) = \prod_{i=1}^{q} \left(z^{\ell_i} + a_{i,1} z^{\ell_i-1} + \dots + a_{i,\ell_i} \right)$.

If the Hypercuboidal filter is used, then Theorem 8.1 can be extended to the case of vector measurements is as follows.

Theorem 9.1: It is possible to stabilize (3) in the mean squared sense with an $(R, \beta)-$anytime

code provided

$$R > R_{v,n} = \frac{1}{n} \sum_{i=1}^{q} \max \left\{ 0, \log \sum_{j=1}^{\ell_i} |a_{i,j}| \right\}, \quad \beta > \beta_{v,n} = \frac{2}{n} \log_2 \rho \left(\overline{F} \right) \qquad \text{(21a)}$$

Proof: See Appendix B2 ∎

The thresholds if one uses an Ellipsoidal filter are given as follows.

Theorem 9.2: It is possible to stabilize (3) in the mean squared sense with an $(R, \beta)-$anytime code provided

$$R > R_{ve,n} = \frac{1}{n} \sum_{i=1}^{q} \max \left\{ 0, \log \left[\sqrt{m_x} \sum_{j=1}^{\ell_i} |a_{i,j}| \theta^{j-1} \right] \right\}, \quad \beta > \beta_{ve,n} = \frac{2}{n} \log_2 \rho \left(F \right) \qquad \text{(22a)}$$

where $\theta = \sqrt{\frac{m_x}{m_x-1}}$ □

We skip the proof for Theorem 9.2 since it is very similar to that of Theorem 9.1.

X. DISCUSSION - ASYMPTOTICS AND THE STABILIZABLE REGION

The sufficient conditions derived above are non-asymptotic in the sense that measurements are encoded every time step. Alternately, one can encode the measurements every, say, ℓ time steps, and consider the asymptotic rate and exponent needed as ℓ grows. This is often the form in which such sufficient conditions appear in the literature [8], [10], [13]. Even though the sufficient conditions in Sections VIII and IX are non-asymptotic, note that they depend only on the system matrices F, H and not on the noise distribution. In order to compare our results with those in the literature, we examine the sufficient conditions in the asymptotic limit of large ℓ.

A. The Limiting Case

Note that encoding once every ℓ measurements amounts to working with the system matrix F^ℓ. So, one can calculate this limiting rate and exponent by writing the eigen values of F, $\{\lambda_i\}_{i=1}^{m}$, as $\lambda_i = \mu_i^n$ and letting n scale. The following asymptotic result allows us to compare the sufficient conditions above with those in the literature (eg., see [8], [10], [13]).

Theorem 10.1 (The Limiting Case): Write the eigen values of F, $\{\lambda_i\}_{i=1}^{m_x}$, in the form $\lambda_i = \mu_i^n$. Letting n scale, R_n, $R_{v,n}$, $R_{e,n}$, $R_{ev,n}$ converge to R^*, and β_n, $\beta_{v,n}$, $\beta_{e,n}$, $\beta_{ev,n}$ converge to β^*,

where

$$R^* = \sum_{i:|\mu_i|>1} \log_2 |\mu_i|, \ \ \beta^* = 2\log_2 \max_i |\mu_i| \tag{23}$$

Proof: See Appendix E. ∎

For stabilizing plants over deterministic rate limited channels, [8] showed that a rate $R > R^*$, where R^* is as in (23), is necessary and sufficient. So, asymptotically the sufficient condition for the rate R in Theorem 8.1 is tight. But it is not clear if one do with an exponent smaller than $\beta^* = 2\log_2 \max_i |\mu_i|$ asymptotically when there is no feedback. Though the above limiting case allows one to obtain a tight and an intuitively pleasing characterization of the rate and exponent needed, it should be noted that this may not be operationally practical. For, if one encodes the measurements every ℓ time steps, even though Theorem 10.1 guarantees stability, the performance of the closed loop system (the LQR cost, say) may be unacceptably large because of the delay we incur. This is what motivated us to present the sufficient condition in the form that we did above.

B. A Comment on the Trade-off Between Rate and Exponent

Once a set of rate-exponent pairs (R, β) that can stabilize a plant is available, one would want to identify the pair that optimizes a given cost function. Higher rates provide finer resolution of the measurements while larger exponents ensure that the controller's estimate of the plant does not drift away; however, we cannot have both. One can either coarsely quantize the measurements and protect the bits heavily or quantize them moderately finely and not protect the bits as much. One can easily cook up examples using an LQR cost function with the balance going either way. Studying this trade-off is integral to making the results practically applicable.

C. Stabilizable Region

Using the thresholds obtained in Theorem 6.2, and the asymptotic sufficient condition in Theorem 10.1, we can discuss the range of the eigen values of F, i.e., $\{|\mu_i|\}_{i=1}^{m_x}$, for which the ηth moment of x_t in (3) can be stabilized over some common channels. Since we are interested in the asymptotics, we assume the same limiting case as in Section X-A. Firstly, consider the scalar case, i.e., $m_x = 1$ and let the eigen value be μ. An anytime reliable code with rate R and
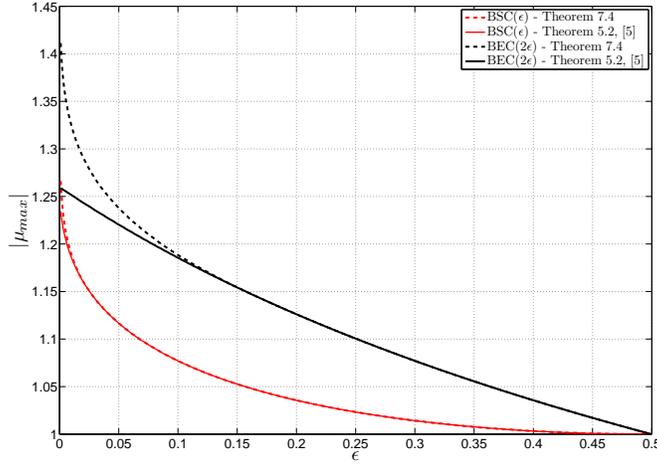
Fig. 4. Comparing the stabilizable regions of BSC and BEC using linear codes

exponent β can stabilize the process in (3) for all μ such that

$$\log_2 |\mu| < \min\left\{R, \frac{\beta}{\eta}\right\}$$

So, a scalar unstable linear process in (3) can be stabilized over a MBIOS channel with Bhattacharya parameter ζ provided

$$\log_2 |\mu| < \log_2 |\mu_{\max}| = \sup_{R<C, \beta<E_\zeta(R)} \min\left\{R, \frac{\beta}{\eta}\right\} \tag{24}$$

The stabilizable region as implied by the threshold in [13] is given by

$$\log_2 |\mu| < \log_2 |\mu_{\max}| = \sup_{R<C, \beta<E_r(R)} \min\left\{R, \frac{\beta}{\eta}\right\}$$

For $\eta = 2$, the stabilizable region for the BEC and BSC is shown in Fig 4 where $|\mu_{max}|$ is plotted against the channel parameter. Consider a vector valued process with unstable eigen values $\{|\mu_i|\}_{i=1}^m$. Such a process can be stabilized by a rate R and exponent β anytime reliable code provided $R > \sum_{i=1}^m \log |\mu_i|$ and $\beta > \log(\max_i |\mu_i|)$. So, given a channel with Bhattacharya parameter ζ for which the rate exponent curve $(R, E_\zeta(R))$ is achievable, the region of unstable eigen values that can be stabilized is given by $\{\mu \in \mathbb{R}^m, \mid \exists R < C \ni \sum_{i=1}^m \log |\mu_i| < R$ and $\log(\max_i |\mu_i|) < E_\zeta(R)\}$, where C is the Shannon capacity of the channel. For
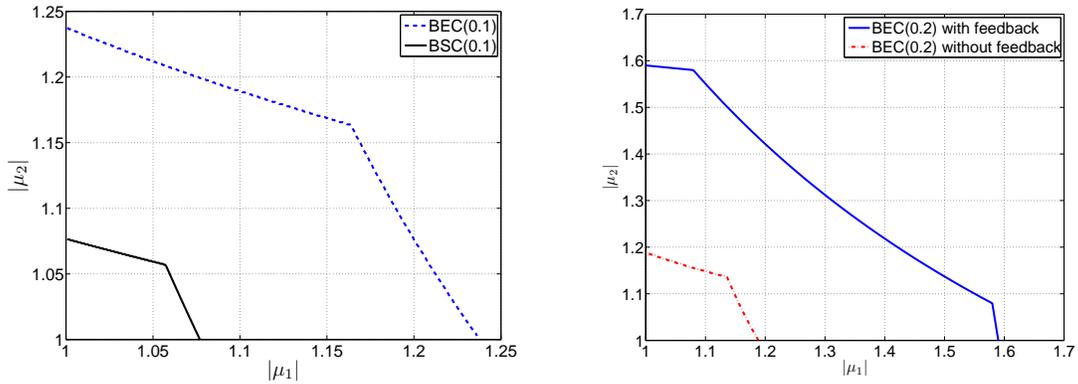
(a) Each curve represents the outer boundary of the stabilizable region.

(b) Stabilizable region with and without feedback

Fig. 5. Comparing the stabilizable region of different channels

example, let $m = 2$ and $\eta = 2$. Fig 5a shows the region of $(|\mu_1|, |\mu_2|)$ that can be stabilized over three different channels, a binary symmetric channel with bit flip probability 0.1 and binary erasure channels with erasure probabilities 0.1 and 0.2 respectively.

We will now compare these results with the case when there is perfect feedback of the channel outputs at the observer/encoder. [27] considered a priority queuing method for stabilizing vector valued unstable processes over channels with perfect feedback. Bits from different unstable subsystems are placed in a FIFO queue. Bits are given preference in decreasing order of the size of the eigen value of the corresponding subsystem. So, bits coming from a subsystem with a larger eigen value are given preference over those from a subsystem with a smaller eigen value. A bit is removed from the queue once it is received correctly. Since the feedback anytime capacity of a binary erasure channel is known [31], one can use Theorem 6.1 in [27] to derive the region of eigen values that can be stabilized by such a scheme. In Fig. 5b, we compare the region of $(|\mu_1|, |\mu_2|)$ that can be stabilized with and without feedback over a binary erasure channel with erasure probability 0.2. As one would expect, the region is much larger when there is feedback. Note that the stabilizable regions in Fig. 5 are only achievable and not necessarily tight.

XI. SIMULATIONS

We present two examples and stabilize them over a binary erasure channel with erasure probability $\epsilon = 0.3$. The number of channel uses per measurement is fixed to $n = 15$. In both cases, time invariant codes $\mathbb{H}_{15,R} \in \mathbb{TZ}_{\frac{1}{2}}$, for an appropriate rate R, were randomly generated and decoded using Algorithm 1. The controller uses the Hypercuboidal filter to estimate the state.

A. Cart-Stick Balancer

The system parameters for a cart-stick balancer (also commonly called the *inverted pendulum on a cart*) with state variables of stick angle, stick angular velocity, and cart velocity, when sampled with sampling duration 0.1s are (Exercise 10.15 in [32])

$$F = \begin{bmatrix} 1.161 & 0.105 & 0 \\ 3.3 & 1.161 & 0.002 \\ -3.265 & -0.160 & 0.979 \end{bmatrix}, \quad G = [-0.003 \; -0.068 \; 0.859]^T, \quad H = [10 \; 0 \; 0]$$

The characteristic polynomial of F is $x^3 - 3.3x^2 + 3.27x - 0.98$ and its eigen values are 1.75, 0.98 and 0.57. So, F is open loop unstable. Each component of the process noise and measurement noise is i.i.d zero mean Gaussian with variance 0.01 truncated to lie in [-0.025,0.025]. The control input is given by $u_t = -K\hat{x}_{t|t}$, where $K = [-81.55 \; -14.37 \; -0.04]$. One can verify that $F - GK$ is stable. In order to apply Theorem 8.1, we write F in the following canonical form

$$F_o = \begin{bmatrix} 3.3 & 1 & 0 \\ -3.27 & 0 & 1 \\ 0.98 & 0 & 0 \end{bmatrix}$$

Applying Theorem 8.1, one can stabilize x_t in the mean squared sense provided the exponent $n\beta > 2\log\left(\rho\left(\overline{F_o}\right)\right) = 4.1035$ and the rate $nR = k > \log\left(3.3 + 3.27 + 0.98\right) = 2.1$. For $k = 5$, there exist anytime reliable codes with exponent upto $n\beta = 4.27$. Fig 6 plots a sample path of the above system for a randomly chosen Toeplitz code. It is clear from Fig 6(b) that the plant is stabilized.
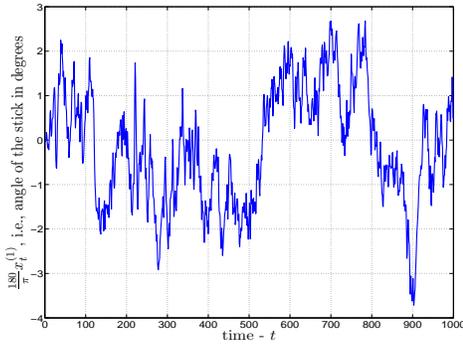
(a) The stick does not deviate by more than 3 degrees from the vertical
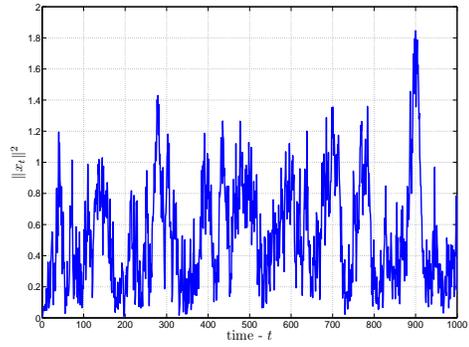
(b) This shows that the plant is stabilized

Fig. 6. A sample path

B. Example 2

This example is aimed at exploring the trade off between the resolution of the quantizer and the error performance of the causal code. Consider a 3-dimensional unstable system (3) with

$$F = \begin{bmatrix} 2 & 1 & 0 \\ 0.25 & 0 & 1 \\ -0.5 & 0 & 0 \end{bmatrix}$$

$G = \mathcal{I}_3$ and $H = [100]$. Each component of w_t and v_t is generated i.i.d $N(0,1)$ and truncated to [-2.5,2.5]. The eigen values of F are $\{2, -0.5, 0.5\}$ while $\lambda(\overline{F}) = 2.215$. The observer has access to the control inputs and we use the hypercuboidal filter outlined in Section B. Using Theorem 8.1, the minimum required bits and exponent are given by $k = nR \geq 2$ and $n\beta \geq 2\log_2 2.215 = 2.29$. The control input is $u_t = -\hat{x}_{t|t-1}$. For $k \leq 7$, $n\beta \geq 2.32$. If $k = 8$, $n\beta = 1.32 < 2.29$. For each value of k ranging from 3 to 7, 1000 codes were generated from the ensemble $\mathbb{TZ}_{\frac{1}{2}}$. For each code, the system was simulated over a horizon of 100 time instants and the LQR cost has been averaged over 100 such runs. For a time horizon T, the LQR cost is defined as $\frac{1}{2T} \sum_{t=0}^{T} \mathbb{E}\left(\|x_t\|^2 + \|u_t\|^2 \right)$. In Fig 7(a), the cumulative distribution function of the LQR cost is plotted for $3 \leq k \leq 7$. The $x-$axis denotes the proportion of codes for which the LQR cost is below a prescribed value, e.g., with $k = 6, n = 15$, the cost was less than 15 for 85% of the codes while with $k = 5, n = 15$, this fraction increases to more than 95%. The
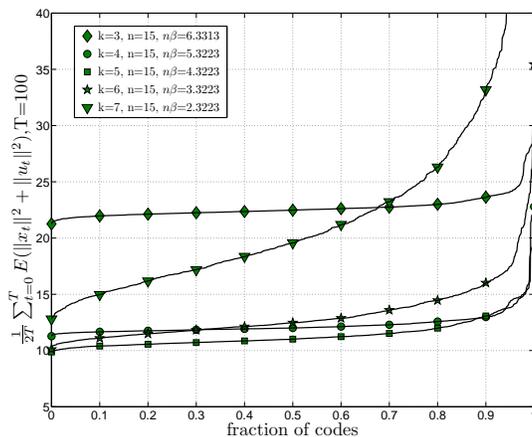
(a) The CDF of the LQR costs for different values of the rate
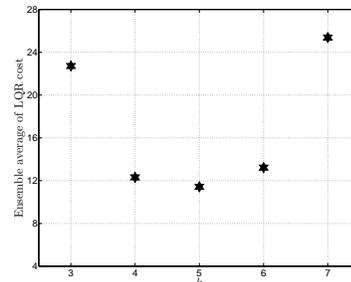
(b) The LQR cost averaged over the 1000 randomly generated codes is plotted against k

Fig. 7. The best choice of the rate is $R = 5/15 = 0.33$

competition between the rate and the exponent in determining the LQR cost is evident when we look at Fig 7(b). When $k = 3$, the error exponent $n\beta = 6.3$ is large. So, at any time t, the decoder decodes all the source bits $\{b_\tau\}_{\tau \leq t-1}$ with a high probability. Hence, the limiting factor on the LQR cost is the resolution that the source bits b_t provide on the measurements. But when $k = 7$, the measurements are quantized to a high resolution but the decoder makes errors in decoding the source bits. So, the best choice appears to be $k = 5$.

REFERENCES

[1] B. Hassibi, A.H. Sayed, and T. Kailath, *Indefinite-Quadratic Estimation and Control*, SIAM, 1999.

[2] R.M. Murray, K.J. Astrom, S.P. Boyd, R.W. Brockett, and G. Stein, "Future directions in control in an information-rich world," *Control Systems, IEEE*, vol. 23, no. 2, apr 2003.

[3] VS Borkar and SK Mitter, "Lqg control with communication constraints," *Communications, computation, control, and signal processing: a tribute to Thomas Kailath*, pp. 365–373.

[4] Wing Shing Wong and R.W. Brockett, "Systems with finite communication bandwidth constraints. i. state estimation problems," *Automatic Control, IEEE Transactions on*, vol. 42, no. 9, sep 1997.

[5] Wing Shing Wong and R.W. Brockett, "Systems with finite communication bandwidth constraints. ii. stabilization with limited information feedback," *Automatic Control, IEEE Transactions on*, vol. 44, no. 5, may 1999.

[6] J Nillson, *Real-time control systems with delays*, Ph.D. thesis, Lund Institute of Technology, 1998.

[7] G. C. Walsh and H. Ye, "Scheduling of networked control systems," *IEEE Control Systems Magazine*, vol. 21, Feb 2001.

[8] GN Nair and RJ Evans, "Stabilizability of stochastic linear systems with finite feedback data rates," *SIAM Journal on Control and Optimization*, vol. 43, no. 2, pp. 413–436, 2005.

[9] Alexey S. Matveev and Andrey V. Savkin, *Estimation and Control over Communication Networks (Control Engineering)*, Birkhauser, 2007.

[10] P. Minero, M. Franceschetti, S. Dey, and G.N. Nair, "Data rate theorem for stabilization over time-varying feedback channels," *Automatic Control, IEEE Transactions on*, vol. 54, no. 2, 2009.

[11] N.C. Martins, M.A. Dahleh, and N. Elia, "Feedback stabilization of uncertain systems in the presence of a direct link," *Automatic Control, IEEE Transactions on*, vol. 51, no. 3, march 2006.

[12] B. Sinopoli, L. Schenato, M. Franceschetti, K. Poolla, M.I. Jordan, and S.S. Sastry, "Kalman filtering with intermittent observations," *Automatic Control, IEEE Transactions on*, vol. 49, no. 9, sept. 2004.

[13] Anant Sahai and Sanjoy Mitter, "The necessity and sufficiency of anytime capacity for stabilization of a linear system over a noisy communication link - part i: Scalar systems," *Information Theory, IEEE Transactions on*, vol. 52, no. 8, pp. 3369–3395, 2006.

[14] LJ Schulman, "Coding for interactive communication," *Information Theory, IEEE Transactions on*, vol. 42, no. 6, pp. 1745 – 1756, 1996.

[15] H Palaiyanur and A Sahai, "A simple encoding and decoding strategy for stabilization over discrete memoryless channels," *Proceedings of the Allerton Conference on Control, Communication and Computing*, 2005.

[16] R Ostrovsky, Y Rabani, and LJ Schulman, "Error-correcting codes for automatic control," *Information Theory, IEEE Transactions on*, vol. 55, no. 7, pp. 2931 – 2941, 2009.

[17] Serdar Yüksel, "A random time stochastic drift result and application to stochastic stabilization over noisy channels," *Communication, Control, and Computing, 2009. Allerton 2009. 47th Annual Allerton Conference on*, pp. 628–635, 2009.

[18] T. Simsek, R. Jain, and P. Varaiya, "Scalar estimation and control with noisy binary observations," *Automatic Control, IEEE Transactions on*, vol. 49, no. 9, pp. 1598 – 1603, 2004.

[19] CE Shannon, "A mathematical theory of communication," *Bell System Technical Journal*, vol. 27, pp. 379 – 423 and 623 – 656, July and Oct 1948.

[20] Mark Braverman and Anup Rao, "Towards coding for maximum errors in interactive communication," in *Proceedings of the 43rd annual ACM symposium on Theory of computing*, 2011, STOC '11.

[21] Ran Gelles and Amit Sahai, "Potent tree codes and their applications: Coding for interactive communication, revisited," *CoRR*, vol. abs/1104.0739, 2011.

[22] Ankur Moitra, "Efficiently coding for interactive communication," *Electronic Colloquium on Computational Complexity (ECCC)*, 2011.

[23] S. Yüksel and T. Başar, "Control over noisy forward and reverse channels," *Automatic Control, IEEE Transactions on*, vol. 56, no. 5, may 2011.

[24] Robert G. Gallager, *Information Theory and Reliable Communication*, 1968.

[25] A. Barg and Jr. Forney, G.D., "Random codes: minimum distances and error exponents," *Information Theory, IEEE Transactions on*, vol. 48, no. 9, pp. 2568 – 2573, sep 2002.

[26] Igal Sason and Shlomo Shamai, "Performance analysis of linear codes under maximum-likelihood decoding: A tutorial," *FNT in Communications and Information Theory*, vol. 3, no. 1/2, pp. 1–222, 2006.

[27] Anant Sahai and Sanjoy K. Mitter, "The necessity and sufficiency of anytime capacity for stabilization of a linear system over a noisy communication link, part ii: vector systems," *http://arxiv.org/abs/cs/0610146*, vol. abs/cs/0610146, 2006.

[28] Y. Bar-Shalom and E. Tse, "Dual effect, certainty equivalence, and separation in stochastic control," *Automatic Control, IEEE Transactions on*, vol. 19, no. 5, oct 1974.

[29] F. Schweppe, "Recursive state estimation: Unknown but bounded errors and system inputs," *Automatic Control, IEEE Transactions on*, vol. 13, no. 1, Feb. 1968.

[30] T. Kailath, *Linear Systems*, Prentice-Hall, Inc., Englewood Cliffs, N.J., 1980.

[31] A. Sahai, "Why do block length and delay behave differently if feedback is present?," *Information Theory, IEEE Transactions on*, vol. 54, no. 5, may 2008.

[32] Gene Franklin, J. D. Powell, and Abbas Emami-Naeini, *Feedback Control of Dynamic Systems*, Pearson Prentice Hall, 5th, edition, 2006.

[33] A. Makhdoumi Kakhaki, H. Karkeh Abadi, Pedram Pad, H. Saeedi, Kasra Alishahi, and Farokh Marvasti, "Capacity achieving random sparse linear codes," *CoRR*, vol. abs/1102.4099, 2011.

[34] Osman Güler and Filiz Gürtuna, "The extremal volume ellipsoids of convex bodies, their symmetry properties, and their determination in some special cases," *arXiv*, vol. math.MG, Sep 2007.

[35] A Sluis, "Upperbounds for roots of polynomials," *Numerische Mathematik*, vol. 15, no. 3, pp. 250–262, 1970.

APPENDIX

A. Proof of Theorem 5.1

We will begin with some preliminary observations.

Lemma A.1 ([33]): Let V be an $m-$dimensional vector space over $\mathbb{GF}_2$ and define a probability function over V such that, for each $v \in V$, $P(v) = p^{\|v\|}(1 - p)^{m-\|v\|}$. If U is an $\ell-$dimensional subspace of V, then

$$P(U) \leq \max(p, 1 - p)^{m-\ell}$$

Proof: Suppose $p \leq 1/2$. The proof for the other case is analogous. Let E be the set of unit vectors, i.e., $E = \{v \in V \mid \|v\| = 1\}$. Then there is a subset, E', of E with $m - \ell$ unit vectors such that $V = U \oplus span(E')$ and $U \cap span(E') = \{0\}$. Let $u' \in span(E')$, then

$$P\left(U + u'\right) = \sum_{u \in U} P(u + u') \geq \sum_{u \in U} P(u) \left(\frac{p}{1 - p}\right)^{\|u'\|} = P(U) \left(\frac{p}{1 - p}\right)^{\|u'\|}$$

Note that for distinct $u'_1, u'_2 \in span(E')$, $(U + u'_1) \cap (U + u'_2) = \emptyset$. Also note that $\|u'\| \leq m - \ell$ $\forall\ u' \in span(E')$.

$$1 = P(V) = P\left(\bigcup_{u' \in span(E')} (U + u')\right) \geq \sum_{u' \in span(E')} P(U) \left(\frac{p}{1 - p}\right)^{\|u'\|}$$

Observe that there are exactly $\binom{m-\ell}{i}$ vectors in $span(E')$ with Hamming weight i. So, we have

$$1 \geq P(U) \sum_{i=0}^{m-\ell} \binom{m-\ell}{i} \left(\frac{p}{1-p}\right)^i = P(U) \left(\frac{1}{1-p}\right)^{m-\ell}$$

This completes the proof. ∎

Remark 1: The Toeplitz parity check matrix $\mathbb{H}_{n,R}^{TZ}$ is full rank if and only if H_1 is full rank. This is why we fix H_1 to be a full rank matrix in the definition of the Toeplitz ensemble.

Recall that we choose the entries of H_i to be i.i.d Bernoulli(p) for $i \geq 2$. Also suppose $p \leq 1/2$. The results for $p \geq 1/2$ are obtained by replacing p with $1-p$ in the subsequent analysis. Consider an arbitrary decoding instant, t. Since $w_{\min,d}^t = w_{\min,d}^{t'}$ and $N_{w,d}^t = N_{w,d}^{t'}$ for all t, t', we will drop these superscripts and write $w_{\min,d}^t = w_{\min,d}$ and $N_{w,d}^t = N_{w,d}$. Let $c = [c_1^T, \ldots, c_t^T]^T$, where $c_i \in \{0,1\}^n$, be a fixed binary word such that $c_{\tau < t-d+1} = 0$ and $c_{t-d+1} \neq 0$. Also, let $\mathbb{H}_{n,R}$ be drawn from the ensemble $\mathbb{TZ}_p$ and let $\mathbb{H}_{n,R}^t$ denote the $\overline{n}t \times nt$ principal minor of $\mathbb{H}_{n,R}$. We examine the probability that c is a codeword of $\mathbb{H}_{n,R}^t$, i.e., $P\left(\mathbb{H}_{n,R}^t c = 0\right)$. Now, since $c_{\tau < t-d+1} = 0$, $\mathbb{H}_{n,R}^t c = 0$ is equivalent to

$$\begin{bmatrix} H_1 & 0 & \ldots & \ldots \\ H_2 & H_1 & 0 & \ldots \\ \vdots & \vdots & \ddots & \vdots \\ H_d & H_{d-1} & \ldots & H_1 \end{bmatrix} \begin{bmatrix} c_{t-d+1} \\ c_{t-d+2} \\ \vdots \\ c_t \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \qquad (25)$$

Note that (25) can be equivalently written as follows

$$\begin{bmatrix} C_{t-d+1} & 0 & \ldots & \ldots \\ C_{t-d+2} & C_{t-d+1} & 0 & \ldots \\ \vdots & \vdots & \ddots & \vdots \\ C_t & C_{t-1} & \ldots & C_{t-d+1} \end{bmatrix} \begin{bmatrix} h_1 \\ h_2 \\ \vdots \\ h_d \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \qquad (26)$$

where $h_i = \text{vec}(H_i^T)$, i.e., h_i is a $n\bar{n} \times 1$ column obtained by stacking the columns of H_i^T one below the other, and $C_i \in \{0,1\}^{\bar{n} \times n\bar{n}}$ is obtained from c_i as follows.

$$
C_i = \begin{bmatrix} c_i^T & 0 & \cdots & \cdots \\ 0 & c_i^T & 0 & \cdots \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & c_i^T \end{bmatrix} \tag{27}
$$

Since H_1 is fixed, we will rewrite (26) as

$$
\underbrace{\begin{bmatrix} C_{t-d+1} & 0 & \cdots & \cdots \\ C_{t-d+2} & C_{t-d+1} & 0 & \cdots \\ \vdots & \vdots & \ddots & \vdots \\ C_{t-1} & C_{t-2} & \cdots & C_{t-d+1} \end{bmatrix}}_{\triangleq C} \underbrace{\begin{bmatrix} h_2 \\ h_3 \\ \vdots \\ h_d \end{bmatrix}}_{\triangleq h} = \begin{bmatrix} C_{t-d+2} \\ C_{t-d+3} \\ \vdots \\ C_t \end{bmatrix} h_1, \quad C_{t-d+1} h_1 = 0 \tag{28}
$$

Since $c_{t-d+1} \neq 0$, C_{t-d+1} has full rank $\bar{n}$ and consequently C has full rank $(d-1)\bar{n}$. Since C is an $(d-1)\bar{n} \times (d-1)n\bar{n}$ matrix, its null space has dimension $(d-1)n\bar{n} - (d-1)\bar{n}$. For (28) to hold, h must lie in an $(d-1)n\bar{n} - (d-1)\bar{n}$ dimensional flat which is contained in an $(d-1)n\bar{n} - (d-1)\bar{n} + 1$ dimensional subspace. Using Lemma A.1, we have

$$
P(\mathbb{H}_{n,R}^t c = 0) \leq (1-p)^{\bar{n}(d-1)-1} \tag{29}
$$

$$
\implies P\left(w_{\min,d} < \alpha nd\right) \leq (1-p)^{\bar{n}(d-1)-1} \sum_{w' \leq \alpha nd} \binom{nd}{w'}
$$

$$
\leq (1-p)^{\bar{n}(d-1)-1} 2^{ndH(\alpha)}
$$

$$
= \eta 2^{-nd((1-R)\log_2(1/(1-p))-H(\alpha))} \tag{30}
$$

where $\eta = (1-p)^{-\bar{n}-1}$. Similarly,

$$
P\left(N_{w,d} > 2^{\theta w}\right) \leq 2^{-\theta w} \mathbb{E} N_{w,d}
$$

$$
\leq \eta 2^{-\theta w} \binom{nd}{w} (1-p)^{\bar{n}d}
$$

$$
\leq \eta 2^{-nd(\theta w/nd - H(w/nd) + (1-R)\log_2(1/(1-p)))} \tag{31}
$$

For convenience, define

$$\delta_1 = (1 - R)\log_2(1/(1-p)) - H(\alpha)$$

$$\delta_{2,w} = \theta \frac{w}{nd} - H\left(\frac{w}{nd}\right) + (1 - R)\log_2(1/(1-p))$$

We need to choose θ such that $\delta_{2,w} > \delta > 0$ for all $\alpha \leq \frac{w}{nd} \leq 1$. Now, define

$$\theta^* = \max_{x \geq \alpha} \frac{H(x) - (1 - R)}{x} \tag{32}$$

Then for each $\theta > \theta^*$, there is a $\delta > 0$ such that $\delta_{2,w} > \delta$ for all $\alpha nd \leq w \leq nd$. A simple calculation gives $\theta^* = \log_2\left(\frac{1}{2^{1-R}-1}\right)$. For such a choice of $\theta > \theta^*$, continuing from (31), we have

$$P\left(\exists \, \alpha nd \leq w \leq nd \ \ni \ N_{w,d} > 2^{\theta w}\right) \leq nd2^{-nd\delta} \tag{33}$$

for some $\delta' > 0$. For some fixed d_o large enough, applying a union bound over $d \geq d_o$ to (30) and (33), we get

$$P\left(\exists \, d \geq d_o \ \ni \ w_{\min,d} < \alpha nd \text{ or } N_{w,d} > 2^{\theta w}\right) \leq 2^{-\Omega(nd_o)} \tag{34}$$

B. Proofs of Theorems 8.1 and 9.1

1) Proof of Theorem 8.1: The analysis will proceed in two steps. We will first determine a sufficient condition on the number of bits per measurement, nR, that are required to track (3) when these bits are available error free. We will then determine the anytime exponent $n\beta$ needed in decoding these source bits when they are communicated over a noisy channel.

Let $\Delta_{t|\tau} \triangleq x_{max,t|\tau} - x_{min,t|\tau}$ be the uncertainty in x_t using $\{b_{\tau'}\}_{\tau' \leq \tau}$, i.e., quantized measurements up to time τ. For convenience, let $\Delta_t \equiv \Delta_{t|t-1}$. Then, the time update is given by the following Lemma.

Lemma A.2 (Time Update): The time update relating Δ_{t+1} and $\Delta_{t|t}$ is given by $\Delta_{t+1} = \overline{F}\Delta_{t|t} + W\mathbf{1}_{m_x}$

Proof: From the system dynamics in (3), the following is immediate

$$\Delta_{t+1}^{(i)} = W + \max\left\{\left|\pm a_i \Delta_{t|t}^{(1)} + \Delta_{t|t}^{(i+1)}\right|, \left|\Delta_{t|t}^{(i+1)}\right|, \left|a_i \Delta_{t|t}^{(1)}\right|\right\}$$

$$= |a_i|\Delta_{t|t}^{(1)} + \Delta_{t|t}^{(i+1)} + W, \ i \leq m-1$$

$$\Delta_{t+1}^{(m)} = |a_m|\Delta_{t|t}^{(1)} + W$$

In short, the above equations amount to $\Delta_{t+1} = \overline{F}\Delta_{t|t} + W\mathbf{1}_{m_x}$. ∎

Towards the measurement update, the observer simply quantizes the measurements y_t according to a 2^{nR}–regular lattice quantizer with bin width δ, i.e., the quantizer is defined by $Q : \mathbb{R} \mapsto \{0, 1, \ldots, 2^{nR} - 1\}$, where $Q(x) = \lfloor \frac{x}{\delta} \rfloor \bmod 2^{nR}$. In order for this to work, we need $\delta 2^{nR} \geq \Delta_t^{(1)}$ for any time t. Assuming that the rate, R, is large enough, we will first find the steady state value of the recursion for Δ_t, which we then use to determine R. At each time t, the observer can communicate the measurement y_t to within an uncertainty of δ, i.e., the estimator knows that the measurement lies in an interval of width δ. Adding to this the effect of the observation noise, $-\frac{V}{2} \leq v_t \leq \frac{V}{2}$, the estimator knows $x_t^{(1)}$ to within an uncertainty of $\Delta_{t|t}^{(1)} = \delta + V$. Note that $\Delta_{t|t}^{(i)} = \Delta_t^{(i)}$ for $i \neq 1$. Combining this observation with Lemma A.2, it is straightforward to see that Δ_t converges, to say Δ_{tu}, in exactly m_x time steps, i.e., $\Delta_t = \Delta_{tu}$ for all $t \geq m_x$. The subscript 'tu' in Δ_{tu} denotes 'time update'. The following result is now immediate.

Lemma A.3 (Steady State value of Δ_t): $\Delta_{tu} = (\delta + V)L_u a + W L_u \mathbf{1}_{m_x}$, where $a = [|a_1|, \ldots, |a_m|]^T$ and $L_u = [\ell_{ij}]_{1 \leq i,j \leq m}$ with $\ell_{ij} = \mathbb{I}_{i \leq j}$.

Now, we need to go back and calculate R. So we just need $\delta 2^{nR} \geq \max\left\{\Delta_0^{(1)}, \Delta_1^{(1)}, \ldots, \Delta_{m_x}^{(1)}\right\}$. Further, a simple calculation gives $\lim_{\delta \to \infty} \frac{\Delta_i^{(1)}}{\delta} = |a_1| + \ldots + |a_i|$. The minimum rate is thus given by $\frac{1}{n}\log_2 \sum_{i=1}^m |a_i|$ and this completes the proof Theorem 8.1.

2) Proof of Theorem 9.1: The proof is very similar to that of Theorem 8.1. The observations are quantized as follows. At any time, for $1 \leq i \leq q$, the ith component of the measurement vector is quantized using a 2^{nR_i}–regular lattice quantizer with bin width δ_i. The remaining components of the measurement vector are ignored. The overall rate, R, is then given by $R = R_1 + R_2 \ldots + R_q$. The time update again is given by $\Delta_{t+1} = \overline{F}\Delta_{t|t} + W\mathbf{1}_{m_x}$. The limiting values of $\{R_i\}_{i=1}^q$ are obtained by letting $\delta_1 \to \infty$ and $\frac{\delta_i}{\delta_{i+1}} \to \infty$. An argument similar to the one in the previous section gives the following threshold, $R_i \geq \frac{1}{n}\max\left\{0, \log\left(|a_{i,1}| + |a_{i,2}| + \ldots + |a_{i,\ell_i}|\right)\right\}$.

C. The Minimum Volume Ellipsoid

Lemma A.4 (Theorem 6.1 [34]): The minimum volume ellipsoid $\mathcal{E}(\hat{P}, c)$ covering

$$\left\{ x \in \mathbb{R}^m | x \in \mathcal{E}(P, 0), \gamma\sqrt{h^T P h} \leq \langle h, x \rangle \leq \delta\sqrt{h^T P h} \right\}$$

where $|\delta| \geq |\gamma|$, is given by

$$\hat{P} = bP - (b - a)\frac{Phh^T P}{h^T P h}, \quad c = \xi\frac{Ph}{\sqrt{h^T P h}} \tag{35}$$

where

1) If $\gamma\delta < -\frac{1}{m}$, then $\xi = 0$, $a = b = 1$

2) If $\gamma + \delta = 0$ and $\gamma\delta > -\frac{1}{m}$, then

$$\xi = 0, \quad a = m\delta^2, \quad b = \frac{m(1 - \delta^2)}{m - 1}$$

3) If $\gamma + \delta \neq 0$ and $\gamma\delta > -\frac{1}{m}$, then

$$\xi = \frac{m(\gamma + \delta)^2 + 2(1 + \gamma\delta) - \sqrt{D}}{2(m + 1)(\gamma + \delta)}$$

$$a = m(\xi - \gamma)(\delta - \xi), \quad b = \frac{a - a\gamma^2}{a - (\xi - \gamma)^2}$$

$$\text{where } D = m^2(\delta^2 - \gamma^2)^2 + 4(1 - \gamma^2)(1 - \delta^2)$$

If $|\delta| < |\gamma|$, change x to $-x$ and apply the above result. And it is easy to verify that $\hat{P}$ is indeed positive semidefinite. Also, a quick calculation shows that $\gamma \leq \xi \leq \delta$. This confirms the intuition that the center of the minimum volume ellipsoid lies within the slab.

D. Proof of Theorem 8.3

The proof is in the same spirit as that of Theorem 8.1. We will first determine a sufficient condition on the number of bits per measurement, nR, that are required to track (3) when these bits are available error free. We will then determine the anytime exponent $n\beta$ needed in decoding these source bits when they are communicated over a noisy channel.

Consider the time update in (18a). Let P_t^{ij} denote the $(i, j)^{th}$ element of P_t, then the time

update implies

$$P_{t+1}^{ii} = (1 + \epsilon')\left(a_i^2 P_{t|t}^{11} + P_{t|t}^{i+1,i+1} - a_i P_{t|t}^{1,i+1} - a_i P_{t|t}^{i+1,1}\right) + \frac{W^2}{4\epsilon'}, \quad 1 \le i \le m_x - 1 \quad (36a)$$

$$P_{t+1}^{m_x,m_x} = (1 + \epsilon')a_{m_x}^2 P_{t|t}^{11} + \frac{W^2}{4\epsilon'} \quad (36b)$$

Since the matrix $P_{t|t}$ is positive semidefinite, we have $P_{t|t}^{1,i+1} = P_{t|t}^{i+1,1}$ and $\left(P_{t|t}^{1,i+1}\right)^2 \le P_{t|t}^{11} P_{t|t}^{i+1,i+1}$. Using this in (36a), for $1 \le i \le m_x - 1$, we get

$$P_{t+1}^{ii} \le (1 + \epsilon')\left(|a_i|\sqrt{P_{t|t}^{11}} + \sqrt{P_{t|t}^{i+1,i+1}}\right)^2 + \frac{W^2}{4\epsilon'} \quad (37)$$

This prompts us to bound the recursion (18) by bounding the diagonal elements of P_t. Now, considering the measurement update (18b), it is easy to see that

$$P_{t|t}^{11} = a_t P_t^{11} \quad (38a)$$

$$a_t P_t^{ii} \le P_{t|t}^{ii} \le b_t P_t^{ii} \quad (38b)$$

We will first show that $b_t \le \frac{m_x}{m_x - 1}$.

Lemma A.5: $b_t \le \frac{m_x}{m_x - 1}$

Proof: To prove this, consider the setup of Lemma A.4 and suppose $|\delta| \ge |\gamma|$. Then, in cases 1) and 2), it is clear that $b \le \frac{m}{m-1}$ since $|\delta|, |\gamma| \le 1$. In case 3), we have

$$b = \frac{1 - \gamma^2}{1 - (\xi - \gamma)^2/a} = \frac{1 - \gamma^2}{1 - \frac{\xi - \gamma}{m(\delta - \xi)}} \le \frac{1}{1 - \frac{\xi - \gamma}{m(\delta - \xi)}}$$

It suffices to show that $\xi - \gamma \le \delta - \xi$. This easily follows from the formulae in case 3). The proof for the case when $|\delta| \le |\gamma|$ is obtained by replacing ξ with $-\xi$. ∎

Like in Section B, the observer quantizes the measurements y_t according to a $2^{nR}-$regular lattice quantizer with bin width δ. In order for the controller to know y_t to within a resolution of δ, it is not hard to see that one needs $\delta 2^{nR} > 2\sqrt{P_t^{11}} + v$. We begin by assuming that the rate R is large enough to provide the same resolution δ on y_t at each time t. The actual rate required to accomplish this will be calculated determining an asymptotic upper bound on P_t^{11}. So, at time t, the controller knows that y_t to within a resolution δ and hence $x_t^{(1)}$ to within a resolution of $\delta + V$. Suppose $\sqrt{P_t^{11}}\gamma_t \le x_t^{(1)} \le \sqrt{P_t^{11}}\delta_t$, where $\sqrt{P_t^{11}}(\delta_t - \gamma_t) \le \delta + V$. Then

using Lemma A.4 and noting that $\gamma_t \leq \xi_t \leq \delta_t$, we have

$$a_t = m_x(\xi_t - \gamma_t)(\delta_t - \xi_t) \leq \frac{m_x}{4}(\delta_t - \gamma_t)^2$$

$$\implies P_t^{11}a_t \leq \frac{m_x}{4}(\delta + V)^2$$

Using this in (38a), we get

$$P_{t|t}^{11} = a_t P_t^{11} \leq \frac{m_x}{4}(\delta + V)^2 \tag{39}$$

Combining Lemma A.5 and (39), we get

$$\sqrt{P_{t|t}^{11}} \leq \frac{\sqrt{m_x}}{2}(\delta + V) \tag{40a}$$

$$\sqrt{P_{t|t}^{ii}} \leq \sqrt{\frac{m_x}{m_x - 1}}\sqrt{P_t^{ii}}, \ \ i \neq 1 \tag{40b}$$

In the following Lemma, we will develop an upper bound on the diagonal elements of P_t which will help us determine an upper bound on P_t^{11}.

Lemma A.6: Let $\Delta_{e,0} \in \mathbb{R}^{m_x}$ be such that $\Delta_{e,0}^{(i)} = P_0^{ii}$ for $1 \leq i \leq m_x$ and suppose its evolution is governed by

$$\Delta_{e,t+1} = (1 + \epsilon')^{\frac{1}{2}}\overline{F}\Delta_{e,t|t} + \frac{W}{2\sqrt{\epsilon'}}\mathbf{1}_{m_x}$$

$$\Delta_{e,t|t}^{(i)} = \begin{cases} \delta + V & i = 1 \\ \theta\Delta_{e,t}^{(i)} & i \neq 1 \end{cases}$$

where $\theta = \sqrt{\frac{m_x}{m_x - 1}}$. Then $\sqrt{P_t^{ii}} \leq \Delta_{e,t}^{(i)}$ and $\sqrt{P_{t|t}^{ii}} \leq \Delta_{e,t|t}^{(i)}$ for all t and $1 \leq i \leq m_x$.

Proof: The proof follows by combining the observations from (36), (37), (40). ∎

Note that the recursion for $\Delta_{e,t}$ above is very similar to that for Δ_t in Section B. So, the steady state value of $\Delta_{e,t}^{(1)}$ can be determined by a calculation similar to that in Lemma A.3. The desired threshold for R is obtained by letting $\delta \to \infty$ for a fixed ϵ'. Since ϵ' can be made arbitrarily small, we get the following bound on R

$$R > \frac{1}{n}\log\left[\sqrt{m_x}\sum_{i=1}^{m_x}|a_i|\theta^{i-1}\right]$$

Now, we need to determine the exponent needed to track (3) with a bounded mean squared error. In the absence of any measurements, it is easy to see from (18a) that the growth of P_t is determined by the spectral radius of $\sqrt{1 + \epsilon'}F$. Since ϵ' can be made arbitrarily small, inorder to track (3) with a bounded mean squared error, we need an anytime exponent $n\beta > 2\log\rho(F)$. This completes the proof.

E. The Limiting Case

There are several bounds in the Mathematics literature on the roots of a polynomial in terms of the polynomial coefficients, a standard and near optimal bound being the Fujiwara's bound which we state below.

Lemma A.7 (Fujiwara's Bound): Consider the monic polynomial with complex coefficients $f(z) = z^m + c_1 z^{m-1} + \ldots + z_m$ and let $\rho(f)$ denote the largest root in magnitude. Then

$$\rho(f) \le K(f) = 2\max\left\{|c_1|, |c_2|^{\frac{1}{2}}, \ldots, |c_{m-1}|^{\frac{1}{m-1}}, \left|\frac{c_m}{2}\right|^{\frac{1}{m}}\right\}$$

We will detail the proof for the case of scalar measurements. The extension to the vector measurements will then suggest itself. Let F is any m_x-dimensional square matrix and $f(z)$ denotes its characteristic polynomial. Then the following bounds hold (for details see [35])

$$\rho(F) \le \rho(\overline{F}) \le \frac{\rho(F)}{\sqrt[m]{2} - 1}, \ \ K(f) \le 2\rho(\overline{F}) \tag{41}$$

By the hypothesis of the Lemma, the eigen values of F_n are of the form $\{\mu_i^n\}_{i=1}^{m_x}$. To emphasize the fact that F depends on n, we write it as F_n and a_i as $a_{i,n}$. Recall that the characteristic polynomial of F_n is given by $f_n(z) = z^{m_x} + a_{1,n} z^{m_x-1} + \ldots + a_{m_x,n}$. Let $\mathcal{I}_u \triangleq \{i \mid |\mu_i| \ge 1\}$, then the following is easy to prove

$$\lim_{n\to\infty} \frac{|a_{i,n}|}{|a_{|\mathcal{I}_u|,n}|} = 0, \ i \ne |\mathcal{I}_u|, \ \lim_{n\to\infty} \frac{1}{n}\log_2\left|a_{|\mathcal{I}_u|,n}\right| = \sum_{i\in\mathcal{I}_u}\log_2|\mu_i| \tag{42}$$

From (42), it is obvious that $\lim_{n\to\infty} R_n = \sum_{i\in\mathcal{I}_u}\log_2|\mu_i|$. The asymptotics of $R_{e,n}$, $R_{v,n}$ and $R_{ev,n}$ can be similarly derived. Also, from (41), it is clear that $\lim_{n\to\infty}\frac{1}{n}\log\rho\left(\overline{F}_n\right) = \lim_{n\to\infty}\frac{1}{n}\log\rho\left(F_n\right)$. The asymptotics of β_n and $\beta_{v,n}$ now follow immediately.