Linear Error Correcting Codes with Anytime Reliability

Ravi Teja Sukhavasi and Babak Hassibi

Abstract

We consider rate $R = \frac{k}{n}$ causal linear codes that map a sequence of k-dimensional binary vectors $\{b_t\}_{t=0}^{\infty}$ to a sequence of n-dimensional binary vectors $\{c_t\}_{t=0}^{\infty}$, such that each c_t is a function of $\{b_\tau\}_{\tau=0}^{t}$. Such a code is called anytime reliable, for a particular binary-input memoryless channel, if at each time instant t, and for all delays $d \geq d_o$, the probability of error $P\left(\hat{b}_{t-d|t} \neq b_{t-d}\right)$ decays exponentially in d, i.e., $P\left(\hat{b}_{t-d|t} \neq b_{t-d}\right) \leq \eta 2^{-\beta nd}$, for some $\beta > 0$. Anytime reliable codes are useful in interactive communication problems and, in particular, can be used to stabilize unstable plants across noisy channels. Schulman proved the existence of such codes which, due to their structure, he called tree codes in [1]; however, to date, no explicit constructions and tractable decoding algorithms have been devised. In this paper, we show the existence of anytime reliable "linear" codes with "high probability", i.e., suitably chosen random linear causal codes are anytime reliable with high probability. The key is to consider time-invariant codes (i.e., ones with Toeplitz generator and parity check matrices) which obviates the need to union bound over all times. For the binary erasure channel we give a simple ML decoding algorithm whose average complexity is constant per time iteration and for which the probability that complexity at a given time t exceeds KC^3 decays exponentially in C. We show the efficacy of the method by simulating the stabilization of an unstable plant across a BEC, and remark on the tradeoffs between the utilization of the communication resources and the control performance.

I. INTRODUCTION

Shannon's information theory, in large part, is concerned with one-way communication of a message, that is available in its entirety, over a noisy communication network. There are increasingly many

applications such as networked control systems [2] and distributed computing [1] where communication is not one-way but interactive. A networked control system is characterized by the measurement unit begin separated from the control unit by a communication channel. The interactive nature of communication can be understood by observing that the measurements to be encoded are determined by the control inputs, which in turn, are determined by the encoded measurements received by the controller. Block encoding of the measurements is not applicable anymore because the controller needs real time information about the system so that an appropriate control input can be applied. This is especially critical when the system being controlled is open loop unstable. Any encoding-decoding delay translates into the system growing increasingly unstable. Hence, the desired reliability of communication is determined by the quality of the control input needed to stabilize the system.

In the context of rate-limited deterministic channels, significant progress has been made (see eg., [3], [4]) in understanding the bandwidth requirements for stabilizing open loop unstable systems. When the communication channel is stochastic, [5] provides a necessary and sufficient condition on the communication reliability needed over a channel that is in the feedback loop of an unstable scalar linear process, and proposes the notion of anytime capacity as the appropriate figure of merit for such channels. In essense, 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 [5], and in a different context, [1] 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. [1] proves the existence of nonlinear tree codes and gives no explicit constructions and/or efficient decoding algorithms. Much more recently [6] proposed efficient error correcting codes for unstable systems where the state grows only polynomially large with time. So, for linear unstable systems that have an exponential growth rate, all that is known in the way of error correction is the existence of tree codes which are, in general, non-linear. When the state of an unstable scalar linear process is available at the encoder, [7] and [8] 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, 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 [9]. 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 $d \geq d_o$, $P(\hat{b}_{t-d|t} \neq b_{t-d}) \leq \eta 2^{-\beta nd}$ 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 III, we come up with a sufficient condition for anytime reliability in terms of the weight distribution of the code. In Section IV, we introduce the ensemble of time invariant codes and use the results from Section III to prove that time invariant codes with anytime reliability exist with a high probability. In Section V, we present a simple decoding algorithm for the BEC and present simulations in Section VI to demonstrate the efficacy of the algorithm.

II. BACKGROUND AND PROBLEM SETUP

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. So, we will illustrate the kind of communication reliability needed for control by analyzing the open loop estimation of the following random walk.

A toy example: Consider estimating the following random walk, $x_{t+1} = \lambda x_t + w_t$, where $w_t = \pm 1$ w.p $\frac{1}{2}$, $x_0 = 0$ and $|\lambda| > 1$. Suppose an observer observes x_t and communicates over a noisy communication channel to an estimator. The observer clearly needs to communicate one bit telling whether w_t is $+1$ or -1[1]. Now 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 $P_{d,t}^e = P\left(\text{argmin}_j(\hat{w}_{j|t} \neq \hat{w}_j) = t - d + 1\right)$, i.e., the position of the earliest erroneous $\hat{w}_{j|t}$ is

---

[1]If the channel is binary input, it should clearly allow at least one usage for each time step of the system evolution

at time $j = t - d + 1$. From (1), we can write $\mathbb{E}\left|x_{t+1} - \hat{x}_{t+1|t}\right|^2$ as

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

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

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

where d_o and t_o are constants that do no depend on t, d. In the above example, noise is discrete valued. If it is not discrete, for eg, a uniform random variable on a bounded interval, then apart from the reliability criterion above, one would also need a minimum rate of communication that depends on the size of $\log_2 |\lambda|$. But codes that communicate at a positive rate and satisfy the above reliability criterion have been more or less elusive. Consider the following unstable scalar linear process
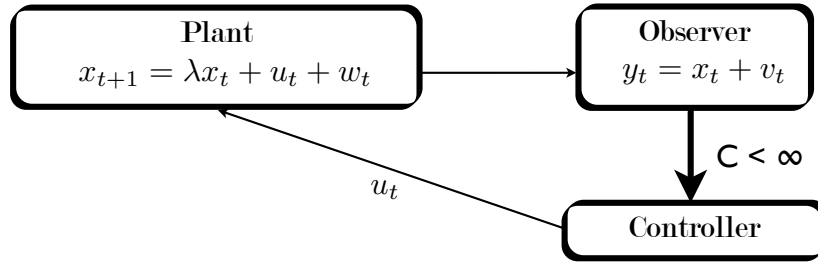

Fig. 1: System model

$$x_{t+1} = \lambda x_t + u_t + w_t, \quad y_t = x_t + v_t \tag{3}$$

where $|\lambda| > 1$, u_t is the control input and, $|w_t| < \frac{W}{2}$ and $|v_t| < \frac{V}{2}$ are bounded process and measurement noise variables. 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. Naturally, the measurements $y_{0:t-1}$ will need to be encoded by the observer to provide protection from the noisy channel while the controller will need to decode the channel output 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.
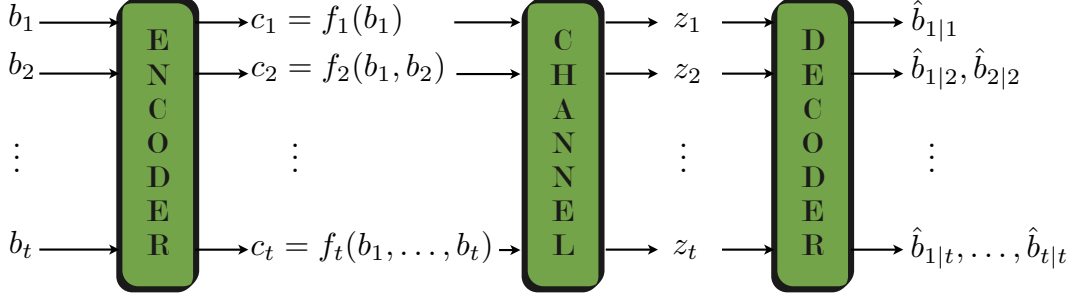
Fig. 2: Causal encoding and decoding

Suppose one time step of system evolution in (3) corresponds to n channel uses[2]. 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, $b_t \in \{0,1\}^k$, that is a causal function of the measurements, i.e., it depends only on $y_{0:t}$. Then the channel encoder causally encodes $b_{0:t} \in \{0,1\}^{kt}$ to generate the n channel inputs $c_t \in \{0,1\}^n$. 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. Note that we do not assume any channel feedback. Then using the lattice quantizer argument presented in [5], in the limit of large n and large λ, we have the following sufficient condition on the performance of the encoder-decoder pair so that the unstable process in (3) can be stabilized

Lemma 2.1 (Theorem 5.2 [5]): It is possible to control the unstable scalar process (3) over a noisy communication channel so that $\limsup_t \mathbb{E}|x_t|^m < \infty$ if, for some rate $R > \frac{1}{n} \log_2 |\lambda|$ and exponent $\beta > \frac{m}{n} \log_2 |\lambda|$, we have

$$P\left(\min\{\tau : \hat{b}_{\tau|t} \neq b_\tau\} = t - d + 1\right) \leq \eta 2^{-\beta n d}, \forall \, d \geq d_o, \, t > t_o$$

where d_o and t_o are constants independent of d, t.

In what follows, we will demonstrate causal linear codes which under maximum likelihood decoding achieve such exponential reliabilities.

---

[2]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.

III. Linear Anytime Codes - A Sufficient Condition

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 1 (Causal Linear Code): A causal linear code is a sequence of linear maps $f_\tau : \{0,1\}^{k\tau} \mapsto \{0,1\}^n$ and hence can be represented as

$$f_\tau(b_{1:\tau}) = G_{\tau 1}b_1 + G_{\tau 2}b_2 + \ldots + G_{\tau\tau}b_\tau \tag{4}$$

where $G_{ij} \in \{0,1\}^{n \times k}$

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 $R = \frac{k}{n}$. One can alternately represent a causal linear code by an infinite dimensional block lower triangular generator matrix $\mathbb{G}_{n,R}$ or equivalently as an infinite dimensional block lower triangular parity check matrix, $\mathbb{H}_{n,R}$.

$$\mathbb{G}_{n,R} = \begin{bmatrix} G_{11} & 0 & \ldots & \ldots & \ldots \\ G_{21} & G_{22} & 0 & \ldots & \ldots \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ G_{\tau 1} & G_{\tau 2} & \ldots & G_{\tau\tau} & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix}, \quad \mathbb{H}_{n,R} = \begin{bmatrix} H_{11} & 0 & \ldots & \ldots & \ldots \\ H_{21} & H_{22} & 0 & \ldots & \ldots \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ H_{\tau 1} & H_{\tau 2} & \ldots & H_{\tau\tau} & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix} \tag{5}$$

where $H_{ij} \in \{0,1\}^{\bar{n} \times n}$ and $\bar{n} = n(1-R)^3$. In fact, we present all our results in terms of the parity check matrix. Before proceeding further, it is useful to introduce some notation

A. Notation

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

where $\|c\|$ denotes the Hamming weight of c.

---

[3]While for a given generator matrix, the parity check matrix is not unique, when $G_{n,R}$ is block lower, it is easy to see that $H_{n,R}$ can also be chosen to be block lower.

B. A Sufficient Condition

The objective is to study the existence of causal linear codes which under ML decoding guarantee

$$P_{d,t}^e \le \eta 2^{-\beta d}, \ \forall \ t, \ d \ge d_o \tag{6}$$

where d_o is a constant independent of d, t. In what follows, we will develop a sufficient condition for a linear code to be anytime reliable in terms of its weight distribution. 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 $\tau \le t$. 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 $\mathbb{H}_{n,R}^t$ having full rank so that $\hat{c}$ can be uniquely mapped to a transmitted sequence $\hat{b}$. Then we have

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

Now, it is well known (for eg, see [10]) that, under maximum likelihood decoding, $P\left(0 \text{ is decoded as } c \right) \le \zeta^{\|c\|}$, where ζ is the Bhattacharya parameter, i.e.,

$$\zeta = \int_{-\infty}^{\infty} \sqrt{p(z|X=1)p(z|X=0)} dz$$

where, z and X denote the channel output and input respectively. From (7a), it follows that

$$P_{t,d}^e \le \sum_{w_{\min,d}^t \le w \le nd} N_{w,d}^t \zeta^w$$

If $w_{\min,d}^t \ge \alpha nd$ and $N_{w,d}^t \le 2^{\theta w}$ for some $\theta < \log_2(1/\zeta)$, then

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

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

$$w_{\min,d}^t \ge \alpha nd, \quad N_{w,d}^t \le 2^{\theta w} \ \forall \ t, \ d \ge d_o \tag{9a}$$

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

Definition 2 (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 $d \geq d_o$.

Also, we say that a code $\mathbb{H}_{n,R}$ is $(R, \beta, d_o)-$anytime reliable if, under ML decoding

$$P_{t,d}^e \leq \eta 2^{-\beta nd}, \ \ \forall \ t > 0, \ \ d \geq d_o \tag{9b}$$

where, d_o is a constant independent of t, d.

IV. LINEAR ANYTIME CODES - EXISTENCE

We will begin by proving the existence of such codes over a finite time horizon, T, i.e., $P_{d,t}^e \leq \eta 2^{-\beta d}, \ \forall \ t \leq T, \ d \geq d_o$. We will then prove their existence for all time.

A. Finite Time Horizon

Over a finite time horizon, T, a causal linear code is represented by a block lower triangular parity check matrix $\mathbb{H}_{n,R,T} \in \{0,1\}^{\bar{n}T \times nT}$. The following Theorem guarantees the existence of a $\mathbb{H}_{n,R,T}$ such that (9) is true for all $t \leq T$.

Theorem 4.1 (Appropriate Weight Distribution): For each time $T > 0$, rate $R > 0$, $\alpha < H^{-1}(1 - R)$ and $\theta > \log_2(1/(2^{1-R} - 1))$, there exists a causal linear code $H(n, k, T)$ that has $(\alpha, \theta, d_o)-$anytime distance, where d_o is a constant independent of d, t and T.

$H^{-1}(1 - R)$ is the smaller root of the equation $H(x) = 1 - R$, where $H(.)$ is the binary entropy function. The proof is by induction and is detailed in the Appendix. Theorem 4.1 proves the existence of finite dimensional causal linear codes, $\mathbb{H}_{n,R,T}$, that are anytime reliable for decoding instants upto time T. In the following subsection, we demonstrate the existence of infinite dimensional causal linear codes, $\mathbb{H}_{n,R}$, that are anytime reliable for all decoding instants. We also show that such codes drawn from an appropriate ensemble are anytime reliable with a high probability.

B. Time Invariant Codes

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 (5) by setting $H_{ij} = H_{i-j+1}$ for $i \geq j$. Due to the Toeplitz structure, we have the following invariance, $w_{\min,d}^t = w_{\min,d}^{t'}$ and $N_{w,d}^t = N_{w,d}^{t'}$ for all t, t'. The code $\mathbb{H}_{n,R}^{TZ}$ 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 [1]. This time invariance allows one to prove that such codes which are anytime reliable are abundant.

Definition 3 (The ensemble $\mathbb{TZ}_p$): The ensemble $\mathbb{TZ}_p$ of time-invariant codes, $\mathbb{H}_{n,R}^{TZ}$, is obtained as follows, H_1 is any 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.

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

Theorem 4.2 (Abundance of time-invariant codes): For each $R > 0$, $\alpha < H^{-1}\left[(1-R)\log_2(1/(1-p))\right]$ and $\theta > -\log_2\left[(1-p)^{-(1-R)} - 1\right]$, we have

$$P\left(\mathbb{H}_{n,R}^{TZ} \text{ has } (\alpha, \theta, d_o) - \text{anytime distance}\right) \geq 1 - 2^{-\Omega(nd_o)} \tag{10}$$

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. To determine the values of R that will satisfy (8), note that we need

$$\log_2(1/(2^{1-R} - 1)) < \log_2(1/\zeta) \implies R < 1 - \log_2(1 + \zeta)$$

With this observation, we have the following Corollary.

Corollary 4.3: For any rate R and exponent β such that

$$R < 1 - \frac{\log_2(1 + \zeta)}{\log_2(1/(1-p))} \quad \text{and} \quad \beta < H^{-1}(1 - R)\left(\log_2\left(\frac{1}{\zeta}\right) - \log_2\left[(1-p)^{-(1-R)} - 1\right]\right) \tag{11a}$$

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

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

Note that by choosing p small, we can trade off better rates and exponents with sparser parity check matrices. Note that for BEC(ϵ), $\zeta = \epsilon$ and for BSC(ϵ), $\zeta = 2\sqrt{\epsilon(1-\epsilon)}$. For BSC($\epsilon$) with $p = \frac{1}{2}$, the threshold for rate in Corollary 4.3 becomes $R < 1 - 2\log_2(\sqrt{\epsilon} + \sqrt{1-\epsilon})$. It turns out that this can be strengthened as follows.

Theorem 4.4 (Tighter bounds for BSC(ϵ)): For any rate R and exponent β such that

$$R < 1 - H(\epsilon), \quad \beta < KL\left(H^{-1}(1-R)\| \min\{\epsilon, 1-\epsilon\}\right)$$

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

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

C. Stabilizable Region

Using the thresholds obtained in Corollary 4.3 and Theorem 4.4, we can discuss the range of $|\lambda|$ for which the mth moment of x_t in (3) can be stabilized over some common channels. Using Lemma 2.1, an anytime reliable code with rate R and exponent β can stabilize the process in (3) for all λ such that

$$\log_2|\lambda| < \min\{nR, \frac{n\beta}{m}\}$$

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

$$\log_2|\lambda| < \log_2|\lambda_{max}| = \sup_{R<R_\zeta, \beta<\beta_{\zeta,R}} \min\{nR, \frac{\beta n}{m}\} \tag{12}$$

where $R_\zeta = 1 - \log_2(1+\zeta)$ and $\beta_{\zeta,R} = H^{-1}(1-R)\left(\log_2\left(\frac{1}{\zeta}\right) + \log_2\left(2^{1-R} - 1\right)\right)$ are obtained by setting $p = \frac{1}{2}$ in (11). For the BSC(ϵ), using Theorem 4.4, one can tighten (12) by replacing R_ζ with $1 - H(\epsilon)$ and $\beta_{\zeta,R}$ with $KL\left(H^{-1}(1-R)\| \min\{\epsilon, 1-\epsilon\}\right)$. For $m = 2$, the stabilizable region for the BEC and BSC is shown in Fig 3 where $|\lambda_{max}|^{\frac{1}{n}}$ is plotted against the channel parameter.

V. Decoding over the BEC

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. We will show that the average complexity of
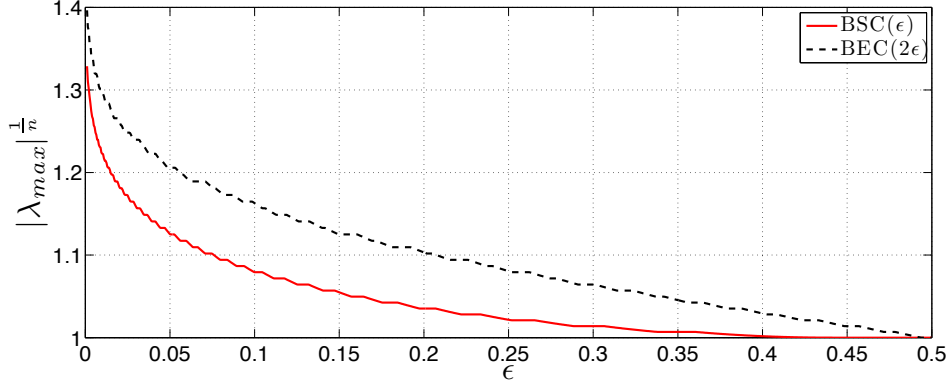
Fig. 3: Comparing the stabilizable regions of BSC and BEC using linear codes

the decoding operation at any time t is constant and that it being larger than KC^3 decays exponentially in C. 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 corresponding channel outputs. For each $z_\tau \in \{0, 1, e\}^n$, let $z_{\tau j}$ denote the jth entry for $1 \leq j \leq n$. Similarly, define $c_{\tau j}$. Let z_e denote the erasures in z and let H_e denote the columns of $\mathbb{H}_{n,R}^t$ 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 $\mathbb{H}_{n,R}$ excluding H_e. For $z_{\tau j} \in \tilde{z}_e$, $z_{\tau j} = c_{\tau j}$. 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

$$H_e z_e = s \tag{13}$$

Due to the lower triangular nature of H_e, unlike in the case of traditional block coding, (13) will typically not have a unique solution, since H_e will typically not be full 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 $z_{e,2}$. Let $H_e = [H_{e,1} H_{e,2}]$ and let $H_{e,2}^\perp$ denote a matrix whose rows form a basis for the subspace othorgonal to the column span of $H_{e,2}$, in short, $H_{e,2}^\perp H_{e,2} = 0$. Then we can write (13) as

$$H_{e,1} z_{e,1} + H_{e,2} z_{e,2} = s \implies H_{e,2}^\perp H_{e,1} z_{e,1} = H_{e,2}^\perp s$$

If $H_{e,2}^\perp H_{e,1}$ has full column rank, then $z_{e,1}$ can be recovered exactly. The decoding algorithm now suggests itself, i.e., find the largest possible $H_{e,1}$ such that $H_{e,2}^\perp H_{e,1}$ is full rank. Suppose the earliest entry in

---
Algorithm 1 Decoder for the BEC
---
1) Using the notation above, at time t, identify H_e and let $H_e = [h_{e,1}, \ldots, h_{e,\ell}]$. Determine

$$j^* = \underset{1 \leq j \leq \ell}{\text{argmax}} \left(h_{e,j+1:\ell}^\perp h_{e,1:j} \text{ has full column rank} \right) \tag{14}$$

2) Determine $[z_{S_t^*}]^T = \left[h_{e,j^*+1:\ell}^\perp h_{e,1:j^*} \right]^{-1} h_{S_t^c} z_{S_t^c}$ and increment t.
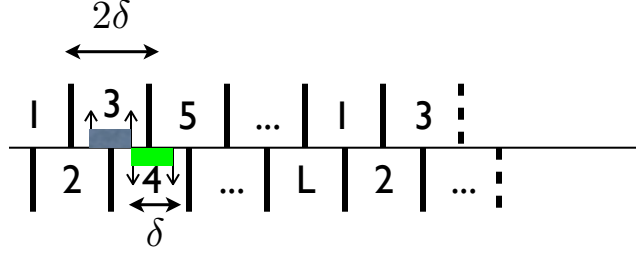---

Fig. 4: L-regular lattice quantizer with bin width δ

z_e, i.e., j^* in (14), corresponds to time $t - d + 1$, then the complexity of the above decoding operation is $\text{O}(d^3)$. Now, 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 $P_{d-1,t-1}^e \leq \eta 2^{-\alpha(d-1)}$. Hence, the average decoding complexity is atmost $K \sum_{d>0} d^3 2^{-\alpha d}$ which is bounded and is independent of t. In particular, the probability of the decoding complexity being Kd^3 is atmost $\eta 2^{-\alpha d}$. The decoder is easy to implement and its performance is simulated in the next section.

Note that the encoding complexity per time iteration increases linear with time. This can also be made constant on average if the decoder can send acks back to the encoder whenever it decodes correctly.

VI. SIMULATIONS

Consider stabilizing the scalar unstable process of (3) with $\lambda = 2$, and w_t and v_t being uniform on $[-30, 30]$ and $[-1, 1]$ respectively, over a binary erasure channel with erasure probability $\epsilon = 0.3$. Also, let $n = 15$, i.e., we get 15 channel uses per time instant. To quantize the measurements, we shall use a L-regular lattice quantizer with bin width $\delta > V$ which is depicted in Fig 4 (see for e.g., [5]) If at some time t, $x_t \in (-\delta, \delta)$ and the controller is aware of it, then, $y_{t+1} \in \left[-\frac{V+W}{2} - |\lambda|\delta, |\lambda|\delta + \frac{W+V}{2} \right]$ and hence can lie in atmost one of $\frac{V+W+2|\lambda|\delta}{\delta}$ possible bins. So, if $L > \frac{V+W+2|\lambda|\delta}{\delta}$, then the observer can encode the bin label of y_{t+1} as $b_t \in \{0,1\}^k$, while the channel coder produces the channel inputs $c_t \in \{0,1\}^n$. Clearly we need $2^k \geq L$. So, for $\delta = V$, $L = 35$ and hence $k = 6$, the associated code rate $R = 6/15 < 1 - \log_2(1 + 0.3) = 0.6215$. Using Lemma 2.1, inorder to stabilize $|x_t|$, one

needs a code with exponent $\beta \geq \frac{1}{n} = 0.0667$. Using Corollary 4.3, causal linear codes exist for $\beta < \beta^* = H^{-1}(1 - R)\left(\log_2\left(\frac{1}{\zeta}\right) + \log_2\left(2^{1-R} - 1\right)\right)$. A quick calculation shows that for $k = 6, n = 15$, $\beta^* = \frac{1.1413}{n} = 0.0761 > 0.0667$. A time invariant code $\mathbb{H}_{15,0.4} \in \mathbb{TZ}_{\frac{1}{2}}$ was randomly generated and used with the decoding algorithm in Section V. Fig 5 shows the plot of a sample path of the above process before and after closing the loop, the fact that the plant has been stabilized is clear.
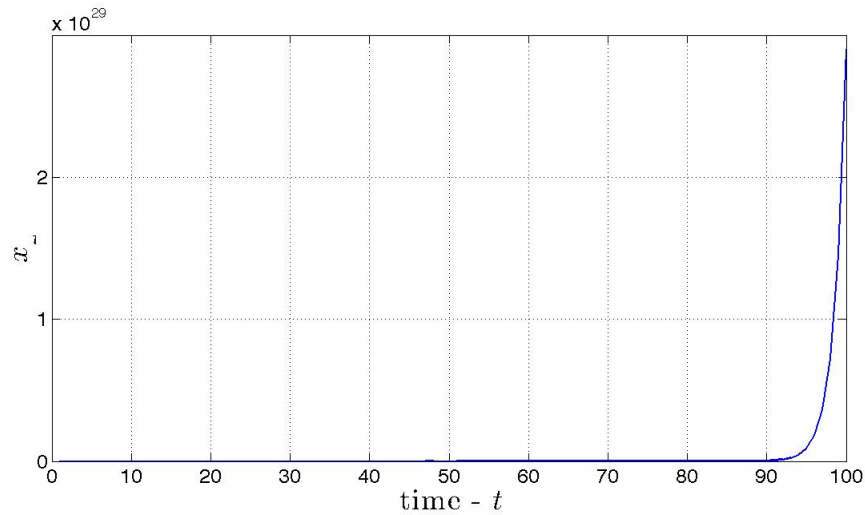
(a) Open loop trajectory
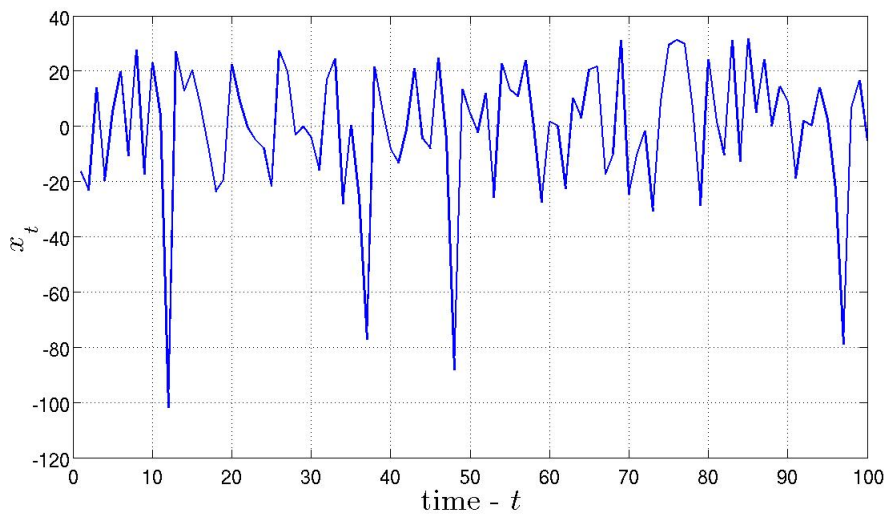
(b) Trajectory after closing the loop

Fig. 5

By easing up on the rate R, the control performance of the code ensemble improves. This is demonstrated in Fig 6. In the example considered above, one can decrease the rate by increasing δ, e.g., increasing
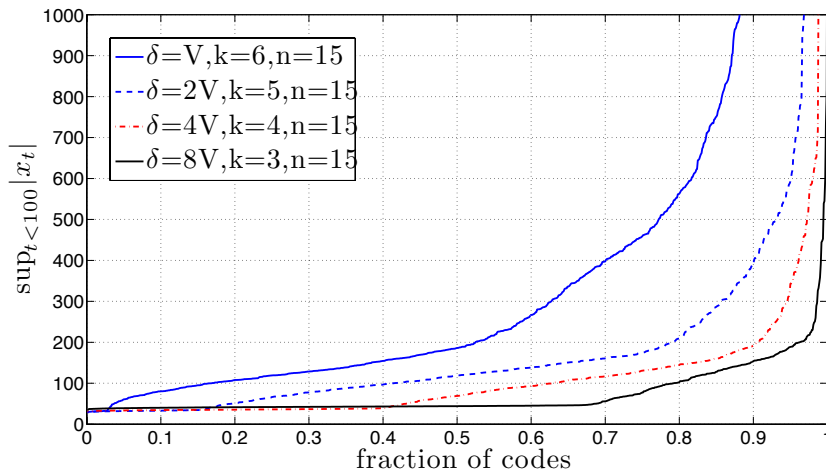
Fig. 6: The control performance of the code ensemble improves as the rate decreases

δ from V to $4V$, k drops from 6 to 4 bits and the resulting rate drops from 0.4 to 0.267. For each value of k from 3 to 6, 1000 time invariant codes were generated at random from $\mathbb{TZ}_{\frac{1}{2}}$. Each such code was used to control the process above over a time horizon of $T = 100$. The $y-$axis denotes $\sup_{t<100} |x_t|$ while the $x-$axis denotes the proportion of codes for which $\sup_{t<100} |x_t|$ is below a prescribed value. For example, with $k = 6, n = 15$, $\sup_{t<100} |x_t|$ was less than 200 for 50% of the codes while with $k = 3, n = 15$, this fraction increases to more than 95%. The $y-$axis has been capped at 1000 for clarity. This shows that one can trade-off utilization of communication resources and control performance. A rate $R = 0.267$ is much smaller than the capacity $1 - \log_2(1 + 0.3) = 0.6215$ but allows codes with much better control performance.

VII. CONCLUSION

We prove the existence of linear anytime codes with high probability. Our analysis considered binary alphabet, but is easily extendable. This is a significant step, since prior work only demonstrated the existence of such codes. For the BEC, we also propose an efficient decoding algorithm with constant average complexity per iteration, and for which the probability of having a complexity of KC^3 at some given iteration decays exponentially in C. Simulations validate the efficacy of the method. Constructing code families with efficient decoding for other channels, such as the BSC or AWGN remains an interesting open problem.

REFERENCES

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

[2] J. Baillieul and P.J. Antsaklis, "Control and communication challenges in networked real-time systems," *Proceedings of the IEEE*, vol. 95, no. 1, pp. 9 –28, Jan 2007.

[3] 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.

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

[5] 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.

[6] 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.

[7] Serdar Yuksel, "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.

[8] 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.

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

[10] 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.

APPENDIX

Lemma A.1: Let $V = \{0,1\}^m$ 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 a $\ell-$dimensional subspace of V, then

$$P(U) \leq \begin{cases} (1-p)^{m-\ell} & \text{if } p \leq 1/2 \\ p^{m-\ell} & \text{if } p \geq 1/2 \end{cases}$$

The proof is easy.

Remark 1 (Rank of a block lower traingular matrix): Let A be the following block lower triangular matrix

$$A = \begin{bmatrix} A_{11} & 0 & \dots & \dots \\ A_{21} & A_{22} & 0 & \dots \\ \vdots & \vdots & \ddots & \vdots \\ A_{t1} & A_{t2} & \dots & A_{tt} \end{bmatrix}$$

If each of $\{A_{jj}\}_{j=1}^{t}$ has full rank, then A also has full rank.

Proof of Theorem 4.2: 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 $\bar{n}t \times nt$ 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

$$
\begin{bmatrix}
H_1 & 0 & \cdots & \cdots \\
H_2 & H_1 & 0 & \cdots \\
\vdots & \vdots & \ddots & \vdots \\
H_d & H_{d-1} & \cdots & H_1
\end{bmatrix}
\begin{bmatrix}
c_{t-d+1} \\
c_{t-d+2} \\
\vdots \\
c_t
\end{bmatrix}
=
\begin{bmatrix}
0 \\
0 \\
\vdots \\
0
\end{bmatrix}
\tag{15}
$$

Note that (15) can be equivalently written as follows

$$
\underbrace{\begin{bmatrix}
C_{t-d+1} & 0 & \cdots & \cdots \\
C_{t-d+2} & C_{t-d+1} & 0 & \cdots \\
\vdots & \vdots & \ddots & \vdots \\
C_T & C_{t-1} & \cdots & C_{t-d+1}
\end{bmatrix}}_{\triangleq C}
\underbrace{\begin{bmatrix}
h_1 \\
h_2 \\
\vdots \\
h_d
\end{bmatrix}}_{\triangleq h}
=
\begin{bmatrix}
0 \\
0 \\
\vdots \\
0
\end{bmatrix}
\tag{16}
$$

where $h_i = \text{vec}(H_i^T)$, i.e., 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.

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

Since $c_{t-d+1} \neq 0$, C_{t-d+1} has full rank $\bar{n}$ and consequently C has full rank $d\bar{n}$. So, $P(\mathbb{H}_{n,R}^{t}c = 0) = P(Ch = 0)$. This is the probability that h lies in a $dn\bar{n} - d\bar{n}$ dimensional null space of C. This further implies that $h' = [h_2^T, \ldots, h_d^T]^T$ lies in a $(d-1)n\bar{n} - (d-1)\bar{n}$ dimensional subspace. Note that h' is an

$(d-1)n\bar{n}$ dimensional Bernoulli(p) vector and using Lemma A.1, we have

$$P(\mathbb{H}_{n,R}^t c = 0) \leq (1-p)^{\bar{n}(d-1)} \tag{17}$$

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

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

$$= \eta 2^{-nd((1-R)\log_2(1/(1-p))-H(\alpha))} \tag{18}$$

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

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

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

$$\leq \eta 2^{-nd(\theta w/nd - H(w/nd) + (1-R)\log_2(1/(1-p)))} \tag{19}$$

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

$$\theta^* = \max_{x \geq \alpha} \frac{H(x) - (1-R)}{x} \tag{20}$$

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 (19), we have

$$P\left(\exists\, \alpha nd \leq w \leq nd \;\ni\; N_{w,d} > 2^{\theta w}\right) \leq nd2^{-nd\delta} \tag{21}$$

for some $\delta' > 0$. For some fixed d_o large enough, applying a union bound over $d \geq d_o$ to (18) and (21), we get

$$P\left(\exists\, d \geq d_o \;\ni\; w_{\min,d} < \alpha nd \text{ or } N_{w,d} > 2^{\theta w}\right) \leq 2^{-\Omega(nd_o)} \tag{22}$$

Recall that for the above argument to be complete, we also require $\mathbb{H}_{n,R}^t$ to have full rank for each t. By Remark 1, this is guaranteed if H_1 has full rank. So, ∎

Proof of Theorem 4.1: The proof is by induction. Suppose $\mathbb{H}_{n,R,T-1}$ has $(\alpha, \theta, d_o)-$anytime distance. Construct $\mathbb{H}_{n,R,T}$ as follows.

$$
\mathbb{H}_{n,R,T} = \left[\begin{array}{c|ccc}
H_{11} & 0 & \cdots & \cdots \\
\hline
H_{21} & & & \\
\vdots & & \mathbb{H}_{n,R,T-1} & \\
H_{T1} & & &
\end{array} \right]
$$

where H_{11} is chosen to be a full rank matrix and the entries of $H_{j1} \in \{0,1\}^{\bar{n} \times n}$, $j \geq 2$, are drawn according to i.i.d Bernoulli($\frac{1}{2}$). We will show that if $\mathbb{H}_{n,R,T-1}$ has $(\alpha, \theta, d_o)-$anytime distance, then $\mathbb{H}_{n,R,T}$ will is also have $(\alpha, \theta, d_o)-$anytime distance with a probability $1 - 2^{-\Omega(nd_o)}$. Note that the probability is over the choice of $\{H_{j1}\}_{j=1}^{T}$. Let $\{w_{\min,d}^{t}, N_{w,d}^{t}\}_{d \geq d_o, t \leq T}$ be the weight distribution parameters associated to $\mathbb{H}_{n,R,T}$. Since $\mathbb{H}_{n,R,T-1}$ has $(\alpha, \theta, d_o)-$anytime distance, we have the following

$$
w_{\min,d}^{t} \geq \alpha n d, \ \forall \ d_o \leq d \leq t-1, \ t \geq d_o + 1
$$
$$
N_{w,d}^{t} \leq 2^{\theta w}, \ \forall \ w \geq \alpha n d, \ d_o \leq d \leq t-1, \ t \geq d_o + 1
$$

Towards proving that $\mathbb{H}_{n,R,T}$ has $(\alpha, \theta, d_o)-$anytime distance, it remains to show the following holds with a positive probability.

$$
\text{For } t \geq d_o, \ w_{\min,t}^{t} \geq \alpha n t, \ N_{w,t}^{t} \leq 2^{\theta w}, \ \forall \ w \geq \alpha n t \tag{23}
$$

Let $c \in \{0,1\}^{nt}$ such that $c_{\tau < t-d+1} = 0$ and $c_{t-d+1} \neq 0$, then it is easy to see that $P\left(\mathbb{H}_{n,R,T}^{t} c = 0 \right) = 2^{-n(d-1)}$. The rest of the analysis follows exactly along the lines of the proof of Theorem 4.2 starting from (17) with $p = \frac{1}{2}$. This gives the following result

$$
P\left(\mathbb{H}_{n,R,T} \text{ is bad} | \mathbb{H}_{n,R,T-1} \text{ is good} \right) =
$$
$$
P\left(\{w_{\min,d}^{t}, N_{w,d}^{t}\} \text{ do not satisfy (23)} \right) \leq 1 - 2^{-\Omega(nd_o)}
$$

In particular, there exists a choice of $\{H_{j1}\}_{j=1}^{T}$ such that $\mathbb{H}_{n,R,T}$ has $(\alpha, \theta, d_o)-$anytime distance, whenever $\mathbb{H}_{n,R,T-1}$ has $(\alpha, \theta, d_o)-$anytime distance. For the inductive argument to be complete, one needs to prove that there exists a $\mathbb{H}_{n,R,d_o}$ that has $(\alpha, \theta, d_o)-$anytime distance. This is already covered in the proof of the above inductive step. ∎

Proof of Theorem 4.4: The result can be obtained by a slight modification of the proof of Theorem 4.2. Consider an arbitrary decoding instant t, if the all zero codeword is transmitted is the transmitted codeword, $z = [z_1^T, \ldots, z_t^T]^T$ is the channel output and $v = [v_1, \ldots, v_t^T]^T$ is the bit flip pattern induced by the channel BSC(ϵ), $\epsilon < 1/2$, then $z = v$ and for any $c \in \mathcal{C}_{t,d}$, we have

$$
P_{t,d}^e \leq P \left(\bigcup_{c \in \mathcal{C}_{t,d}} 0 \text{ decoded as } c, \|v_{t-d+1:t}\| \leq nd(\epsilon + \delta) \right)
$$

$$
+ P\big(\|v_{t-d+1:t}\| > nd(\epsilon + \delta)\big)
$$

$$
\leq \sum_{w_{\min,d} \leq w \leq nd(\epsilon + \delta)} N_{w,d} \zeta^w + 2^{-ndKL(\epsilon + \delta \| \epsilon)} \tag{24}
$$

Since we know from Theorem 4.2 that $w_{\min,d}$ can scale as αnd for α upto $H^{-1}(1 - R)$, for any choice of R such that $H^{-1}(1 - R) > \epsilon$, there exists a $\delta > 0$ small enough such that $\epsilon + \delta < \alpha < H^{-1}(1 - R)$. So (24) gives $P_{d,t}^e \leq 2^{-ndKL(\epsilon + \delta \| \epsilon)}$. The proof for $\epsilon > 1/2$ follows by symmetry. ∎