Anytime reliable transmission of continuous information through digital noisy channels [*]

Giacomo Como[†], Fabio Fagnani[‡] and Sandro Zampieri[§]

January 4, 2009

Abstract

The problem of reliably transmitting a real-valued random vector through a digital noisy channel is relevant for the design of distributed estimation and control techniques over networked systems. One important example consists in the remote state estimation under communication constraints. In this case, the coding consists of an encoder –which maps the real vector into a sequence of channel inputs– and a decoder –which sequentially updates the estimation of the transmitted data as more and more channel outputs are observed. The encoder performs both source and channel coding of the data. Assuming that no channel feedback is available at the transmitter, this paper studies the rates of convergence to zero of the mean squared error. Two coding strategies are analyzed: the first one has exponential convergence rate but it is expensive in terms of encoder/decoder computational complexity, while the second one has a convenient computational complexity, but sub-exponential convergence rate. General bounds are obtained describing the convergence properties of these classes of methods.

1 Introduction

Reliable transmission of information among the nodes of a network is known to be a relevant problem in information engineering. It is indeed fundamental both when the network is designed for pure information transmission, as well as in scenarios in which the network is deputed to accomplish some specific tasks requiring information exchange. Important examples include: networks of processors performing parallel and distributed computation [2, 34] or load balancing [9, 10, 24]; wireless sensor networks, in which the final goal is estimation and decision making from distributed measurements [15, 17, 38, 11]; sensors/actuators networks, such as mobile multi-agent networks, in which the final goal is control [16, 25, 23, 26]. Distributed algorithms to accomplish synchronization, estimation or localization tasks necessarily need to exchange quantities among the agents which are often real valued. Assuming that transmission links are digital, a fundamental problem is thus to transmit a continuous quantity, namely a real number or, possibly, a vector, through a digital noisy channel up to a certain degree of precision.

This paper is concerned with the problem of efficiently transmitting a finite-dimensional Euclidean-space-valued state through a noisy digital channel. We shall focus on anytime coding algorithms, namely algorithms which can be stopped anytime while providing estimations of increasing quality. These algorithms are particularly suitable for applications in problems of distributed control.

As especially pointed out in a series of works by Sahai and Mitter [29, 30, 31], there is a specific feature distinguishing the problem of information transmission for control from the problem of pure information transmission. This is related to the different sensitivity to delay typically occurring in the two scenarios. Indeed, while the presence of sensible delays can often be tolerated in the communication performance evaluation, it typically has disastrous effects in control applications. Here, the important question is not only

---

where information is available, but also when. For this reason, while in the standard digital communication framework data are requested to be available at the receiver only at the end of the transmission (block coding), transmission systems for control applications need to be able to produce a reasonable partial information transmission also in case the process is stopped before the end. Consequently, it is desirable to design coding/decoding schemes which are able to provide an estimate whose precision increases with time.

On the other hand, the computational complexity of the transmission schemes is a central issue. In fact, sensors in distributed networks are usually very simple devices with limited computational abilities and severe energy constraints. Applicable transmission systems should be designed performing a number of operations which remains bounded in time. Hence, an analysis of the tradeoffs between performance and complexity of the transmission schemes is required.

In many problems of information transmission, there is the possibility to take advantage of the feedback information naturally available to the transmitter. Feedback can be helpful in enlarging the achievable capacity regions, improving the trade-off between performance and latency, as well as in reducing the computational complexity. In many cases, however, feedback information is incomplete, difficult to be used or, as for instance in the wireless network scenario, there are situations in which the transmitter needs to broadcast his information to many different receivers and hence feedback strategies to acknowledge the receipt of past transmissions could be unfeasible. For these reasons, in the present paper we shall restrict to the case in which there is no feedback information.

A fundamental characteristic of digital communication for control applications concerns the nature of information bits. In the traditional communication theory, bits are usually assumed to be equally valuable, and they are consequently given the same priority by the transmission-system designer. While such an assumption is typically justified by the source-channel separation principle, such a principle does not generally hold when delay is a primary concern. For instance, it is known that separate source-channel coding is suboptimal in terms of the joint source-channel error exponent [6, 7]. In fact, in problems of information transmission for control or estimation, different bits typically require significantly different treatment. This motivates the study of unequal error protection codes [22]: we refer to the recent work [4] for some information-theoretical aspects of unequal error protection and further references on the subject. While modern low-complexity codes [21, 27], based on random sparse graphical models and iterative decoding algorithms, are typically analyzed and designed for standard information transmission problems, one of the challenges posed by information transmission for control applications is to come up with design paradigms providing the required unequal error protection at low computational costs.

In this paper we propose two classes of coding strategies for the anytime transmission of real-valued random vectors through a digital noisy channel. In both cases the coding scheme consists of an encoder mapping the real vector into a sequence of bits and of a decoder sequentially refining the estimate of the vector as more and more channel outputs are observed. The first strategy is characterized by good performance in terms of the convergence of the mean squared error, but it is expensive in terms of encoder/decoder computational complexity. On the other hand, the second class of strategies has a convenient computational complexity, but worse convergence rate.

In order to keep the use of information-theoretical techniques at a minimum, we shall confine our exposition to the binary erasure channel (BEC), where a bit is either transmitted correctly or erased with some probability ϵ.[1] While this channel allows for an elementary treatment, it is of its own interest in many scenarios. For instance, it well models the situation of mobile agents which, depending on their current position, may or may not be in the range of transmission of the other agents, as well as a communication network -like the internet- in which information packets are either correctly received or lost in the transmission.

The rest of this paper is organized as follows. Sect.2 formally states the problem. In Sect.3 we briefly address the case of noiseless digital channels, for which our problem reduces to that of efficient vector quantization. In Sect.4, we find some information-theoretic limits of the coding schemes: an upper bound on the best error exponent achievable is presented in Sect.4.1, while, in Sect.4.2, random linear convolutional codes are shown to achieve exponential error rates at the cost of computational complexity growing quadratically in time. In Sect.5, trade-offs between performance and computational complexity are investigated. First, a simple linear-time encodable/decodable repetition scheme is analyzed in Sect.5.1. Then, the main result is presented in Sect.5.2, showing that finite-window coding schemes are able to achieve only sub-exponential error decays. Finally, some Monte Carlo simulations of finite-window coding schemes with linear complexity are reported in Sect.5.3.

---

[1]It will be pointed out how the results can be extended to general discrete memoryless channels.

2 Problem formulation

This section provides a formal description of the problem. Let $\mathcal{X} \subseteq \mathbb{R}^d$ be a non-empty subset where the random vector to be transmitted is known to take values, equipped with an a-priori probability density $f(x)$. We shall assume in the sequel that $\mathcal{X}$ is a bounded set, which, with no loss of generality, can be identified with $[0,1]^d$. The case when $\mathcal{X}$ is unbounded can be treated in an analogous way, but some more technicalities are needed. The communication channel is assumed to have binary input alphabet $\mathcal{Y} = \{0,1\}$ and a finite output alphabet $\mathcal{Z}$ which is assumed to contain $\mathcal{Y}$, namely we assume that $\mathcal{Y} \subseteq \mathcal{Z}$. We will consider in detail the binary erasure channel (BEC) in which $\mathcal{Z} = \{0,1,?\}$, where ? stays for the erasure event. The channel is described by two probability distributions on $\mathcal{Z}$, denoted by $p(z|0)$, and $p(z|1)$, where $z \in \mathcal{Z}$, and to be interpreted as the probability distribution of the output, assuming that the input has been equal to 0 or 1, respectively. In the case of the BEC we have that

$$p(?|0) = p(?|1) = \epsilon, \quad p(0|0) = p(1|1) = 1 - \epsilon, \quad p(1|0) = p(0|1) = 0.$$

We assume, for the sake of simplicity, that at every time instant t, we can transmit a bit through the channel, and, moreover, that the channel is memoryless, namely, the output values of repeated transmissions are independent among each other.

The coding scheme Our transmission scheme consists of an encoder

$$\mathcal{E} : \mathcal{X} \to \mathcal{Y}^{\mathbb{N}},$$

and of a decoder

$$\mathcal{D} : \mathcal{Z}^{\mathbb{N}} \to \mathcal{X}^{\mathbb{N}}.$$

The overall sequence of maps is described by the following scheme

$$\mathcal{X} \xrightarrow{\mathcal{E}} \mathcal{Y}^{\mathbb{N}} \xrightarrow{\text{Channel}} \mathcal{Z}^{\mathbb{N}} \xrightarrow{\mathcal{D}} \mathcal{X}^{\mathbb{N}}.$$

More precisely, the decoder is defined by a family of maps

$$\mathcal{D}_t : \mathcal{Z}^t \to \mathcal{X},$$

so that, for any $(z_s)_{s=1}^{\infty} \in \mathcal{Z}^{\mathbb{N}}$, the value at time t of $\mathcal{D}((z_s)_{s=1}^{\infty})$ is $\mathcal{D}_t((z_s)_{s=1}^t)$. In other words, if $\pi_t : \mathcal{Y}^{\mathbb{N}} \to \mathcal{Y}^t$ is the projection of a sequence in $\mathcal{Y}^{\mathbb{N}}$ into its first t symbols, then, for any $x \in \mathcal{X}$, the string $\pi_t(\mathcal{E}(x)) = (y_s)_{s=1}^t = (y_1, \ldots, y_t) \in \mathcal{Y}^t$ is transmitted along the channel and the output $(z_s)_{s=1}^t = (z_1, \ldots, z_t) \in \mathcal{Z}^t$ is then received by the decoder $\mathcal{D}_t$ which provides an estimate of x at time t

$$\hat{x}_t = \mathcal{D}_t((y_s)_{s=1}^t).$$

This is described by the following scheme

$$\mathcal{X} \xrightarrow{\mathcal{E}_t} \mathcal{Y}^t \xrightarrow{\text{Channel}} \mathcal{Z}^t \xrightarrow{\mathcal{D}_t} \mathcal{X}$$

$$x \longrightarrow (y_s)_{s=1}^t \longrightarrow (z_s)_{s=1}^t \longrightarrow \hat{x}_t$$

(1)

where $\mathcal{E}_t := \pi_t \circ \mathcal{E}$.

Performance evaluation In order to evaluate the performance of a scheme, we define the mean squared error (mean with respect to both the randomness of $x \in \mathcal{X}$ and with respect to the possible randomness of the communication channel) at time t by

$$\Delta_t := (\mathbb{E}||x - \hat{x}_t||^2)^{1/2}.$$

(2)

In this paper we want to understand how fast Δ_t decreases as t tends to infinity. In this paper we shall analyze different encoding and decoding strategies and we shall compare them by analyzing their performance in terms of convergence rate of Δ_t and their complexity of the encoding and decoding algorithms as functions

of t. All the coding strategies which will be analyzed in the present paper are characterized by mean squared error Δ_t converging to zero and such that[2]

$$\liminf_{t \to \infty} \left[-\frac{1}{t^\alpha} \log \Delta_t \right] \geq \beta \,, \tag{3}$$

for some constants $\beta > 0$ and $0 < \alpha \leq 1$. When (3) holds the coding strategy will be said to achieve a degree of convergence α and a rate of convergence β. When $\alpha = 1$ we shall simply say that we have an exponential convergence and that β is the exponential convergence rate. In the sequel, various strategies will be compared in terms of the parameters α and β that can be achieved, and such parameters will be related to the required computational complexity.

2.1 Application to state estimation under communication constraints

The problem illustrated in the previous paragraph is related to the state estimation problem under communication constraints (see [20, 19, 32, 33, 28]). Assume we are given a discrete time stochastic linear system

$$x(t+1) \quad = \quad Ax(t) + v(t) \qquad x(0) = x_0 \tag{4}$$

where $x_0 \in \mathbb{R}^n$ is a random vector with zero mean, $v(t) \in \mathbb{R}^n$ is a zero-mean white noise, $x(t) \in \mathbb{R}^n$ is the state sequence and $A \in \mathbb{R}^{n \times n}$.

Suppose that a remotely positioned receiver is required to estimate the state of the system, but it can receive information from it only through a binary erasure channel. We then need to design a family of encoders E_t and of decoders $\mathcal{D}_t$. At each time $t \geq 0$, the encoder E_t takes $x(0), \ldots, x(t)$ as input, and returns the symbol $y_t \in \{0, 1\}$, which is in turn fed as an input to the channel. The receiver observes the channel output symbols $z_0, \ldots, z_t$, from which the decoder $\mathcal{D}_t$ has to obtain an estimate $\hat{x}(t)$ of the current state. We distinguish two cases:

1. Assume that the variance of $v(t)$ is big with respect to the variance of $x(0)$ or that we are interested in the steady state performance. In this case it is the asymptotic value of $\mathbb{E}[||x(t) - \hat{x}(t)||^2]$ the most relevant parameter to be considered in designing the encoders E_t and the decoders $\mathcal{D}_t$.

2. Assume that the variance of $v(t)$ is small with respect to the variance of $x(0)$ and that we are interested in the transient behavior. In this case the prominent role is taken by the speed of convergence of $\mathbb{E}[||x(t) - \hat{x}(t)||^2]$ towards its asymptotic value and hence the role of the noise $v(t)$ is negligible. If this is the case, it makes sense to assume that $v(t) = 0$.

The results presented in this paper are relevant for the second scenario. In fact, since we can assume that $v(t) = 0$, the only source of uncertainty is due to the initial condition x_0 and so the encoder/decoder task reduces to obtain good estimates of x_0 at the receiver side. Indeed, in order to obtain a good estimate $\hat{x}(t)$ of $x(t)$, the decoder has to obtain the best possible estimate $\hat{x}(0|t)$ of the initial condition $x(0)$ from the received data $y_0, \ldots, y_t$, and then it can define

$$\hat{x}(t) := A^t \hat{x}(0|t) \,.$$

In this way we have

$$x(t) - \hat{x}(t) = A^t (x(0) - \hat{x}(0|t)) \,,$$

so that the problem reduces to finding the best way of coding $x(0)$ in such a way that expansion of A^t is well dominated by the contraction of $x(0) - \hat{x}(0|t)$.

3 Quantized encoding schemes

In this paper we shall propose and compare different coding strategies. All of them are based on a preliminary quantization of the real vector x into an infinite binary sequence. More precisely we shall first consider a map

$$\mathcal{S} : \mathcal{X} \to \mathcal{Y}^{\mathbb{N}} \,.$$

---

[2]Throughout the paper log denotes the logarithm in base 2, while ln denotes the natural logarithm.

This map is called a quantizer. There are many ways to design efficient quantizers. Below we shall propose a particularly simple and natural one and we shall stick to it in the following. In this paper, indeed, the focus is rather on the construction of the encoding scheme starting from the sequence $\mathcal{S}(x)$ which will be considered fully available at the transmitter.

The construction of $\mathcal{S}(x)$ works as follows. First, by considering the binary expansion of each component of x, we can find a sequence of binary vectors $b_1, b_2, \ldots \in \{0,1\}^d$ such that

$$x = \sum_{i=1}^{\infty} b_i 2^{-i} \tag{5}$$

Then we define[3] $\mathcal{S}(x) := (b_1^T, b_2^T, \ldots) \in \{0,1\}^{\mathbb{N}}$ namely as the infinite binary sequence obtained by concatenating the finite binary vectors b_i^T. We shall call the map $\mathcal{S}$ a dyadic quantizer and $\mathcal{S}(x)$ a dyadic expansion of x. It is clear that the formula (5) can be used also for defining the inverse $\mathcal{S}^{-1}$ of $\mathcal{S}$.

Let now $\mathcal{S}_t := \pi_t \circ \mathcal{S}$, where π_t is the truncation operator defined above. We define a right inverse $\mathcal{S}_t^{-1}$ of $\mathcal{S}_t$ as follows. Given a finite binary sequence $w = (w_1, \ldots, w_t) \in \{0,1\}^t$, first we expand it to a sequence $\bar{w} \in \{0,1\}^{\mathbb{N}}$ by adding infinitely many zeroes. Then from this sequence we define $\mathcal{S}_t^{-1}(w) := \mathcal{S}^{-1}(\bar{w})$.

Notice that, if we have two infinite sequences $w' = (w_1', w_2', \ldots)$ and $w'' = (w_1'', w_2'', \ldots)$ in $\{0,1\}^{\mathbb{N}}$ are such that $w_1' = w_1'', \ldots, w_t' = w_t''$, then

$$||\mathcal{S}^{-1}(w') - \mathcal{S}^{-1}(w'')|| \leq 2d^{1/2} 2^{-t/d} . \tag{6}$$

From this we can argue that

$$||x - \mathcal{S}_t^{-1} \circ \mathcal{S}_t(x)|| \leq 2d^{1/2} 2^{-t/d} ,$$

which implies that

$$(\mathbb{E}||x - \mathcal{S}_t^{-1} \circ \mathcal{S}_t(x)||^2)^{1/2} \leq 2d^{1/2} 2^{-t/d} .$$

Classical results in quantization theory show that the optimal quantizer has exponential rate of convergence $1/d$.

Theorem 1 (Theorem 6.2 pag.78 in [14]). *Suppose that $\mathbb{E}||x||^{2+\delta} < +\infty$ for some $\delta > 0$. Then, there exists $C > 0$ such that, for all $t \geq 0$ and $Q_t : \mathcal{X} \to \mathcal{X}$ with $|Q_t(\mathcal{X})| \leq 2^t$,*

$$(\mathbb{E}||x - Q_t(x)||^2)^{1/2} \geq C 2^{-t/d} \tag{7}$$

We now show how an encoder/decoder scheme can be built starting from the quantizer $\mathcal{S}$ and the family of inverses $\mathcal{S}_t^{-1}$.

Consider a sequence of integers $m_1, m_2, \ldots \in \mathbb{N}$ such that $m_{t-1} \leq m_t \leq t$ for all t and a family of maps

$$E_t : \mathcal{Y}^{m_t} \to \mathcal{Y}, \qquad \tilde{\mathcal{D}}_t : \mathcal{Z}^t \to \mathcal{Y}^{m_t} . \tag{8}$$

We can define the map $\tilde{\mathcal{E}} : \mathcal{Y}^{\mathbb{N}} \to \mathcal{Y}^{\mathbb{N}}$ by putting the value of $\tilde{\mathcal{E}}((w_s)_{s=1}^{\infty})$ at time t equal to $E_t(w_1, \ldots, w_{m_t})$. We also put $\tilde{\mathcal{E}}_t := \pi_t \circ \tilde{\mathcal{E}}$. Notice that, since $\tilde{\mathcal{E}}_t((w_s)_{s=1}^{\infty}))$ depends on $w_1, \ldots, w_{m_t}$ only, then $\tilde{\mathcal{E}}_t$ is actually a map from $\mathcal{Y}^{m_t}$ to $\mathcal{Y}^t$. Finally encoders and decoders are defined by $\mathcal{E}_t := \tilde{\mathcal{E}}_t \circ \mathcal{S}_{m_t}$ and $\mathcal{D}_t := \mathcal{S}_{m_t}^{-1} \circ \tilde{\mathcal{D}}_t$. The overall sequence of maps is described by the following scheme

$$\mathcal{X} \xrightarrow{\mathcal{S}_{m_t}} \mathcal{Y}^{m_t} \xrightarrow{\tilde{\mathcal{E}}_t} \mathcal{Y}^t \xrightarrow{\text{Channel}} \mathcal{Z}^t \xrightarrow{\tilde{\mathcal{D}}_t} \mathcal{Y}^{m_t} \xrightarrow{\mathcal{S}_{m_t}^{-1}} \mathcal{X}$$

$$\tag{9}$$

$$x \longrightarrow (w_s)_{s=1}^{m_t} \longrightarrow (y_s)_{s=1}^{t} \longrightarrow (z_s)_{s=1}^{t} \longrightarrow (\hat{w}_s(t))_{s=1}^{m_t} \longrightarrow \hat{x}_t$$

More specifically, in this scheme we first use a quantizer to transform x into a string of bits $(w_1, w_2, \ldots, w_{m_t})$ and then we use a block encoder. The received data are decoded by a block decoder providing an estimated version $(\hat{w}_1(t), \hat{w}_2(t), \ldots, \hat{w}_{m_t}(t))$ of $(w_1, w_2, \ldots, w_{m_t})$ (whose components in general depend on t) which is translated to an estimate $\hat{x}_t$ of x.

---

[3]Here and throughout the paper for a column vector $v \in \mathbb{R}^d$, v^T will denote its transposed row vector.

4   Information-theoretical limits

In the previous section it has been shown that quantizers attain exponential convergence ($\alpha = 1$) with rate $\beta = 1/d$ on noiseless binary channels. In this section it will be shown that exponential convergence is achievable on the BEC without feedback, at the cost of introducing non-trivial coding schemes ($\tilde{\mathcal{E}}_t, \tilde{\mathcal{D}}_t$).

First, in Sect.4.1 we shall prove a simple upper bound on the achievable exponential convergence rate β. Such a bound will show that, even with perfect feedback, no rates β larger than some $\overline{\beta}(\varepsilon, d)$ (see (13)) are achievable in the estimation of a d-dimensional random vector through a BEC with erasure probability ε. The quantity $\overline{\beta}(\varepsilon, d)$ will be shown to be strictly smaller than the normalized channel capacity.

Then, in Sect.4.2, it will be shown that exponential convergence rates β larger than or equal to some quantity $\underline{\beta}(\varepsilon, d)$ (see (29)) are indeed achievable on the BEC without feedback. The proposed schemes, based on random binary-linear convolutional codes, have computational complexity of the encoder quadratic in t, and decoder complexity $O(t^3)$.[4]

The results presented in Sect.4.2 constitute a refinement for the BEC of those proved in [31] for general memoryless discrete channels using non-linear convolutional codes. Indeed, the schemes proposed in [31] yield exponential convergence with encoder and decoder complexity growing exponentially in t. Also, we show that the use of binary-linear convolutional codes allows to achieve rates β larger than those achieved by non-linear convolutional codes for a whole range of values of ε and d.

Remark 1. Notice that a computational complexity growing linearly in t poses in principle no limitation on the reachable precision of $\hat{x}$, since it is natural to assume that a computing unit can perform a number of operations which grows linearly in time with a rate which depends on its computational power. If, on the other hand, we have a computational complexity growing more than linearly in t, then the algorithm will necessarily stop when the number of required operations will exceed the number of operations which the computing unit is able to do. This fact poses a limit on the reachable precision on $\hat{x}$.

4.1   A lower bound on the estimation error

From Sect.3 we know that it is not possible to obtain a convergence degree α greater than 1 and a convergence rate β larger than $1/d$. In this section we shall present a tighter upper bound on the convergence of Δ_t on the BEC with erasure probability ε.

Consider the general scheme (1). The error pattern associated to the output sequence $(z_t) \in \mathcal{Z}^{\mathbb{N}}$ is the sequence $(\xi_t) \in \{c, ?\}^{\mathbb{N}}$ componentwise defined by $\xi_t = c$ if $z_t \in \{0, 1\}$ and $\xi_t = c$ if $z_t =?$. Observe that the error pattern $(\xi_t)_{t \in \mathbb{N}}$ is a random variable independent from the random vector x, as well as from the encoder $\mathcal{E}$ and the decoder $\mathcal{D}$. This property will allow us to present for the BEC almost elementary proofs of results holding true also for more general channels. In particular, for $j \leq t$, let

$$\lambda_j^t := \sum_{j \leq s \leq t} \mathbb{1}_{\{\xi_s = c\}} \tag{10}$$

be the random variable describing the number of non-erased outputs observed between time j and t. Clearly,

$$\mathbb{P}(\lambda_j^t = l) = \binom{t - j + 1}{l} \epsilon^{t-l}(1 - \epsilon)^l, \qquad l = 0, \dots, t - j + 1. \tag{11}$$

The simple observation above allows to prove the following result.

Theorem 2. *Assume transmission over the BEC with erasure probability $\varepsilon \in [0, 1]$. Then, the estimation error of any coding scheme as in (1) satisfies*

$$\Delta_t \geq C \, 2^{-t\overline{\beta}(d,\varepsilon)}, \tag{12}$$

for all $t \geq 0$, where

$$\overline{\beta}(d, \varepsilon) := -\frac{1}{2} \log \left(\epsilon + (1 - \epsilon)2^{-2/d} \right) \tag{13}$$

and C is a constant depending only on the probability density of the random vector x.

---

[4]Here and throughout the paper, for two sequences $(a_t)_{t \in \mathbb{N}}$ and $(b_t)_{t \in \mathbb{N}}$, both the notations $a_t = O(b_t)$ and $b_t = \Theta(a_t)$ will mean that $a_t \leq Kb_t$ for some constant K, while $a_t = o(b_t)$ will mean that $\lim_t a_t/b_t = 0$.

Proof Conditioned on the infinite error pattern $(\xi_s)_{s\in\mathbb{N}}$, the channel reduces to a deterministic map, so that the composition of all the maps in (1) becomes a quantizer from $\mathcal{X}$ to itself with a range of cardinality $2^{\lambda_1^t}$. From this fact and from Theorem 1, we can deduce that

$$\mathbb{E}\left[||x-\hat{x}_t||^2\,|\,\lambda_1^t=l\right]\geq C^2 2^{-2l/d}\,.$$

It follows that

$$
\begin{aligned}
\mathbb{E}\left[||x-\hat{x}_t||^2\right] &= \textstyle\sum_{l=0}^t \mathbb{E}\left[||x-\hat{x}_t||^2\,|\,\lambda_1^t=l\right]\mathbb{P}(\lambda_1^t=l)\\
&\geq C^2\textstyle\sum_{l=0}^t 2^{-2l/d}\binom{t}{l}\epsilon^{t-l}(1-\epsilon)^l \qquad\qquad (14)\\
&= C^2\left(\epsilon+(1-\epsilon)2^{-2/d}\right)^t
\end{aligned}
$$

From this inequality the thesis follows.

Remark 2. The Shannon capacity of the BEC (measured in bits per channel use) equals $1-\varepsilon$, which is the average number of non-erased bits per channel use. It can be directly verified that[5]

$$\overline{\beta}(d,\varepsilon)<\frac{1}{d}(1-\varepsilon)\,,\qquad \forall\varepsilon\in]0,1[\,. \qquad\qquad (15)$$

The inequality (15) shows that the estimation error of any coding scheme after t uses of a digital noisy channel is exponentially larger than that of a quantizer whose image has cardinality t times the capacity of the original channel. In other words, (15) shows that the Shannon capacity is not sufficient in order to characterize the achievable exponential rates of the estimation error on a noisy channel. Indeed, a closer look at (14) reveals that the second summation is asymptotically dominated by the term corresponding to $l=l_t^*:=\lfloor t\frac{1-\varepsilon}{2^{1/d}\varepsilon+1-\varepsilon}\rfloor$, while the average number of unerased bits is given by $\mathbb{E}[\lambda_1^t]=(1-\varepsilon)t$. Hence, the exponential rate is dominated by atypical channel realizations, namely by the events $\{\lambda_1^t=l_t^*\}$ of probability exponentially vanishing in t. In fact, using finer information-theoretic arguments, Theorem 2 can be extended to general discrete memoryless channels, providing an upper bound $\overline{\beta}$ on the achievable error rate which can be written as a function of the sphere-packing exponent of the channel [13, pag.158]. Such a bound turns out to be strictly smaller than the Shannon capacity of the channel, whenever the sphere-packing exponent is finite at rates below capacity. The insufficiency of channel capacity for control/estimation problems with communication constraints and mean squared error distortion criteria[6] has already been observed in [29]. On the other hand, as we have seen in Sect.3, this is not the case for noiseless digital channels: in fact for such channels the Shannon capacity has been proven to be a sufficient measure in more general control/estimation problems [32].

It is not hard to see that (12) continues to hold true even if the encoder has access to noiseless (even non-causal) output feedback.[7] A fortiori, (12) holds in the case of partial or noisy feedback, which is the typical situation occurring in the network scenarios outlined in Sect.1. In the case of perfect causal output feedback, the bound (12) can be achieved using the coding scheme which repeats the transmission of the most significant bit of the dyadic expansion until it is correctly received. However, if the feedback is noisy, partial, or not available (as in the applications outlined in Sect.1), then the answer is not a priori clear. In Sect.4.2 we shall present schemes achieving exponential error rates at the cost of higher computational complexity, while in Sect.5.1 we shall propose some simple schemes which are not able to achieve exponential error rates, but have a lower computational complexity.

4.2 A coding scheme with exponential error rates

We shall now propose an encoding/decoding scheme achieving exponential convergence rates over the BEC, and requiring quadratic computational complexity at the encoder and cubic complexity at the decoder. We shall use random coding arguments employing anytime linear codes over the binary field $\mathbb{Z}_2$. These

---

[5]See also Fig.1.

[6]Or any other finite moment of the estimation error.

[7]In fact, it is tempting to conjecture that a tighter bound could possibly be proven for the exponent in the absence of feedback. It has been shown in [30] that the anytime reliability function of the BEC with feedback significantly exceeds the one without feedback. However, the proof of the upper bound on the exponent without feedback in [30] strongly relies on the causality of the coding scheme, an assumption which is not justified in our setting since the whole dyadic expansion of the random vector is assumed to be available at the transmitter at the beginning of the communication process.

arguments were first developed in the context of convolutional codes [36, 37, 12], and recently applied in the framework of anytime reliability [29, 31]. For the reader's convenience, and since those results have not appeared anywhere else in this form, we shall present self-contained proofs. The coding strategy we shall propose is very close in spirit to those in [29, Th.5.1] and [31, Th.5.1], the main difference being that we use linear convolutional codes instead of general random convolutional codes. Our choice has the double advantage of lowering the memory and complexity requirements for the encoder and the decoder (see Remark 4), and improving the achievable error rate for a significant range of values of ε (see Theorem 4 and Remark 3).

4.2.1 A random causal linear coding scheme

In this section we shall identify the binary set $\mathcal{Y} = \{0, 1\}$ with the binary field $\mathbb{Z}_2$ of the integers modulo 2.

Fix a rate $0 < R < 1$ and any t let $m_t := \lfloor Rt \rfloor$. Consider a random, doubly infinite, binary matrix $\phi \in \mathbb{Z}_2^{\mathbb{N} \times \mathbb{N}}$ distributed as follows: $\phi_{ij} = 0$ for all $j > Ri$ (namely for all $j \geq m_i + 1$), while $\{\phi_{ij}\}_{1 \leq j \leq Ri}$ is a family of mutually independent random with identical uniform distribution over $\mathbb{Z}_2$. As customary in random coding arguments, we shall assume the random matrix ϕ to be independent from the source vector x as well as from the channel, and known a priori both at the transmitting and receiving ends. Let us naturally identify the random matrix ϕ with the corresponding random $\mathbb{Z}_2$-linear operator $\tilde{\mathcal{E}} : \mathbb{Z}_2^{\mathbb{N}} \to \mathbb{Z}_2^{\mathbb{N}}$. Consider the truncated encoder

$$\tilde{\mathcal{E}}_t : \mathbb{Z}_2^{m_t} \to \mathbb{Z}_2^t, \qquad \tilde{\mathcal{E}}_t \left((w_s)_{s=1}^{m_t} \right) := \pi_t(\phi \mathbf{w}), \tag{16}$$

where $\mathbf{w} \in \mathbb{Z}_2^{\mathbb{N}}$ is such that $\pi_{m_t} \mathbf{w} = (w_s)_{s=1}^{m_t}$. Observe that the definition (16) is consistent, since it is independent on the choice of $\mathbf{w}$. Now, let $\mathcal{S} : \mathcal{X} \to \mathbb{Z}_2^{\mathbb{N}}$ be a dyadic quantizer defined as in (5), and define, as usual, the encoding scheme $\mathcal{E} : \mathcal{X} \to \mathbb{Z}_2^{\mathbb{N}}$ as the composition $\mathcal{E} = \tilde{\mathcal{E}} \circ \mathcal{S}$.

4.2.2 Maximum a posteriori decoding

Let $w_s, y_s, z_s, \hat{w}_s(t)$ be the sequences introduced in (9) and let ξ_s be the error pattern associated with z_s as defined in Sect.4.1. When an error pattern is fixed, the channel becomes a deterministic map. A maximum a posteriori decoder $\tilde{\mathcal{D}}_t$ for $\tilde{\mathcal{E}}_t$ is any map such that, for any fixed error pattern,

$$\text{Channel} \left[\tilde{\mathcal{E}}_t((\hat{w}_s(t))_{s=1}^{m_t}) \right] = (z_s)_{s=1}^t = \text{Channel} \left[\tilde{\mathcal{E}}_t((w_s)_{s=1}^{m_t}) \right]. \tag{17}$$

In other words, a maximum a posteriori decoder produces an estimate which is one of the possible encoder inputs which are mapped by the encoder and the channel into $(z_s)_{s=1}^t$. If we let $(\hat{y}_s)_{s=1}^t := \tilde{\mathcal{E}}_t((\hat{w}_s(t))_{s=1}^{m_t})$, then condition (18) is equivalent to impose that $\hat{y}_s = y_s$ for all s such that $z_s \neq ?$.

In order to express condition (17) more formally, we need to introduce for any fixed error pattern the set

$$\Xi_t := \{s \ : \ 1 \leq s \leq t : z_s \neq ?\} = \{s \ : \ 1 \leq s \leq t : \xi_s \neq ?\},$$

which is the set of non-erased positions up to time t, and the canonical projection $\pi_{\Xi_t} : \mathbb{Z}_2^t \to \mathbb{Z}_2^{\Xi_t}$. Condition (17) is equivalent to the following

$$\pi_{\Xi_t} \left[\tilde{\mathcal{E}}_t \tilde{\mathcal{D}}_t((z_s)_{s=1}^t) \right] = \pi_{\Xi_t} \left[(z_s)_{s=1}^t \right]. \tag{18}$$

Finally, the overall decoder is defined as the composition $\mathcal{D}_t := \mathcal{S}_{m_t}^{-1} \circ \tilde{\mathcal{D}}_t$.

4.2.3 Performance analysis

Assume now that $\tilde{\mathcal{D}}_t$ is a maximum a posteriori decoder as in Sect.4.2.2. Then, the decoded block $(\hat{w}_s(t))_{s=1}^{m_t} = \tilde{\mathcal{D}}_t((z_s)_{s=1}^t) \in \mathbb{Z}_2^{m_t}$ is uniquely defined, and correct, whenever the linear map $\pi_{\Xi_t}\tilde{\mathcal{E}}_t : \mathbb{Z}_2^{m_t} \to \mathbb{Z}_2^{\Xi_t}$ is injective. However, our analysis requires more detailed information regarding the location of the uncorrectly decoded information bits when injectivity is lost. To this end, let $\{\delta_1, \delta_2, \ldots, \delta_{m_t}\}$ be the canonical basis of $\mathbb{Z}_2^{m_t}$, and, for $0 \leq j \leq m_t$, consider the subspace[8]

$$K_j := \text{span}(\delta_{j+1}, \ldots, \delta_{m_t}) \subseteq \mathbb{Z}_2^{m_t}.$$

---

[8]We shall use the standard convention $\text{span}(\emptyset) := \{0\}$.

Define the events

$$A_j := \{\ker(\pi_{\Xi_t}\tilde{\mathcal{E}}_t) \subseteq K_j\}, \qquad 0 \le j \le m_t, \tag{19}$$

$$B_j := A_{j-1} \setminus A_j, \qquad 1 \le j \le m_t. \tag{20}$$

Observe that $A_j \subseteq A_{j-1}$, and that A_0 coincides with the whole sample space Ω. Hence, for every $t \in \mathbb{N}$, the sample space admits the partition

$$\Omega = \bigcup_{1 \le j \le m_t} B_j \bigcup A_{m_t}. \tag{21}$$

Notice now that, from (18) we can deduce that

$$\pi_{\Xi_t}\tilde{\mathcal{E}}_t((\hat{w}_s(t))_{s=1}^{m_t}) = \pi_{\Xi_t}\tilde{\mathcal{E}}_t((w_s)_{s=1}^{m_t})$$

and so $(w_s - \hat{w}_s(t))_{s=1}^{m_t} \in \ker \pi_{\Xi_t}\tilde{\mathcal{E}}_t$. Therefore, if A_j holds true, then $(\hat{w}_s(t))_{s=1}^j = (w_s)_{s=1}^j$, i.e. the first j bits of the quantization of x are correctly decoded. Hence, we immediately get from (6) that, if A_j holds true, then

$$||\hat{x}_t - x||^2 \le 4d2^{-2j/d}, \qquad 0 \le j \le m_t. \tag{22}$$

The following result characterizes the average mean squared error of the random coding scheme $(\mathcal{E}, \mathcal{D})$ over the BEC. Here the average has to be considered with respect to the randomness of the vector x, the channel, as well as the matrix ϕ. For $\varepsilon \in [0, 1]$ and $d \in \mathbb{N}$, define

$$\underline{\beta}'(d, \varepsilon, R) := \min\{\tfrac{1}{d}R, \tfrac{1}{2} \min_{0 \le \eta \le 1} D(\eta||1-\varepsilon) + \lfloor \eta - R \rfloor_+\}, \tag{23}$$

where $D(x||y) := x \log \frac{x}{y} + (1-x) \log \frac{1-x}{1-y}$ denotes the binary Kullback-Leiber distance[9] and where $\lfloor x \rfloor_+ := \max\{0, x\}$.

Theorem 3. *Assume transmission over the BEC. Then, for all $0 < R < 1$, the average estimation error of the above-described random coding scheme satisfies*

$$\mathbb{E}[||\hat{x}_t - x||^2] \le Ct2^{-2t\underline{\beta}'(d,\varepsilon,R)} \tag{24}$$

for all $t \in \mathbb{N}$. where $C > 0$ is a constant depending only on d, R and ϵ.

Proof Using (21) and (22), we obtain

$$\begin{aligned}
\mathbb{E}\left[||\hat{x}_t - x||^2\right] &= \sum_{j=1}^{m_t} \mathbb{E}\left[||\hat{x}_t - x||^2 \,|\, B_j\right] \mathbb{P}(B_j) + \mathbb{E}\left[||\hat{x}_t - x||^2 \,|\, A_{m_t}\right] \mathbb{P}(A_{m_t}) \\
&\le \sum_{j=1}^{m_t} \mathbb{P}(B_j) 4d2^{-2(j-1)/d} + 4d2^{-2m_t/d}.
\end{aligned} \tag{25}$$

In order to estimate $\mathbb{P}(B_j)$, first we claim that the event B_j implies that the column $\pi_{\Xi_t}\tilde{\mathcal{E}}_t\delta_j$ belongs to the subspace $\pi_{\Xi_t}\tilde{\mathcal{E}}_t K_j$, namely

$$B_j \subseteq \{\pi_{\Xi_t}\tilde{\mathcal{E}}_t\delta_j \in \pi_{\Xi_t}\tilde{\mathcal{E}}_t K_j\}$$

Indeed, if $\pi_{\Xi_t}\tilde{\mathcal{E}}_t v = 0$ for some $v \in \mathbb{Z}_2^{m_t}$, then A_{j-1} implies that $v_i = 0$ for all $i < j$, while $\overline{A_j}$[10] implies that $v_j \ne 0$.

Fix now an error pattern ξ_s and let $\lambda_{\lceil j/R \rceil}^t$ be defined in (10). Observe that $\tilde{\mathcal{E}}_t\delta_j$ is a random variable uniformly distributed over the subspace $H_j := \operatorname{span}(\delta_{\lceil j/R \rceil}, \ldots, \delta_t) \subseteq \mathbb{Z}_2^t$, and independent from the error pattern. It follows that $\pi_{\Xi_t}\tilde{\mathcal{E}}_t\delta_j$ takes any value in $\pi_{\Xi_t} H_j$ with probability $2^{-\lambda_{\lceil j/R \rceil}^t}$. Since $|\pi_{\Xi_t}\tilde{\mathcal{E}}_t K_j| \le |K_j| = 2^{m_t - j}$, we have that, for every $k = 0, \ldots, t - \lceil j/R \rceil + 1$,

$$\mathbb{P}(B_j | \lambda_{\lceil j/R \rceil}^t = k) \le \mathbb{P}(\pi_{\Xi_t}\tilde{\mathcal{E}}_t\delta_j \in \pi_{\Xi_t}\tilde{\mathcal{E}}_t K_j | \lambda_{\lceil j/R \rceil}^t = k) \le \min\{1, |K_j|2^{-k}\} = 2^{-\lfloor j+k-m_t \rfloor_+}.$$

---

[9]With the standard convention $0 \log 0 = 0$.
[10]For an event A, $\overline{A}$ denotes its complement.

From (11) it follows that

$$
\begin{aligned}
\mathbb{P}(B_j) &= \sum_{k=0}^{t-\lceil j/R\rceil+1} P(B_j|\lambda_{\lceil j/R\rceil}^t = k)\mathbb{P}(\lambda_{\lceil j/R\rceil}^t = k) \\
&\leq \sum_{k=0}^{t-\lceil j/R\rceil+1} 2^{-\lfloor j+k-m_t\rfloor_+}\binom{t-\lceil j/R\rceil+1}{k}\varepsilon^{t-\lceil j/R\rceil+1-k}(1-\varepsilon)^k \\
&\leq \sum_{k=0}^{t-\lceil j/R\rceil+1} 2^{-\lfloor j+k-m_t\rfloor_+}2^{-(t-\lceil j/R\rceil+1)D(\frac{k}{t-\lceil j/R\rceil+1}||1-\varepsilon)} \\
&\leq t2^{-(t-j/R)\min_{0\leq\eta\leq 1} D(\eta||1-\varepsilon)+\lfloor\eta-R\rfloor_+\}}
\end{aligned}
\tag{26}
$$

where the second inequality follows from standard estimations of the binomial coefficient (see e.g. [8]). Finally, (24) follows by substituting (26) into (25). ∎

Standard probabilistic arguments allow to prove the following corollary of Theorem 3, characterizing the exponential error rate of a typical realization of the random coding scheme $(\mathcal{E}, \mathcal{D})$. Observe that the mean square error of the coding scheme is given by

$$
\left(\mathbb{E}\left[||\hat{x}_t - x||^2|\phi\right]\right)^{1/2},
$$

which is a function of ϕ, and hence it is itself a random variable.

Corollary 1. *Assume transmission over the BEC with erasure probability ε. Then, for all $0 < R < 1$,*

$$
\liminf_t \left[-\frac{1}{t}\log\mathbb{E}[||x - \hat{x}_t||^2|\phi]\right] \geq 2\underline{\beta}'(d,\varepsilon,R), \tag{27}
$$

with probability one.

Proof Fix some $\eta > 0$ and consider the events

$$
A_{\eta,n} := \left\{\mathbb{E}[||x - \hat{x}_t||^2|\phi] \geq 2^{-2t(\underline{\beta}'(d,\varepsilon,R)-\eta)}\right\}
$$

for $n \in \mathbb{N}$. By applying Markov's inequality and Theorem 3, we get

$$
\mathbb{P}[A_{\eta,n}] \leq 2^{2t(\underline{\beta}'(d,\varepsilon,R)-\eta)}\mathbb{E}[||x - \hat{x}_t||^2] \leq K2^{-2t\eta},
$$

so that the series $\sum_{t\in\mathbb{N}}\mathbb{P}[A_{\eta,n}]$ is convergent and the Borel-Cantelli lemma implies that, with probability one, $A_{\eta,t}$ occurs only for finitely many values of $t \in \mathbb{N}$. Therefore, with probability one,

$$
\liminf_t \left[-\frac{1}{t}\log\mathbb{E}[||x - \hat{x}_t||^2|\phi]\right] \geq 2(\underline{\beta}'(d,\varepsilon,R) - \eta),
$$

and (27) follows by the arbitrariness of $\eta > 0$. ∎

It is possible to derive another lower bound on the typical-case exponential error rate achieved by the random scheme $(\mathcal{E}, \mathcal{D})$, which turns out to be tighter than that provided by Corollary 1 for certain values of R and ε. For every $0 \leq R \leq 1$ define[11]

$$
\gamma(R) := \min\{x \in [0,1] : \mathrm{H}(x) \geq 1 - R\},
$$

$$
\underline{\beta}''(d,\varepsilon,R) := \min\left\{\frac{1}{d}R, \ \frac{1}{2}\min_{\gamma(R)\leq\eta\leq 1}\{\mathrm{H}(\eta) - 1 + R - \eta\log\varepsilon\}\right\}.
$$

The following result is proved in Appendix A.

Theorem 4. *Assume transmission over the BEC with erasure probability ε. Then, for all $0 < R < 1$,*

$$
\liminf_t \left[-\frac{1}{t}\log\mathbb{E}[||x - \hat{x}_t||^2|\phi]\right] \geq 2\underline{\beta}''(d,\varepsilon,R), \tag{28}
$$

with probability one.

---

[11]Throughout, for $x \in [0,1]$, we shall use the notation $\mathrm{H}(x) := -x\log x - (1-x)\log(1-x)$ for the binary entropy of x with the standard convention $0\log 0 = 1$.
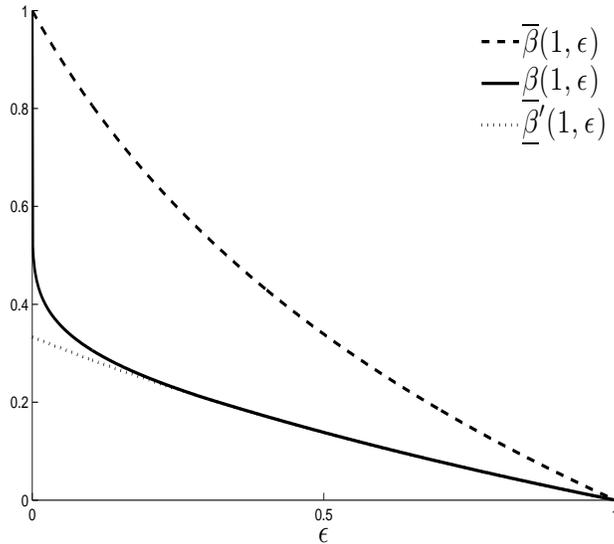
Figure 1: Upper and lower bounds to the achievable estimation error exponent achievable on the BEC for $d = 1$.

It then follows from Corollary 1 and Theorem 4 that the exponent

$$\underline{\beta}(d,\varepsilon) := \max_{0 \leq R \leq 1} \max\{\underline{\beta}'(d,\varepsilon,R), \underline{\beta}''(d,\varepsilon,R)\}, \tag{29}$$

is achievable by random causal linear codes. In Fig.1 the upper and lower bounds to the error exponent, i.e. $\overline{\beta}(d,\varepsilon)$ and $\underline{\beta}(d,\varepsilon)$, are plotted as functions of the erasure probability ε, in the case $d = 1$.

Remark 3. It is not difficult to see that

$$\lim_{\varepsilon \downarrow 0} \max_{0 \leq R \leq 1} \{\underline{\beta}'(d,\varepsilon,R)\} = \frac{1}{d+2}, \qquad \lim_{\varepsilon \downarrow 0} \max_{0 \leq R \leq 1} \{\underline{\beta}''(d,\varepsilon,R)\} = \frac{1}{d}.$$

Hence, Theorem 4 becomes particularly relevant for small erasure probabilities, showing that the noiseless error exponent $1/d$ (see Sect.3) is recovered in the limit of vanishing noise: this does not follow from the average-code analysis of Theorem 2. Using arguments as in [37], Theorem 4 for random linear convolutional codes can be extended to the class of discrete memoryless channels which are symmetric with respect to the action of the additive group of some finite field, showing the achievability of the exponential error rate $\min\left\{\frac{1}{d}R, \frac{1}{2}E_x(R)\right\}$, where $E_x(R)$ is the expurgated exponent of the channel [13].

4.3   Computational complexity of the scheme

Observe that the number n_t of binary operations required in order to compute the channel input $y_t = \tilde{\mathcal{E}}_t((w_s)_{s=1}^{m_t})$, equals the number of non-zero entries of the t-th row of the infinite random matrix ϕ. By the way ϕ has been defined, n_t is a binomial random variable of parameters m_t and $1/2$. Hence, the number of binary operations required by the encoder up to time t, $\chi_t := \sum_{s \leq t} n_s$, has binomial distribution of parameters $\frac{1}{2}m_t(m_t+1)$ and $1/2$. Therefore, the worst-case encoding complexity (worst case with respect to the realization of ϕ) grows like $\frac{1}{2}R^2t^2$, while the strong law of large numbers implies that the typical encoder complexity χ_t is such that $\chi_t/\frac{1}{4}R^2t^2$ converges to 1 with probability one. Thus, the encoder complexity (both worst-case and typical-case) is quadratic in t. Further, observe that the memory requirements of the encoder are quadratic in t for it is necessary to store $m_t t$ binary values in order to memorize the finite truncation $\mathcal{E}_t$ of the encoder $\mathcal{E}$.

In order to evaluate the decoder's computational complexity, observe that $\tilde{\mathcal{D}}_t$ is required to solve the $\mathbb{Z}_2$-linear system

$$\pi_{\Xi_t}\tilde{\mathcal{E}}_t((w_s)_{s=1}^{m_t}) = \pi_{\Xi_t}(z_s)_{s=1}^t. \tag{30}$$

11

at each time step t. This can be performed using Gaussian elimination techniques in order to reduce the matrix $\pi_{\Xi_t}\tilde{\mathcal{E}}_t$ to a lower-diagonal form. Notice that a sequential implementation is possible, i.e. the part of $\pi_{\Xi_t}\tilde{\mathcal{E}}_t$ which has been reduced in lower triangular form at time t does not require to be further processed in future times $s > t$. Since Gaussian elimination techniques require $O(t^3)$ operations, we can conclude that the decoder complexity is at most $O(t^3)$. On the other hand, it might be possible to find algorithms for solving a linear system like (30) with number of operations $o(t^3)$: see [35, pagg.247-248] for the analogous problem for linear systems over the reals. However, the system (30) cannot be solved using fewer operations than those required to verify that a given string $v \in \mathbb{Z}_2^{m_t}$ is a solution. Using arguments similar to those outlined above, it is possible to show that, with probability one, this requires $\Theta(t^2)$ binary operations. In summary, the complexity of maximum a posteriori decoding of linear convolutional codes on the BEC is at most $O(t^3)$ and at least $\Theta(t^2)$.

Remark 4. It is possible to extend Theorem 3 to arbitrary discrete memoryless channels, using a random coset approach possibly followed by a quantization as in [13, pagg.206-209] showing that the error rate $\underline{\beta}'(d,\varepsilon,R) := \min\left\{\frac{1}{d}R, \frac{1}{2}E_r(R)\right\}$ is achievable, where $E_r(R)$ is the random coding exponent of the channel [13]. On arbitrary discrete memoryless channels, linear (or coset) convolutional codes maintain linear encoding complexity, but their maximum a posteriori decoding is known to be an NP-hard problem [1].

The error rate $\underline{\beta}'(d,\varepsilon,R) := \min\left\{\frac{1}{d}R, \frac{1}{2}E_r(R)\right\}$ can be shown to be achievable, on general discrete memoryless channels, by using random non-linear convolutional codes as in [12, 29, 31]. However, observe that non-linear convolutional codes require exponential memory for the encoder, while their maximum a posteriori decoding is also an NP-hard problem. Moreover, at our knowledge, no result analogous to Theorem 4 is known to hold for non-linear random convolutional codes.

5 Low-complexity coding schemes

In this section, tradeoffs between computational complexity and performance of the coding schemes are investigated. First, in Sect.5.1, a simple linear-time encodable/decodable scheme is analyzed, showing that the estimation error converges to zero sub-exponentially fast with degree $\alpha = 1/2$. Then, in Sect.5.2, lower bounds on the estimation error are obtained: it is shown that encoding schemes with finite memory (finite-state automata), have estimation error bounded away from zero, while finite-window linear-time encodable encoders cannot achieve a convergence degree larger than $1/2$. Finally, in Sect.5.3, Monte Carlo simulations of finite-window coding schemes with iterative decoding are presented, showing that, while not improving the convergence degree $1/2$, they can provide better convergence rates.

5.1 A repetition coding scheme

We shall propose a simple repetition-coded scheme characterized by encoding and decoding complexity growing linearly in t. It will be shown that the convergence degree achievable in this case is $\alpha = 1/2$, with convergence rate $\beta = \sqrt{\frac{1}{d}\log\varepsilon^{-1}}$.

Let $\mathcal{S}_t : [0,1]^d \to \{0,1\}^t$ be the truncation of the dyadic quantizer $\mathcal{S} : [0,1]^d \to \{0,1\}^{\mathbb{N}}$ introduced in Sect.3, and let $\mathcal{S}_t^{-1} : \{0,1\}^t \to [0,1]^d$ be one of its right inverses. If a coding scheme with $\mathcal{E} = \mathcal{S}$ were simply used, namely if we send through the channel the bits directly coming from the quantizer and decode the erasures in an arbitrary way, then the estimation error Δ_t would not converge to 0 as $t \to \infty$. Indeed, with probability ϵ the first bit of $\mathcal{S}(x)$ would be lost with no possibility of recovering it. As we have already seen in Sect.4, it is necessary to introduce redundancy in order to cope with channel erasures. The simplest way to do that consists in using repetition schemes. Of course, since the different bits of the binary expansion $\mathcal{S}(x)$ require different levels of protection, they need to be repeated with a frequency monotonically decreasing in their significance.

Fix a positive real number q and the sequence of positive integers

$$\tau_0 = 0, \qquad \tau_k = \lceil q \rceil + \lceil 2q \rceil + \cdots + \lceil kq \rceil, \quad \forall\, k > 0.$$

Consider the encoder $\tilde{\mathcal{E}} : \{0,1\}^{\mathbb{N}} \to \{0,1\}^{\mathbb{N}}$ informally described by

$$\tilde{\mathcal{E}}((w_s)_{s=1}^{\infty}) = (w_1, w_2, \ldots, w_{\lceil q \rceil}, w_1, w_2, \ldots, w_{\lceil 2q \rceil}, w_1, w_2, \ldots, w_{\lceil 3q \rceil}, \ldots\ldots).$$

More precisely, notice that, for any $t \in \mathbb{N}$, there exist unique $m \in \mathbb{N}$ and $j \in \{1, 2, \dots, \lceil mq \rceil\}$ such that $t = \tau_{m-1} + j$. Denote these numbers by $m(t)$ and $j(t)$. Then,

$$\tilde{\mathcal{E}}((w_s)_{s \in \mathbb{N}}) := (w_{j(t)})_{t \in \mathbb{N}} . \tag{31}$$

Notice that this encoder fits in the scheme (9) by taking $m_t = \lceil qm(t) \rceil$. We construct the decoders $\tilde{\mathcal{D}}_t : \{0, 1, ?\}^t \to \{0, 1\}^{m_t}$ as follows. If $(\hat{w}_j(t))_{j=1}^{m_t} = \tilde{\mathcal{D}}_t((z_s)_{s=1}^t)$, then

$$\hat{w}_j(t) = \begin{cases} z_s & \text{if } \exists s \leq t \text{ such that } j(s) = j \text{ and } z_s \neq ? \\ ? & \text{otherwise} . \end{cases}$$

In the computation of $\hat{x}_t$ the symbols ? can be transformed arbitrarily to 0 or to 1. Notice that this decoding scheme has complexity growing linearly in t. Indeed, it admits the following natural recursive implementation. Assume that $(\hat{w}_j(t))_{j=1}^{m_t}$ has already be computed and that we receive z_{t+1}. Then, we compute $(\hat{w}_j(t+1))_{j=1}^{m_{t+1}}$ as

$$\hat{w}_j(t+1) = \begin{cases} z_{t+1} & \text{if } j = j(t+1) \text{ and } z_{t+1} \neq ? \\ \hat{w}_j(t) & \text{otherwise} \end{cases} . \tag{32}$$

Proposition 1. *Consider the repetition coding scheme defined by (32) and (32) on the BEC with erasure probability ε and assume that $q > \frac{d \log \epsilon^{-1}}{2}$. Then the mean squared error of satisfies*

$$\Delta_t \leq C 2^{-\log \epsilon^{-1} \sqrt{\frac{t}{2q}}} . \tag{33}$$

where $C > 0$ is a constant depending only on q, ε and d.

Proof Let us fix some $t \in \mathbb{N}$. Define $v_j := |\{1 \leq \tau \leq t \mid j(\tau) = j\}|$, and observe that $\mathbb{P}(\hat{w}_j(t) \neq w_j) = \epsilon^{v_j}$. Introduce the following event

$$A_j = \{\hat{w}_1(t) = w_1, \dots, \hat{w}_j(t) = w_j, \hat{w}_{j+1}(t) \neq w_{j+1}\} .$$

for $j = 0, 1, \dots, m_t$. Notice that these events are disjoint and $\mathbb{P}\left(\bigcup_{j=0}^{\lceil qm \rceil} A_j\right) = 1$. Moreover, observe that, for all $j = 0, 1, \dots, \lceil qm \rceil$,

$$\mathbb{P}(A_j^t) = \prod_{i=1}^j \mathbb{P}(\hat{w}_i(t) = w_j) \mathbb{P}(\hat{w}_{j+1}(t) \neq w_{j+1}) = \prod_{i=1}^j (1 - \epsilon^{v_i^t}) \epsilon^{v_{j+1}^t} \leq \epsilon^{v_{j+1}^t} . \tag{34}$$

Notice that, under the constraints posed by the event A_j^t we have that the first j bits of $S(x)$ and of $S(\hat{x}_t)$ coincide. Hence, by (6), we have that

$$\mathbb{E}[||x - \hat{x}_t||^2 \mid A_j^t] \leq 4d 2^{-2j/d} .$$

From this it follows that

$$\Delta_t^2 = \sum_{j=0}^{\lceil qm \rceil} \mathbb{E}\left[||x - \hat{x}_t||^2 \mid A_j\right] \mathbb{P}(A_j) \leq 4d \sum_{j=0}^{\lceil qm \rceil} 2^{-2j/d} \epsilon^{v_{j+1}} . \tag{35}$$

We need now to estimate the value of v_j. For simplicity we assume that $t = \tau_m$. In this case we have that $v_j = m + 1 - \min\{h \mid \lceil qh \rceil \geq j\}$. Observe now that, from the fact that for any positive real x we have that $j \leq \lceil x \rceil$ if and only if $j < x + 1$, we can argue that

$$\min\{h \mid \lceil qh \rceil \geq j\} = \min\{h \mid qh + 1 > j\} = \min\{h \mid h > (j-1)/q\} = \left\lfloor \tfrac{1}{q}(j-1) \right\rfloor + 1 .$$

This implies that, for $j = 0, 1, \dots, \lceil qm \rceil$, we have that $v_j = m - \left\lfloor \frac{j-1}{q} \right\rfloor$. Applying this argument to (35) and considering that for $j = \lceil qm \rceil + 1$ we have that $v_j = 0$, we obtain

$$\Delta_t^2 \leq 4d \left[\sum_{j=0}^{\lceil qm \rceil - 1} 2^{-2j/d} \epsilon^{m - \lfloor j/q \rfloor} + 2^{-2 \lceil qm \rceil / d} \right] \leq 4d \left[\sum_{j=0}^{\lceil qm \rceil - 1} 2^{-2j/d} \epsilon^{m - j/q} + 2^{-2qm/d} \right]$$

13

Now we take any $q > \frac{d \log \epsilon^{-1}}{2}$. Then we have that $\epsilon^{-1/q} 2^{-2/d} < 1$ and so

$$\Delta_t^2 \leq 4d \left[\frac{1}{1 - \epsilon^{-1/q} 2^{-2/d}} + 1 \right] \epsilon^m$$

Observe finally that

$$t = \tau_m = \sum_{j=1}^{m} \lceil qj \rceil \leq \sum_{j=1}^{m} (qj + 1) = \frac{q}{2} m^2 + \frac{q+2}{2} m$$

This implies that $m \geq \sqrt{\frac{2t}{q}} - \frac{q+2}{2q}$, so that the claim follows. ∎

Notice that the constant C in the previous proposition tends to infinity as q tends $\frac{d \log \epsilon^{-1}}{2}$. Theorem 1 implies that repetition coding schemes allow to achieve

$$\alpha = 1/2, \qquad \beta = \sqrt{\frac{\log \epsilon^{-1}}{d}}.$$

In the next subsection we shall see that, using repetition encoding schemes as the one above such a performance can not be beaten.

5.2 Bounds on the performance of low-complexity coding schemes

We now consider a more general class of encoders encompassing the previous example. As our main result, we shall show that, in any case, with such bounded complexity schemes, exponential decay of error can never be achieved. We shall first consider finite-state automata encoders and then finite-window encoders. As above we assume that $\mathcal{S} : [0,1]^d \to \{0,1\}$ is the dyadic quantizer introduced in Sect.3 and we consider encoders $\tilde{\mathcal{E}} : \{0,1\}^{\mathbb{N}} \to \{0,1\}^{\mathbb{N}}$.

5.2.1 Finite-state automata encoders

Encoders which can be implemented as finite state automata yield very poor performance. In fact, the mean square error Δ_t in this case does not converge to 0 as $t \to +\infty$. Indeed, assume we are given a finite state alphabet Z and two maps

$$\xi : Z \times \{0,1\} \to Z, \quad \rho : Z \times \{0,1\} \to \{0,1\}$$

Fix moreover an initial state $z^* \in Z$. To the quadruple (Z, ξ, ρ, z^*) we can naturally associate an encoder $\tilde{\mathcal{E}}$: given $(w_s)_{s=1}^{\infty} \in \{0,1\}^{\mathbb{N}}$, then we can define $(y_s)_{s=1}^{\infty} = \tilde{\mathcal{E}}((w_s)_{s=1}^{\infty})$ recursively by

$$\begin{cases} z_{t+1} &= \xi(z_t, w_t) \quad z_0 = z^* \\ y_t &= \rho(z_t, w_t) \end{cases}$$

Notice that the state updating map ξ together with the initial condition $z_0 = z^*$ yield a sequence of maps $\xi^{(t)} : \{0,1\}^t \to Z$ such that $z_{t+1} = \xi^{(t)}(w_1, \dots w_t)$. If we choose $t = t_0$ in such a way that $2^{t_0} > |Z|$, the map $\xi^{(t_0)}$ is, for sure, not injective. Hence, there exist two different input truncated sequences $(w_1', \dots, w_{t_0}')$ and $(w_1'', \dots, w_{t_0}'')$ such that $\xi^{(t_0)}(w_1', \dots, w_{t_0}') = \xi^{(t_0)}(w_1'', \dots, w_{t_0}'')$. Consider the event $A = \{w_k = w_k', z_k = ?$ for $k = 1, \dots, t_0\}$. Clearly, conditioned to A, the decoder, for any $t \geq t_0$, will decode incorrectly at least one information bit in the first t_0 position with positive probability independent from t. Hence,

$$\Delta_t^2 \geq \mathbb{E}\left[\|x - \hat{x}_t\|^2 \mid A \right] \mathbb{P}(A) \geq 2^{-2t_0/d} \mathbb{P}(A) > 0.$$

5.2.2 Finite-window encoders

Finite-window encoders are encoders of the form

$$\tilde{\mathcal{E}}((w_s)_{s=1}^{m_t})_t = f_t((w_s)_{s \in \Theta_t}) \tag{36}$$

where $\Theta_t \subseteq \mathbb{N}$ has finite cardinality $|\Theta_t| = n_t$, and where $f_t : \{0,1\}^{\Theta_t} \to \{0,1\}$. With each finite-window encoder it is possible to associate, for every $j, t \in \mathbb{N}$, the quantity

$$\omega_j(t) := \sum_{s=1}^{t} \mathbb{1}_{\Theta_s}(j)$$

14

counting the number of channel inputs up to time t, which have been affected by w_j. Notice that

$$\chi_t := \sum_{j \in \mathbb{N}} \omega_j(t) = \sum_{s \leq t} n_s \,.$$

The quantity χ_t can be thought of as measuring the complexity of the encoder $\tilde{\mathcal{E}}$. Indeed, if the maps f_t are $\mathbb{Z}_2$-linear, then χ_t coincides with the number of binary operations implemented by the encoder up to time t.

The following is our main result, relating the mean square error Δ_t to the complexity parameter χ_t.

Theorem 5. *For any transmission scheme for the BEC, with erasure probability ϵ, consisting of a finite window encoder of the form (36) with complexity function χ_t, it holds*

$$\Delta_t \geq C \, 2^{-\sqrt{\frac{1}{d}\chi_t \log \epsilon^{-1}}} \,, \tag{37}$$

where $C > 0$ is a constant depending only on d, the erasure probability ε and the density function f of the random vector x.

Proof Assume that, at time t, all the $\omega_j(t)$ channel inputs affected by the j-th bit w_j have been erased. Then, there is clearly no way for the decoder to reliably recover w_j from the channel output. This gives the following lower bound to the squared estimation error, independent of the way the decoders are chosen

$$\Delta_t^2 \geq C_1 \sup_{j \in \mathbb{N}} \left\{ 2^{-2j/d} \, \varepsilon^{\omega_j(t)} \right\} \,,$$

for some constant $C_1 > 0$ only depending on d and f. It will be convenient to consider the looser bounds

$$\Delta_t^2 \geq C_1 \sup_{1 \leq j \leq s} \left\{ 2^{-2j} \varepsilon^{\omega_j(t)} \right\} \geq C_1 \psi_s(\omega_1(t), \dots, \omega_s(t)) \,, \qquad \forall \, s \in \mathbb{N} \,,$$

where $\psi_s : (\mathbb{R}^+)^s \to \mathbb{R}$ is defined as follows

$$\psi_s(\omega_1, \dots, \omega_s) := \frac{1}{s} \sum_{j=1}^{s} 2^{-2j/d} \, \varepsilon^{\omega_j} \,.$$

Hence, for every possible s,

$$\Delta_t^2 \geq C_1 \inf_{\omega \in M_s} \psi_s(\omega_1, \dots, \omega_s) \tag{38}$$

where $M_s := \left\{ \omega_1, \dots, \omega_s \in (\mathbb{R}^+)^s \mid \sum_j \omega_j = \chi_t \right\}$. Since the function ψ_s is strictly convex, it admits a unique minimum on the convex compact set M_s. Using Lagrange multipliers we can characterize the unique stationary point of $\psi_s(\omega_1, \dots, \omega_s)$ on the hyperplane M_s

$$\omega_j^* = \varsigma - \rho j \,, \qquad \forall j \leq s \,,$$

where $\rho := \frac{\ln 4}{d \ln \varepsilon^{-1}} = \frac{2}{d \log \epsilon^{-1}} > 0$, and $\varsigma = \frac{\chi_t}{s} + \rho \frac{s+1}{2}$. We have that $\omega^* \in M_s$ if and only if $\omega_s^* \geq 0$ which is equivalent to

$$s \leq \frac{1}{2} \left(1 + \sqrt{1 + \frac{8\chi_t}{\rho}} \right) \,.$$

A possible choice is provided by $s^* = \left\lfloor \sqrt{2\chi_t/\rho} \right\rfloor$. We thus obtain

$$\Delta_t^2 \geq C_1 \inf_{\omega \in M_{s^*}} \psi_{s^*}(\omega_1, \dots, \omega_s) = \psi_{s^*}(\omega_1^*, \dots, \omega_s^*) = C_1 e^{-\varsigma \ln \epsilon^{-1}} \,. \tag{39}$$

We can estimate ζ^* as follows

$$\begin{aligned}
\varsigma^* &= \frac{\chi_t}{\left\lfloor \sqrt{\frac{2\chi_t}{\rho}} \right\rfloor} + \rho \frac{\left\lfloor \sqrt{\frac{2\chi_t}{\rho}} \right\rfloor + 1}{2} \leq \frac{\chi_t}{\sqrt{\frac{2\chi_t}{\rho}} - 1} + \frac{\rho}{2} \left(\sqrt{\frac{2\chi_t}{\rho}} + 1 \right) \\
&= \sqrt{\rho} \frac{2\chi_t - \rho/2}{\sqrt{2\chi_t} - \sqrt{\rho}} \leq \rho \left(\sqrt{\frac{2\chi_t}{\rho}} + \frac{2\sqrt{2} - 1}{2\sqrt{2} - 2} \right) \,,
\end{aligned}$$

the last equality following from the assumption $\chi_t \geq \rho$. Inserting this last estimation inside (39), the claim follows. ∎

15

Remark 5. In the case of the repetition encoders treated in Sect.5.1, we have that $\chi_t = t$. If we compare (37) with (33), considering the fact that ς can be picked arbitrarily close to 0, we have thus established that among the repetition schemes ($\chi_t = t$), the example treated in Sect.5.1 is optimal from the point of view of the asymptotic performance.

Remark 6. The bound (37) implies that, for the estimation error Δ_t to decrease to zero exponentially fast in t, then χ_t needs to grow quadratically in t. Hence, in order to obtain exponential convergence of the error, it is necessary that $\frac{1}{t}\chi_t$, i.e. the average number of bits of the dyadic expansion $\mathcal{S}(x)$ the channel inputs depend on, grows linearly in t. Indeed, observe that the random linear codes proposed in Sect.4.2 have exactly this property. However, observe that this does not imply that linear-time encodable schemes cannot attain exponential error decays in any case, for this might be achieved, for instance, by encoders obtained as concatenation of finite-window with finite-state automata schemes.

5.3 Simulation results for finite-window coding schemes

We shall now present Monte Carlo simulation results for some finite-window $\mathbb{Z}_2$-linear coding schemes with low-complexity iterative decoding. These schemes are based on ideas similar to those of digital fountain codes (see [18][21, Ch.50]). The latter are widely used in many applications, such as data storage, or reliable transmission on broadcast channels with erasures. The main additional challenge posed by our application consists in providing unequal error protection to the source bits.

We propose the following random construction for finite-window encoders fitting in the framework of Sect.5.2.2. As usual, assume that we have a dyadic quantizer $\mathcal{S}$ which maps the vector x into an infinite string of bits $(w_s)_{s=1}^{\infty}$. We imagine that at each time t the encoder produces a bit y_t which is the (modulo-2) sum of a random number of randomly chosen w_s, namely

$$y_t = \sum_{s \in \Theta_t} w_s \, .$$

where Θ_t is a random subset of $\mathbb{N}$. We assume that the cardinality of Θ_t is bounded, namely $|\Theta_t| \leq n_{\max}$.

More precisely, fix $n_{\max} \in \mathbb{N}$, and a probability distribution $\mu(\cdot)$ on $\{1, \ldots, n_{\max}\}$. Randomly generate a sequence $(n_t)_{t \in \mathbb{N}}$ of independent random variables distributed accordingly to $\mu(\cdot)$. Let $(\nu_t(\cdot))_{t \in \mathbb{N}}$ be a sequence of probability distributions over $\mathbb{N}$, with $\nu_t(\cdot)$ possibly depending on $(n_s)_{s \leq t}$. Then, for all $t \geq 1$, we let

$$\Theta_t := \{\theta_{1,t}, \theta_{2,t}, \ldots, \theta_{n_t,t}\}$$

where $\theta_{i,t}$ are independent random variables unformly distributed according to $\nu_t(\cdot)$. Notice that in this way we have that $|\Theta_t| \leq n_t \leq n_{\max}$ and so the encoder complexity is linear in t.

For the decoding, a sequential implementation of the peeling algorithm is used, this being the standard decoding technique for digital fountain codes [18][21, Ch.50]. Such an algorithm works on an iteratively updated infinite hypergraph[12] $\mathcal{G}_t = (\mathcal{V}_t, \mathcal{H}_t)$ as explained below. At $t = 0$, $\mathcal{G}_0$ is initialized with vertex set $\mathcal{V}_0 = \mathbb{N}$ and empty hyperedge set $\mathcal{H}_0 = \emptyset$. The estimates $(\hat{w}_s(0))_{s \in \mathbb{N}}$ of the dyadic expansion $\mathcal{S}(x)$ are in turn initialized arbitrarily in $\{0, 1\}^{\mathbb{N}}$. At each time $t \geq 1$, first update $\mathcal{V}_t = \mathcal{V}_{t-1}$, $\mathcal{H}_t = \mathcal{H}_{t-1}$, and $\hat{w}_s(t) = \hat{w}_s(t+1)$ for all $s \in \mathbb{N}$. Then:

- if $z_t =?$, then quit; if $z_t \neq ?$, update $\mathcal{H}_t = \mathcal{H}_t \cup \{B_t\}$, where $B_t := \Theta_t \cap \mathcal{V}_t$;

- if $|B_t| > 1$, then quit; otherwise if $B_t = \{v\}$ for some $v \in \mathcal{V}_t$, set $\hat{w}_v(t) = z_t + \sum_{j \in \Theta_t \setminus \{v\}} \hat{w}_j(t)$, eliminate v from $\mathcal{V}_t$ as well as from all the hyperedges $h \in \mathcal{H}_t$ containing it;

- if $|h| \neq 1$ for all $h \in \mathcal{H}_t$, quit; otherwise, if there is some $h = \{v\} \in \mathcal{H}_t$, repeat the previous step.

The above-described algorithm requires an order of $\chi_t = \sum_{s \leq t} n_s$ operations up to time t, hence it has linear complexity in t. It is suboptimal with respect to the maximum a posteriori decoding analyzed in Sect.4.2, as it may fail to correctly estimate the first j bits of the dyadic expansion $\mathcal{S}(x)$ even when that would be possible using the maximum a posteriori decoder introduced in Sect.4.2.2.

---

[12]The term hypergraph [3, pag.7] refers to a pair $(\mathcal{V}, \mathcal{H})$, where $\mathcal{V}$ is a discrete set and $\mathcal{H}$ is a subset of $\mathcal{P}(\mathcal{V})$, the power set of $\mathcal{V}$.

| (a) | (b) |
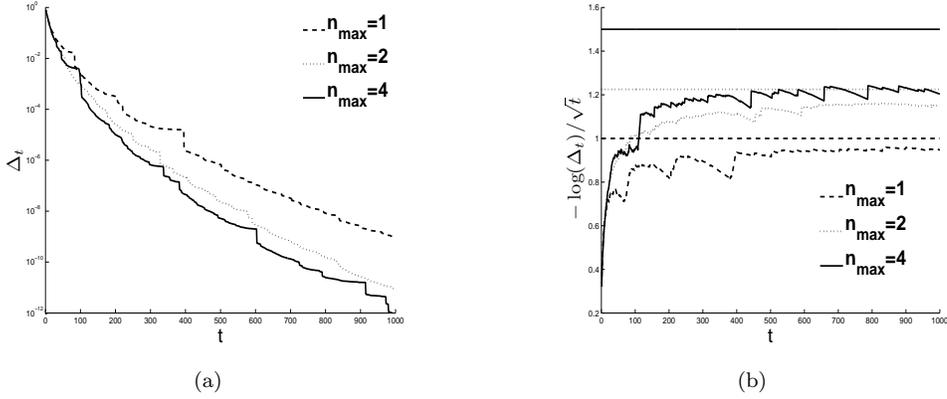
Figure 2: Monte Carlo simulations of finite-window coding schemes on the BEC, with erasure probability $\varepsilon = 0.5$. The performance of three coding schemes are compared: these schemes were randomly generated accordingly to (40) and (41) with $n_{\max} = 1, 2, 4$ respectively. In (a) the mean squared error Δ_t is plotted as a function of the time t in log-linear scale. In (b) $-\frac{1}{\sqrt{t}} \log \Delta_t$ is plotted as a function of t, together with the corresponding upper bounds $\sqrt{\chi_t \log \varepsilon^{-1}}$ provided by Theorem 5. The number of samples used is 200000.

In Fig.2 we report Monte Carlo simulations of three finite-windows encoding schemes, with $n_{\max} = 1, 2, 4$ respectively. The degree distribution $\mu(\cdot)$ was chosen to be the truncated solyton one [21, pag.592]

$$\mu(1) := \frac{1}{n_{\max}}, \qquad \mu(n) := \frac{1}{n(n-1)} \quad \forall\, 2 \le n \le n_{\max}. \tag{40}$$

The distributions ν_t have been selected as follows. We let

$$\rho = \frac{2}{d \log \varepsilon^{-1}}, \qquad s_t := \lfloor \sqrt{2\chi_t \rho^{-1}} \rfloor, \qquad \varsigma_t = \frac{\chi_t}{s_t} + \rho \frac{s_t + 1}{2},$$

where $\chi_t = \sum_{s \le t} n_s$. Then we have chosen

$$\nu_t(j) := \begin{cases} \eta(\varsigma_t - \rho j) & \text{if} \quad j \le s_t \\ 0 & \text{if} \quad j > s_t, \end{cases} \tag{41}$$

Such a choice was suggested by the optimization problem in the right-hand side of (38).

It is clear from Fig.2(a) that the three schemes have subexponential error decay and that increasing the degree allows to obtain better convergence rates. Fig.2(b) shows that the convergence degree is $\alpha = 1/2$, as expected from the theory, while it is possible to recognize the different values of β of the three schemes, in the asymptotic limit of $-\frac{1}{\sqrt{t}} \log \Delta_t$.

It should be underlined as the choices of the distributions μ and ν_t were not optimized, but rather suggested by the literature on digital fountain codes and by Theorem 5, respectively. A theoretical analysis of the behavior of finite-window schemes, hopefully providing hints on the design of μ and ν_t, is left as a topic for future research.

6 Conclusions

The problem of anytime reliable transmission of a real-valued random vector through a digital noisy channel has been addressed. Upper and lower bounds on the highest exponential rate achievable for the mean squared error have been obtained assuming transmission over the BEC. Moreover, a lower bound on the performance achievable by low-complexity coding schemes have been derived. This bound shows that if we want that the mean squared error decreases exponentially fast in the number of channel uses, than we need to adopt an encoder in which the channel input depends on a number of bits of the vector expansion which grows linearly. Finally, simulation results for linear-complexity coding/decoding schemes have been proposed.

Many of the questions raised in this paper have been left open. Among them, a particularly relevant issue is the analysis and design of linear-complexity coding schemes achieving exponential error rates. Another problem consists in tightening the upper bound on the achievable error exponent proved in Theorem 2, by better exploiting the absence of feedback. Current work includes extension of the theory to distributed estimation/computation problems over networks of agents communicating through noisy digital channels.

A   Proof of Theorem 4

We shall now prove Theorem 4 by means of so-called code-expurgation arguments. The Hamming weight of a binary string $\boldsymbol{y} \in \mathbb{Z}_2^t$ will be denoted by $w_H(\boldsymbol{y}) := |\{1 \le j \le t : y_j = 1\}|$. For $t \in \mathbb{N}$, $0 \le j \le m_t$, and $h \ge 0$, let us consider the number of binary strings $\boldsymbol{y}$ whose first non-zero bit is the $(j+1)$-th and such that $\tilde{\mathcal{E}}_t \boldsymbol{y}$ has weight h. Since ϕ is random, the aforementioned is a random variable, which will be denoted by

$$\Upsilon_j^t(h) := \left| \left\{ \boldsymbol{y} \in K_j \setminus K_{j+1} : w_H(\tilde{\mathcal{E}}_t \boldsymbol{y}) = h \right\} \right| , \qquad h \ge 0 ,$$

where we recall that $K_j := \mathrm{span}(\delta_{j+1}, \dots, \delta_{m_t}) \le \mathbb{Z}_2^{m_t}$.

Observe that the causality of ϕ implies that, if $\boldsymbol{y} \in K_j$, then $\tilde{\mathcal{E}}_t \boldsymbol{y}$ belongs to $L_j := \mathrm{span}\left(\delta_s \mid \lceil (j+1)/R \rceil \le s \le t\right) \le \mathbb{Z}_2^t$. Further, since $\phi \delta_{j+1}$ is uniformly distributed over L_j, and since the columns of ϕ are independent, we have that, if $\boldsymbol{y} \in K_j \setminus K_{j+1}$, then $\tilde{\mathcal{E}}_t \boldsymbol{y}$ is a random variable uniformly distributed over L_j. It follows that

$$\begin{aligned}
\mathbb{E}[\Upsilon_j^t(h)] &= \sum_{\boldsymbol{y} \in K_j \setminus K_{j+1}} \mathbb{P}(w_H(\tilde{\mathcal{E}}_t \boldsymbol{y}) = h) \\
&= |K_j \setminus K_{j+1}| \binom{l_j}{h} |L_j|^{-1} \\
&\le 2^{l_j(\mathrm{H}(\eta) - 1 + R)} ,
\end{aligned}$$

where $l_j := (t - \lceil (j+1)/R \rceil + 1)$ and $\eta := h/l$.

For every $\lambda, \varphi > 0$, by using the union bound and Markov's inequality, we can estimate the probability of the event

$$F_t := \bigcup_{j=1}^{(1-\lambda)Rt} \bigcup_{h=0}^{l_j \gamma (R+\varphi)} \{\Upsilon_j^t(h) \ge 1\}$$

as follows:

$$\mathbb{P}(F_t) \le \sum_{j,h} \mathbb{E}[\Upsilon_j^t(h)] \le t^2 2^{-t\lambda\varphi} .$$

Then, the series $\sum_{n \in \mathbb{N}} \mathbb{P}(F_n)$ is convergent, and the Borel-Cantelli lemma implies that, with probability one, F_n occurs finitely many times, i.e. there exists $t_0 \in \mathbb{N}$ such that

$$\Upsilon_j^t(h) = 0 , \qquad \forall\, h < l_j \gamma(R + \varphi) .$$

for all $t \ge t_0$ and $1 \le j \le (1 - \lambda)Rt$. An analogous argument shows that with probability one

$$\Upsilon_j^t(h) \le 2^{l_j(\mathrm{H}(\eta) - 1 + R + \varphi))} , \qquad \forall\, l_j \gamma(R + \varphi) \le h \le l_j ,$$

for sufficiently large t.

We are now ready to prove Theorem 4. For this, fix $\lambda, \varphi \in (0,1)$, and consider the event $H_t := \bigcup_{j=1}^{\lfloor (1-\lambda)Rt \rfloor} G_t^j$, where

$$G_t^j := \bigcup_{h=0}^{l_j \gamma(R+\varphi)} \{\Upsilon_j^t(h) \ge 1\} \bigcup_{h=l_j \gamma(R+\varphi)}^{l_j} \left\{ \Upsilon_j^t(h) \le 2^{(t-\lceil j/R \rceil)(\mathrm{H}(\eta)-1+R+\eta))} \right\} .$$

Then, for $j = 1, \dots, \lfloor (1-\lambda)Rt \rfloor$, the union bound for the event B_j defined in (20) yields the estimation

$$\mathbb{P}(B_j | H_t) \le \sum_{h=0}^{l_j} \varepsilon^h \mathbb{E}\left[\Upsilon_j^t(h) | H_t\right] \le \sum_{h=l_j \gamma(R+\varphi)}^{l_j} \varepsilon^h 2^{l_j(\mathrm{H}(\eta)-1+R+\eta))} .$$

18

Hence, (21) and (22) imply that

$$
\begin{aligned}
\mathbb{E}[||x - \hat{x}_t||^2 | H_t] &= \sum_{j=1}^{\lfloor (1-\lambda)Rt \rfloor} \mathbb{E}[||x - \hat{x}_t||^2 | H_t \cap B_j] \mathbb{P}(B_j | H_t) + \mathbb{E}[||x - \hat{x}_t||^2 \mathbb{1}_{A_{\lfloor R(1-\lambda)t \rfloor}} | H_t] \\
&\leq \sum_{j=1}^{\lfloor (1-\lambda)Rt \rfloor} 16d2^{-2j/d} \sum_{h=l_j \gamma(R+\varphi)}^{l_j} \varepsilon^h 2^{l_j(\mathrm{H}(\eta)-1+R+\varphi)} + 16d2^{-2\lfloor (1-\lambda)Rt \rfloor/d} \\
&\leq K't^2 2^{-t(2\underline{\beta}''(d,\varepsilon,R)-\varphi)} + K''t2^{-2(1-\lambda)Rt},
\end{aligned}
$$

for some constants $K', K'' > 0$. Since, with probability one, there exists $t_0 \in \mathbb{N}$ such that H_t occurs for all $t \geq t_0$, for all such t we have

$$
\mathbb{E}[||x - \hat{x}_t||^2 | \phi] \leq K't^2 2^{-t(2\underline{\beta}''(d,\varepsilon,R)-\varphi)} + K''t2^{-2(1-\lambda)Rt/d}.
$$

It follows that

$$
\liminf_t -\frac{1}{t} \log \mathbb{E}[||x - \hat{x}_t||^2 | \phi] \geq \min \left\{ 2\underline{\beta}''(d,\varepsilon,R) - \varphi, \tfrac{2}{d}(1-\lambda)R \right\}
$$

with probability one, and the claim of Theorem 4 follows from the arbitrariness of $\varphi, \lambda > 0$. ∎

References

[1] E. R. Berlekamp, R. J. McEliece and H. C. A. Van Tilborg, "On the inherent intractability of certain coding problems", *IEEE Trans. Inf. Theory*, vol. 24, pp. 384–386, 1978.

[2] D. Bertsekas and J. Tsitsiklis, *Parallel and Distributed Computation: Numerical Methods*, Athena Scientific, Belmont (MA), 1997.

[3] B. Bollobas, *Modern graph theory*, Springer Verlag, New York, 1998.

[4] S. Borade, B. Nakiboğlu and L. Zheng, "Unequal error protection: some fundamental limits", submitted, 2008.

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

[6] I. Csiszàr, "Joint source-channel error exponent", *Probl. Control Inf. Theory*, 1980, vol. 9, pp. 315-328.

[7] I. Csiszàr, "On the error exponent of source-channel transmission with a distortion threshold", *IEEE Trans. Inf. Theory*, 1982, vol. 28, pp. 823-828.

[8] I. Csiszàr, "The method of types", *IEEE Trans. Inf. Theory*, vol. 44, pp. 2505-2523, 1998.

[9] G. Cybenko, "Dynamic load balancing for distributed memory multiprocessors", *Journal of parallel and distributed computing*, vol. 7, pp. 279-301, 1989.

[10] R. Diekmann, A. Frommer, and B. Monien, "Efficient schemes for nearest neighbor load balancing", *Parallel computing*, vol. 25, 789-812, 1999.

[11] A. G. Dimakis, A. D. Sarwate, M. J. Wainwright, "Geographic Gossip: Efficient Averaging for Sensor Networks", *IEEE Trans. Sig. Proc.*, vol. 56, pp. 1205-1216, 2008.

[12] G. D. Forney Jr.,"Convolutional codes II. Maximum-likelihood decoding", Inf. Control, 25, pp. 222-266, 1974.

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

[14] S. Graf and H. Luschgy, *Foundations of quantization for probability distributions*, Spriger, Berlin, 2000.

[15] C. Intanagonwiwat, R. Govindan, and D. Estrin, "Directed diffusion: a scalable and robust communication paradigm for sensor networks", In Proc. ACM/IEEE Conf. Mobile Computing and Networking, pp. 56-67, 2000.

[16] A. Jadbabaie, J. Lin and A.S. Morse. "Coordination of groups of mobile autonomous agents using nearest neighbor rules", *IEEE Trans. Automat. Control*, vol. 48, pp. 988-1001, 2003.

[17] D. Kempe, A. Dobra and J. Gehrke, "Gossip-based computation of aggregate information", Proc. 44th IEEE Symp. on Foundations of Computer Science, pp. 1-10, 2003.

[18] M. Luby, "LT codes", in Proc. 43rd Annual IEEE Sym- posium on Foundations of Computer Science, pp. 271-282, November 16-19, 2002.

[19] A.S. Matveev and A.V. Savkin, "Shannon zero error capacity in the problems of state estimation and stabilization via noisy communication channels", *Int. J. Control*, vol. 80, pp. 241255, 2007.

[20] A.S. Matveev, "State estimation via limited capacity noisy communication channels", *Mathematics of Control, Signals, and Systems*, vol. 20, pp. 135, 2008.

[21] D. McKay, *Information theory, inference, and learning algorithms*, Cambridge University Press, Cambridge, 2003.

[22] B. Masnick and J. Wolf, "On linear unequal error protection", *IEEE Trans. Inf. Theory*, vol. 13, pp. 600-607, 1967.

[23] L. Moreau, "Stability of multiagent systems with time-dependent communication links", *IEEE Trans. Automat. Control*, vol. 50, 169-182, 2005.

[24] S. Muthukrishnan, B. Ghosh, and M. Schultz, "First and second order diffusive methods for rapid, coarse, distributed load balancing", *Theory of computing systems*, vol. 31, pp. 331-354, 1998.

[25] R. Olfati-Saber and R.M. Murray, "Consensus problems in networks of agents with switching topology and time-delays", *IEEE Trans. Automat. Control*, vol. 49, pp. 1520-1533, 2004.

[26] W. Ren and R.W. Beard, "Consensus seeking in multiagent systems under dynamically changing interaction topologies", *IEEE Trans. Automat. Control*, vol. 50, pp. 655-661, 2005.

[27] T.J. Richardson and R. Urbanke, *Modern Coding Theory*, Cambridge Univeristy Press, Cambridge, 2007.

[28] T. Simsek, R. Jain and P. Varaiya, "Scalar estimation and control with noisy binary observation", *IEEE Trans. Automat. Control*, vol. 49, pp. 15981603, 2004.

[29] A. Sahai and S. Mitter, "The necessity and sufficiency of anytime capacity for stabilization of a linear system over a noisy communication link–Part I: scalar systems", *IEEE Trans. Inf. Theory*, vol. 52, pp. 3369-3395, 2006.

[30] A. Sahai, "Why do block length and delay behave differently if feedback is present?", *IEEE Trans. Inf. Theory*, vol. 54, pp. 1860-1886, 2008.

[31] A. Sahai and S. Mitter, "Source coding and channel requirements for unstable processes", submitted, 2006.

[32] S. Tatikonda and S. Mitter, "Control under communication constraints", *IEEE Trans. Automat. Control*, vol. 49, pp. 1056-1068, 2004.

[33] S. Tatikonda and S. Mitter, "Control over noisy channels", *IEEE Trans. Automat. Control*, vol. 49, pp. 11961201, 2004.

[34] J. Tsitsiklis, *Problems in decentralized decision making and computation*, Ph.D. dissertation, Dep. Elec. Eng. Comput. Sci., Mass. Inst. Technol., Cambridge, MA, 1984.

[35] L.N. Treften and D. Bau, III, *Numerical linear algebra*, SIAM, Philadelphia (PA), 1997.

[36] A.J. Viterbi, "Error bounds for convolutional codes and an asymptotically optimal decoding algorithm", *IEEE Trans. Inf. Theory*, vol. 13, pp. 260-269, 1967.

[37] A.J. Viterbi, "Further results on optimal decoding of convolutional codes", *IEEE Trans. Inf. Theory*, vol. 15, pp. 732-734, 1969.

[38] J. Zhao, R. Govindan and D. Estrin, "Computing aggregates for monitoring wireless sensor networks", Proc. Int. Workshop on Sensor Net Protocols and Applications, pp. 139-148, 2003.