ANYTIME RELIABLE TRANSMISSION OF REAL-VALUED INFORMATION THROUGH DIGITAL NOISY CHANNELS

GIACOMO COMO *, FABIO FAGNANI †, AND SANDRO ZAMPIERI ‡

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, an anytime transmission scheme consists of an encoder –which maps the real vector into a sequence of channel inputs– and a decoder –which sequentially updates its estimate of the vector 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 its 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, 40], or load balancing [12, 13, 29]; wireless sensor networks, in which the final goal is estimation and decision making from distributed measurements [18, 20, 44, 14]; sensors/actuators networks, such as mobile multi-agent networks, in which the final goal is control [19, 31, 28, 32]. 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, i.e. 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 transmission algorithms, i.e. algorithms which can be stopped anytime while providing estimations of increasing precision. These algorithms are particularly suitable for applications in problems of distributed control.

As especially pointed out in a series of works by A. Sahai and S. Mitter [34, 35, 36], 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, such delays can be detrimental for the system performance in several control applications. Here, the fundamental question is not only where, but also when

*Laboratory for Information and Decision Systems, Massachusetts Institute of Technology, 77 Mass Ave, Cambridge (MA), 02139, US (giacomo@mit.edu).

†Dipartimento di Matematica, Politecnico di Torino, Corso Duca degli Abruzzi 24, 10129 Torino, Italy (fabio.fagnani@polito.it).

‡Dipartimento di Elettronica e Informatica, Università di Padova, via Gradenigo 6/A, 35131, Padova, Italy, (zampi@dei.unipd.it). The research leading to these results has received funding from the European Community's Seventh Framework Programme under agreement n.'FP7-ICT-223866-FeedNetBack.

the information is available. For this reason, it is often desirable to use transmission systems for control applications which are able to provide estimates whose precision increases with time, so as providing a reasonable partial information transmission anytime the process is stopped.

On the other hand, the computational complexity of the transmission schemes is a central issue. In fact, nodes in wireless 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, both in the encoding and in the decoding. 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. Known results in Information Theory [8] show that feedback can improve the capacity of channels with memory, or multiple access channels[1], as well as reduce latency and computational complexity. In many cases of practical interest, however, feedback information is incomplete, or difficult to be used. Also, there are many situations, for instance in the wireless network scenario, in which the transmitter needs to broadcast its 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 ourselves 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, information bits are usually assumed to be equally valuable, and they are consequently given the same priority by the transmission system designer. In fact, design paradigms of modern low-complexity codes [26, 33] –based on random sparse graphical models and iterative decoding algorithms– treat information bits as equally valuable. While such an assumption is typically justified by the source-channel separation principle, this 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 [9, 10]. In fact, in many problems of information transmission for control or estimation, different information bits typically require significantly different treatment.

As an example, particularly relevant for the topics addressed in this paper, assume that a random parameter, uniformly distributed over a unitary interval, has to be reliably transmitted through a digital noisy channel (see [5] and references therein for the analysis of the information theoretic limits of this problem on the bandwidth-unlimited Gaussian channel). Such a parameter may be represented by its dyadic expansion, which is a stream of independent identically distributed bits. Clearly, such information bits are not equally valuable, since the first one is more significant than the second one, and so on. This motivates the study of unequal error protection codes [27, 4]. One of the challenges posed by information transmission for control/estimation applications is to come up with design paradigms for practical, low-complexity, unequal error protection codes.

In this paper, we shall propose two classes of coding strategies for the anytime transmission of real-valued random vectors through a digital noisy channel. In both cases, the transmission scheme consists of an encoder, mapping the real vector into a

[1] Whereas a classic result due to Shannon shows that feedback does not improve the capacity of a discrete memoryless channel.

sequence of channel inputs, 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 have 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), and defer any discussion on the possible extensions to general discrete memoryless channels to the concluding section. In the BEC, a transmitted binary signal is either correctly received, or erased with some probability ε. While this channel allows for an elementary treatment, it is of its own interest in many scenarios. In [6], the techniques proposed here have been applied in order to obtain a version of the average consensus algorithm working in presence of digital erasure communication channels between the nodes.

The rest of this paper is organized as follows. Sect. 2 formally states the problem. Sect. 3 presents an upper bound to the possible error convergence of any coding scheme over the BEC. In Sect. 4 we introduce the class of encoder/decoder schemes used throughout the paper and which are based on a preliminary vector quantization of the continuous vector to be transmitted. 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. In Sect. 6, random linear convolutional codes are shown to achieve exponential error rates at the cost of computational complexity growing quadratically in time. Finally, some Monte Carlo simulations of finite-window coding schemes are reported in Sect. 7.

We end this introduction by establishing some notation. Throughout the paper, $\mathbb{R}$ and $\mathbb{N}$ will denote the sets of reals and naturals, respectively. For a subset $A \subseteq B$, $|A|$ will denote the cardinality of A, $\overline{A} = B \setminus A$ its complement, and $\mathbb{1}_A : B \to \{0,1\}$ its indicator function, defined by $\mathbb{1}_A(x) = 1$ if $x \in A$, and $\mathbb{1}_A(x) = 0$ otherwise. The natural logarithm will be denoted by ln, while log will stand for the logarithm in base 2. 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 $0 \log 0 = 0$. For two sequences of reals $(a_t)_{t\in\mathbb{N}}$ and $(b_t)_{t\in\mathbb{N}}$, both the notations $a_t = O(b_t)$ and $b_t = \Theta(a_t)$ will mean that $a_t \leq K b_t$ for some constant K, while $a_t = o(b_t)$ will mean that $\lim_t a_t/b_t = 0$. A sequence a_t, $t = 1, 2, \ldots$ is sometimes denoted with the symbol $a = (a_t)_{t=1}^{\infty}$, while with the symbol $a = (a_t)_{t=1}^{T}$ we will mean its truncation to $t = 1, \ldots, T$.

2. Problem formulation. We shall now provide a formal description of the problem. Let x be a random variable taking values on $\mathcal{X} \subseteq \mathbb{R}^d$. We shall assume that x has an a priori probability law which is absolutely continuous with respect to the Lebesgue measure, and denote by $f(x)$ the probability density of x. Further, we shall assume that $\mathbb{E}||x||^{2+\delta} < +\infty$ for some $\delta > 0$. At time $t \in \mathbb{N}$, the communication channel has input y_t, and output z_t, taking values in some finite alphabets $\mathcal{Y}$, and $\mathcal{Z}$, respectively. Transmission is assumed to be memoryless, i.e., given the current input y_t, the output z_t is assumed to be conditionally independent from the previous inputs $(y_s)_{s=1}^{t-1}$ and outputs $(z_s)_{s=1}^{t-1}$, as well as from the vector x. The conditional probability of $z_t = z$ given $y_t = y$ will be assumed stationary and denoted by $p(z|y)$. We shall consider in detail the binary erasure channel (BEC) in which $\mathcal{Y} = \{0,1\}$,

$\mathcal{Z} = \{0, 1, ?\}$, and

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

Here, ? stands for the erased signal, and $\varepsilon \in [0, 1[$ for the erasure probability.

The anytime transmission scheme consists of an encoder and a sequential decoder.[2] The encoder consists of a family of maps $E_t : \mathcal{X} \to \mathcal{Y}$, specifying the symbol transmitted through the channel at time t, $y_t = E_t(x)$. With this family of maps we can associate the global map $\mathcal{E} : \mathcal{X} \to \mathcal{Y}^{\mathbb{N}}$ which specifies the infinite string that the encoder generates from x. The decoder instead is given by a family of maps $\mathcal{D}_t : \mathcal{Z}^t \to \mathcal{X}$, describing the estimate $\hat{x}_t = \mathcal{D}_t((z_s)_{s=1}^t)$ of x obtained from the string $(z_s)_{s=1}^t$ that has been received until time t. With this family of maps we can associate naturally the global map $\mathcal{D} : \mathcal{Z}^{\mathbb{N}} \to \mathcal{X}^{\mathbb{N}}$. This is represented in the following scheme

$$\begin{array}{ccccccc} \mathcal{X} & \xrightarrow{\mathcal{E}_t} & \mathcal{Y}^t & \xrightarrow{\text{Channel}} & \mathcal{Z}^t & \xrightarrow{\mathcal{D}_t} & \mathcal{X} \\ x & \longmapsto & (y_s)_{s=1}^t & \longmapsto & (z_s)_{s=1}^t & \longmapsto & \hat{x}_t \end{array} \tag{2.1}$$

where $\mathcal{E}_t := \pi_t \circ \mathcal{E}$ and where $\pi_t : \mathcal{Y}^{\mathbb{N}} \to \mathcal{Y}^t$ is the projection of a sequence in $\mathcal{Y}^{\mathbb{N}}$ into its first t symbols

In order to evaluate the performance of a scheme, we define the root 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

$$\Delta_t := (\mathbb{E}||x - \hat{x}_t||^2)^{1/2}\,. \tag{2.2}$$

In this paper, we shall be concerned with the rate of decay of Δ_t for different anytime transmission schemes. All the coding strategies which will be analyzed are characterized by a root mean squared error Δ_t converging to zero like $2^{-\beta t^\alpha}$ for some constants $\beta > 0$ and $0 < \alpha \leq 1$. More precisely we shall seek to find α, β such that

$$\Delta_t \leq p(t) 2^{-\beta t^\alpha} \tag{2.3}$$

for some polynomial $p(t)$. When (2.3) holds, the coding strategy will be said to achieve a degree of convergence α and rate of convergence β. When $\alpha = 1$ we shall simply say that we have an exponential convergence. In this case β is referred to as 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 [38, 39, 22, 37, 23, 24, 25] and references therein). Assume we are given a discrete time stochastic linear system

$$x(t+1) = Ax(t) + v(t)\,, \qquad x(0) = x_0\,, \tag{2.4}$$

[2]Our definition of anytime transmission scheme does not formally coincide with that in the Anytime Information Theory of S. Mitter and A. Sahai. Our usage of the term "anytime" has to be understood in the broader sense it has in Artificial Intelligence, where anytime algorithms are algorithms whose quality of results improves gradually as computation time increases [45].

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 $A \in \mathbb{R}^{n\times n}$ is a full rank, unstable matrix.

Suppose that a remotely positioned receiver is required to estimate the state of the system, while observing the output of a binary erasure channel only. Then, it is necessary to design a family of encoders E_t and of decoders 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 D_t has to obtain an estimate $\hat{x}(t)$ of the current state.

If we have that $v(t) = 0$ for every $t \geq 0$, then the only source of uncertainty is due to the initial condition x_0. Hence, in this case, the encoder/decoder task reduces to obtaining good estimates of x_0 at the receiver side. Indeed, in order to obtain a good estimate $\hat{x}(t)$ of $x(t)$, the receiver 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, one has $x(t) - \hat{x}(t) = A^t(x(0) - \hat{x}(0|t))$, 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 $x(0) - \hat{x}(0|t)$. The same technique can be applied if $v(t)$ is small with respect to x_0 as clarified by the following example.

EXAMPLE 1. *Consider the following unstable scalar discrete time linear system*

$$x(t+1) = ax(t) + v(t)\,, \qquad x(0) = x_0\,,$$

where $a > 1$ and where x_0 is a random variable with probability density $f(x)$ and $v(t)$ is a sequence of independent, identically distributed random variables with zero mean and variance σ_v^2, which are independent of x_0. Assume that a state estimation algorithm is run, based on the noiseless model $x(t+1) = ax(t)$ by estimating the initial condition x_0 from data transmitted until time t. As before, we shall denote this estimate by $\hat{x}(0|t)$. From $\hat{x}(0|t)$, we form the estimate $\hat{x}(t) := a^t\hat{x}(0|t)$ of $x(t)$. The estimation error at time t will be $e(t) := x(t) - \hat{x}(t) = a^t(x(0) - \hat{x}(0|t)) + \sum_{i=0}^{t-1} a^{t-1-i}v(i)$, so that $\mathbb{E}[e(t)^2] = a^{2t}\mathbb{E}[(x(0) - \hat{x}(0|t))^2] + \sigma_v^2\frac{1-a^{2t}}{1-a^2}$. This error depends both on the error in the estimation of the initial condition, and on the wrong model we used. As we shall see, our techniques yield an estimation error on $x(0)$ of the form $\mathbb{E}[(x(0) - \hat{x}(0|t))^2] = C\zeta(t)$, where C depends only on the probability density $f(x)$ and $\zeta(t)$ is a function converging to zero depending only on the communication channel characteristics and on the coding strategy. Therefore,

$$\mathbb{E}[e(t)^2] = a^{2t}\left[C\zeta(t) + \sigma_v^2\frac{1 - a^{-2t}}{a^2 - 1}\right].$$

In case C is much larger than σ_v^2, there will be an initial time regime in which the error is not influenced by the model noise but only by the estimation of the initial condition $x(0)$.

3. The limit of performance on the binary erasure channel. Observe that, in case of noiseless channel, the function mapping x into $(\hat{x}_0, \ldots, \hat{x}_{t-1})$ is a quantizer assuming at most 2^t values. It is well-known in the theory of vector quantization [17] that, if $\mathcal{Q} : \mathcal{X} \to \mathcal{X}$ is a quantizer assuming m values, then

$$(\mathbb{E}||x - \mathcal{Q}(x)||^2)^{1/2} \geq C_- m^{-1/d}\,, \tag{3.1}$$

where C_- is a positive constant only depending on the dimension d, and the a priori density $f(x)$. This shows that $\Delta_t \geq C_- 2^{-t/d}$ for all $t \in \mathbb{N}$. Hence, it is not possible

to obtain a convergence degree α greater than 1 with an exponential convergence rate β larger than $1/d$. In this section, we shall present a tighter upper bound on the exponential convergence rate of Δ_t on the BEC with erasure probability ε.

Consider the general scheme (2.1). 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 $z_t \in \{0, 1\}$ (this corresponds to a correct transmission), and $\xi_t =?$ if $z_t =?$ (this corresponds to an erased signal). Observe that, given the encoder $\mathcal{E}$ and the decoder $\mathcal{D}$, the error pattern $(\xi_t)_{t\in\mathbb{N}}$ is a random variable independent of the source vector x. This property will allow us to present for the BEC almost elementary proofs of results holding true also for general discrete memoryless channels.

For $j \leq t$, let

$$\lambda_j^t := \sum_{j\leq s\leq t} \mathbb{1}_{\{\xi_s=c\}} \tag{3.2}$$

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

$$\mathbb{P}(\lambda_j^t = l) = \binom{t-j+1}{l} \varepsilon^{t-j+1-l}(1-\varepsilon)^l\,, \qquad l = 0, \ldots, t-j+1\,. \tag{3.3}$$

The simple observation above allows one to prove the following result.

THEOREM 3.1. *Assume transmission over the BEC with erasure probability $\varepsilon \in [0, 1]$. Then, the estimation error of any coding scheme as in (2.1) satisfies*

$$\Delta_t \geq C_-\ 2^{-t\overline{\beta}(d,\varepsilon)}\,, \tag{3.4}$$

for all $t \geq 0$, where

$$\overline{\beta}(d, \varepsilon) := -\frac{1}{2}\log\left(\varepsilon + (1-\varepsilon)2^{-2/d}\right) \tag{3.5}$$

and C_- is a constant depending only on the dimension d and the a priori density $f(x)$.

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 (2.1) becomes a quantizer from $\mathcal{X}$ to itself with a range of cardinality not larger than $2^{\lambda_1^t}$. From this fact and from (3.1), we can deduce that $\mathbb{E}\left[||x - \hat{x}_t||^2 |\, \lambda_1^t = l\right] \geq C_-^2\, 2^{-2l/d}$. Therefore,

$$\begin{aligned} \mathbb{E}\left[||x-\hat{x}_t||^2\right] &= \textstyle\sum_{l=0}^t \mathbb{E}\left[||x-\hat{x}_t||^2 |\, \lambda_1^t = l\right] \mathbb{P}(\lambda_1^t = l) \\ &\geq C_-^2 \textstyle\sum_{l=0}^t 2^{-2l/d}\binom{t}{l}\varepsilon^{t-l}(1-\varepsilon)^l \\ &= C_-^2\left(\varepsilon + (1-\varepsilon)2^{-2/d}\right)^t \end{aligned} \tag{3.6}$$

From the inequality above, the claim follows. □

REMARK 1. *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*[3]

$$\overline{\beta}(d, \varepsilon) < \frac{1}{d}(1-\varepsilon)\,, \qquad \forall \varepsilon \in]0, 1[\,. \tag{3.7}$$

[3]See also Fig.6.1.

The inequality (3.7) 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 $(1-\varepsilon)t$. In fact, the latter is the lower bound one would have obtained by simply using a rate distortion argument. Indeed, a closer look at (3.6) 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, i.e. by the events $\{\lambda_1^t = l_t^*\}$ of probability exponentially vanishing in t.

It is not hard to see that (3.4) continues to hold true even if the encoder has access to noiseless (even non-causal) output feedback.[4] A fortiori, (3.4) holds in the case of partial or noisy feedback, which is the typical situation occurring in the network scenarios outlined in Sect. 1. In case of perfect causal feedback, the bound (3.4) is achieved by the encoder which keeps on transmitting the most significant bit of the dyadic expansion (see Sect. 4) of x until this is correctly received. However, it is not clear what can be done if the feedback is noisy, partial, or not available (as in the applications outlined in Sect. 1). In Sect. 5.1 we shall propose some simple schemes which are not able to achieve exponential error rates, but have low computational complexity, while in Sect. 6 we shall present schemes achieving exponential error rates at the cost of higher computational complexity.

4. Quantized encoding schemes. In this paper, we shall propose and compare different coding strategies. All of them are based on a separation between the quantization of the continuous vector and the channel coding. In the literature, vector quantizers with special structure have been proposed, called tree-structured vector quantizers [17]. Consider a map $\mathcal{S} : \mathcal{X} \to \{0,1\}^{\mathbb{N}}$, and, for all $t \in \mathbb{N}$, the map $\mathcal{S}_t := \pi_t \circ \mathcal{S}$, where $\pi_t : \{0,1\}^{\mathbb{N}} \to \{0,1\}^t$ is the truncation operator defined above. Finally, let $\mathcal{S}_t^{-1}$ be a right inverse of $\mathcal{S}_t$. Then, we can define a tree-structured vector quantizer [17, pag.410] which is the family of maps $\mathcal{Q}_t : \mathcal{X} \to \mathcal{X}$ defined as $\mathcal{Q}_t := \mathcal{S}_t^{-1} \circ \mathcal{S}_t$. It can be seen [17] that if $\mathbb{E}||x||^{2+\delta} < +\infty$ for some $\delta > 0$, then, there exists a tree-structured vector quantizer $(\mathcal{Q}_t)$ such that

$$(\mathbb{E}||x - \mathcal{Q}_t(x)||^2)^{1/2} \leq C_+ 2^{-t/d}\,, \tag{4.1}$$

where C_+ is a positive constant depending only on the dimension d and the a priori density $f(x)$.

REMARK 2. *The upper bound (4.1) is easy to be obtained if $\mathcal{X} = [0,1]$. Indeed, in this case, one can take $\mathcal{S}$ to be the map which associates with x its binary expansion. We can apply this argument in case $\mathcal{X}$ is a bounded subset of $\mathbb{R}$. In case $\mathcal{X}$ is unbounded, tree-structured quantizers can be determined satisfying the upper bound (4.1) (see Lemma 5.2 in [30]). The extension from the scalar to the vector case is straightforward.*

Notice that, if $x', x'' \in \mathcal{X}$ are such that $\mathcal{S}_t(x') = \mathcal{S}_t(x'')$, then

$$\begin{aligned} \mathbb{E}||x' - x''||^2 &\leq \mathbb{E}(||x' - \mathcal{Q}_t(x')|| + ||x'' - \mathcal{Q}_t(x'')||)^2 \\ &\leq 2\mathbb{E}||x' - \mathcal{Q}_t(x')||^2 + 2\mathbb{E}||x'' - \mathcal{Q}_t(x'')||^2 \leq 2C_+^2 2^{-2t/d}\,. \end{aligned} \tag{4.2}$$

With a slight abuse of terminology, the map $\mathcal{S}$ associated with a tree-structured vector quantizer will be called a dyadic expansion map. We now show how a transmission scheme can be built starting from $\mathcal{S}$ and a family of its truncations' right inverses $\mathcal{S}_t^{-1}$.

[4] In fact, it is tempting to conjecture that a tighter bound could possibly be proven for the exponent in the absence of feedback.

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

$$\tilde{E}_t : \mathcal{Y}^{m_t} \to \mathcal{Y}\,, \qquad \tilde{\mathcal{D}}_t : \mathcal{Z}^t \to \mathcal{Y}^{m_t}\,. \tag{4.3}$$

We can define the map $\tilde{\mathcal{E}} : \mathcal{Y}^{\mathbb{N}} \to \mathcal{Y}^{\mathbb{N}}$ by letting the value of $\tilde{\mathcal{E}}((w_s)_{s=1}^{\infty})$ at time t equal to $\tilde{E}_t(w_1, \ldots, w_{m_t})$. 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

$$\begin{array}{ccccccccccc} \mathcal{X} & \xrightarrow{\mathcal{S}_{m_t}} & \mathcal{Y}^{m_t} & \xrightarrow{\tilde{\mathcal{E}}_t} & \mathcal{Y}^t & \xrightarrow{\text{Channel}} & \mathcal{Z}^t & \xrightarrow{\tilde{\mathcal{D}}_t} & \mathcal{Y}^{m_t} & \xrightarrow{\mathcal{S}_{m_t}^{-1}} & \mathcal{X} \\ x & \longmapsto & (w_s)_{s=1}^{m_t} & \longmapsto & (y_s)_{s=1}^{t} & \longmapsto & (z_s)_{s=1}^{t} & \longmapsto & (\hat{w}_s(t))_{s=1}^{m_t} & \longmapsto & \hat{x}_t\,. \end{array} \tag{4.4}$$

In other words, in this scheme we first use the dyadic expansion map 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.

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 $1/2$.

5.1. A repetition coding scheme. Let $\mathcal{S}_t : \mathcal{X} \to \{0,1\}^t$ be the truncated dyadic expansion map introduced in Sect. 4, and let $\mathcal{S}_t^{-1} : \{0,1\}^t \to \mathcal{X}$ be one of its right inverses. If a coding scheme with $\mathcal{E} = \mathcal{S}$ were simply used, i.e. if the bits of the dyadic expansion were directly sent through the channel, 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. It is therefore 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 which is monotonically increasing in their significance.

The encoder we propose here is of the following type: at time t, the bit y_t to be sent through the channel coincides with w_{j_t}, the bit in position j_t of the dyadic expansion $\mathcal{S}(x)$. The encoder in this way will depend on the choice of j_t and fits in the scheme proposed in (4.3) simply by taking $m_t := \max\{j_1, j_2, \ldots, j_t\}$.

In the scheme we propose j_t is selected as follows. Fix a positive real q and define $\tau_0 = 0$ and $\tau_k = \lceil q \rceil + \lceil 2q \rceil + \cdots + \lceil kq \rceil$ for $k \in \mathbb{N}$. Notice that, for any $t \in \mathbb{N}$, there exists a unique k such that $\tau_{k-1} + 1 \leq t \leq \tau_k$. Then, define $j_t := t - \tau_{k-1}$. In other words, we have

$$(y_s)_{s=1}^{\infty} = \tilde{\mathcal{E}}((w_s)_{s=1}^{\infty}) = (w_1, w_2, \ldots, w_{\lceil q \rceil}, w_1, w_2, \ldots, w_{\lceil 2q \rceil}, w_1, w_2, \ldots, w_{\lceil 3q \rceil}, \ldots\ldots)\,. \tag{5.1}$$

In any scheme of this kind the decoding is elementary. The output of the decoder $(\hat{w}_j(t))_{j=1}^{m_t} \in \{0,1\}^{m_t}$ may be given by

$$\hat{w}_j(t) = \begin{cases} z_s & \text{if } \exists s \leq t \text{ such that } j(s) = j \text{ and } z_s \neq ? \\ 0 & \text{otherwise}\,. \end{cases}$$

Notice that this decoding scheme has complexity growing linearly in t. Indeed, it admits the following natural recursive implementation. First, initialize $\hat{w}_j(0) = 0$ for all $j = 0$. Then, for all $t \geq 0$, upon receiving z_{t+1} we compute $(\hat{w}_j(t+1))_{j=1}^{m_{t+1}}$ as

$$\hat{w}_j(t+1) = \begin{cases} z_{t+1} & \text{if } j = j(t+1) \text{ and } z_{t+1} \neq ? \\ \hat{w}_j(t) & \text{otherwise} \end{cases} . \tag{5.2}$$

PROPOSITION 5.1. *Consider the repetition coding scheme defined by (5.1) and (5.2) on the BEC with erasure probability ε. Then, the root mean squared error satisfies*

$$\Delta_t \leq p(t) 2^{-\beta t^{1/2}}\,, \tag{5.3}$$

where

$$\begin{aligned} &\beta = \tfrac{\sqrt{2q}}{d}, && p(t) = C_1 && \text{if } q < \tfrac{d \log \varepsilon^{-1}}{2} \\ &\beta = \tfrac{\sqrt{2q}}{d}, && p(t) = C_2\sqrt{t} && \text{if } q = \tfrac{d \log \varepsilon^{-1}}{2} \\ &\beta = \tfrac{\log \varepsilon^{-1}}{\sqrt{2q}}, && p(t) = C_3 && \text{if } q > \tfrac{d \log \varepsilon^{-1}}{2} \end{aligned}$$

with C_1, C_2, C_3 a positive constants depending only on q, ε and d.

Proof. Let us fix some $t \in \mathbb{N}$. Define $\upsilon_j := |\{1 \leq \tau \leq t \mid j(\tau) = j\}|$, and observe that $\upsilon_j = 0$ if $j > m_t$. For $j = 0, 1, \ldots, m_t$, consider the events $A_j := \{\hat{w}_1(t) = w_1, \ldots, \hat{w}_j(t) = w_j,\}$, $B_j := A_j \setminus A_{j+1}$. Notice that the events B_j are pairwise disjoint, and $\mathbb{P}\left(\bigcup_{j=0}^{m_t-1} B_j \bigcup A_{m_t}\right) = 1$. Moreover, observe that $\mathbb{P}(B_j) \leq \mathbb{P}(\hat{w}_{j+1}(t) \neq w_{j+1}) \leq \varepsilon^{\upsilon_{j+1}}$.

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 (4.2), $\mathbb{E}[||x - \hat{x}_t||^2 \mid A_j^t] \leq 2C_+^2 2^{-2j/d}$. Now for simplicity we assume that $t = \tau_k$ for some k. In this case we have that $m_t = \lceil kq \rceil$. Moreover we have that $\upsilon_j = k + 1 - \min\{h : \lceil qh \rceil \geq j\}$. Observe now that, since $j \leq \lceil x \rceil$ if and only if $j < x + 1$, we can argue that

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

This implies that, for $j = 0, 1, \ldots, \lceil kq \rceil$, we have $\upsilon_j = k - \left\lfloor \frac{j-1}{q} \right\rfloor$. Therefore,

$$\begin{aligned} \Delta_t^2 &= \textstyle\sum_{j=0}^{\lceil kq \rceil - 1} \mathbb{E}\left[||x - \hat{x}_t||^2 \,|\, B_j\right] \mathbb{P}(B_j) + \mathbb{E}[||x - \hat{x}_t||^2 \,|\, A_{\lceil kq \rceil}] \mathbb{P}(A_{\lceil kq \rceil}) \\ &\leq 2C_+^2 \left[\textstyle\sum_{j=0}^{\lceil kq \rceil - 1} 2^{-2j/d} \varepsilon^{\upsilon_{j+1}} + 2^{-2\lceil kq \rceil / d}\right] \\ &\leq 2C_+^2 \left[\textstyle\sum_{j=0}^{\lceil kq \rceil - 1} 2^{-2j/d} \varepsilon^{k - \lfloor j/q \rfloor} + 2^{-2\lceil kq \rceil / d}\right] \\ &\leq 2C_+^2 \left[\textstyle\sum_{j=0}^{\lceil kq \rceil - 1} 2^{-2j/d} \varepsilon^{k - j/q} + 2^{-2qk/d}\right]. \end{aligned}$$

Since $t = \tau_k = \sum_{j=1}^{k} \lceil jq \rceil \leq \sum_{j=1}^{k} (qj+1) = \frac{q}{2}k^2 + \frac{q+2}{2}k$, we have $k \geq \sqrt{\frac{2t}{q}} - \frac{q+2}{2q}$. Then, the claim easily follows from a straightforward computation. ∎

REMARK 3. *Notice that Proposition 5.1 implies that a convergence degree $\alpha = 1/2$ is achievable for any choice of the positive parameter q, without any knowledge of the value of the erasure probability $\varepsilon \in [0, 1[$. If one knows ε, then it is possible to optimize the convergence rate β by choosing $q = \frac{d \log \varepsilon^{-1}}{2}$.*

5.2. A trade-off result between performance and complexity. We shall now show how complexity limitations imply lower bounds to the error decay stronger than Theorem 3.1. In particular we shall prove that, for certain class of encoders (finite-window and finite-state automata), exponential decay of error can never be achieved. As before, let us assume that $\mathcal{S} : \mathcal{X} \to \{0,1\}^{\mathbb{N}}$ is the dyadic expansion map introduced in Sect. 4, and consider encoders $\tilde{\mathcal{E}} : \{0,1\}^{\mathbb{N}} \to \{0,1\}^{\mathbb{N}}$ of the form $\tilde{\mathcal{E}}((w_s)_{s=1}^{\infty})_t = \tilde{E}_t(w_1, \ldots, w_{m_t})$, for some finite integer m_t, and a map $\tilde{E}_t : \mathcal{Y}^{m_t} \to \mathcal{Y}$.

In general, $\tilde{E}_t$ may actually depend on a proper subset of the m_t bits $\{1, 2, \ldots, m_t\}$. Consider the minimal $\Theta_t \subseteq \{1, 2, \ldots, m_t\}$ which allows one to write

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

for a suitable function $f_t : \Theta_t \to \{0,1\}$. Let $n_t = |\Theta_t|$. The encoder $\tilde{\mathcal{E}}$ is called finite-window if n_t is bounded in t. With each encoder it is possible to associate, for every $j, t \in \mathbb{N}$, the quantity $\omega_j(t) := \sum_{1 \leq s \leq t} \mathbb{1}_{\Theta_s}(j)$, 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 is related to the complexity of the encoder $\tilde{\mathcal{E}}$. If the maps f_t are $\mathbb{Z}_2$-linear and separately computed, then χ_t provides an upper bound to the number of binary operations implemented by the encoder up to time t. However, there could be hidden recursive links among the f_t capable to lower the real computational complexity. In any case, for brevity, we shall refer to χ_t as the complexity function of the encoder. The following is our main result, relating the root mean squared error Δ_t to the complexity function χ_t.

THEOREM 5.2. *For any transmission scheme for the BEC, with erasure probability ε, consisting of an encoder with complexity function χ_t, it holds*

$$\Delta_t \geq C \; 2^{-\sqrt{\frac{1}{d}\chi_t \log \varepsilon^{-1}}} \,, \tag{5.4}$$

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 implies that $\Delta_t^2 \geq C_1 \sup_{j \in \mathbb{N}} \left\{ 2^{-2j/d} \; \varepsilon^{\omega_j(t)} \right\}$, for some constant $C_1 > 0$ only depending on d and f, independently from the way the decoders are chosen. It will prove convenient to consider the looser bounds

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

where $\psi_s(\omega_1, \ldots, \omega_s) := \frac{1}{s} \sum_{j=1}^{s} 2^{-2j/d} \; \varepsilon^{\omega_j}$. Hence, for every possible s,

$$\Delta_t^2 \geq C_1 \inf\{\psi_s(\omega_1, \ldots, \omega_s) : \, \omega \in M_s\} \tag{5.5}$$

where $M_s := \left\{\omega_1, \ldots, \omega_s \in (\mathbb{R}^+)^s \,:\, \sum_j \omega_j = \chi_t\right\}$. Since the function ψ_s is strictly convex, it admits a unique minimum on the convex compact set M_s. Using Lagrange multipliers, the unique stationary point (ω_j^*) of $\psi_s(\omega_1, \ldots, \omega_s)$ on the hyperplane M_s has to satisfy, for all $j \leq s$, $\omega_j^* = \varsigma - \rho j$, where, $\rho := \frac{\ln 4}{d \ln \varepsilon^{-1}} = \frac{2}{d \log \varepsilon^{-1}} > 0$, 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

$$\Delta_t^2 \geq C_1 \inf_{\omega \in M_{s^*}} \psi_{s^*}(\omega_1, \ldots, \omega_s) = \psi_{s^*}(\omega_1^*, \ldots, \omega_s^*) = C_1 e^{-\varsigma \ln \varepsilon^{-1}} \,. \tag{5.6}$$

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 (5.6), the claim follows. ■

We have the following straightforward consequence for finite-window encoders which show that the degree $\alpha = 1/2$ can not be beaten.

COROLLARY 5.3. *For any transmission scheme for the BEC, with erasure probability ε, consisting of a finite-window encoder with $n_t \leq n_{\max}$ for every t, it holds*

$$\Delta_t \geq C \; 2^{-\beta t^{1/2}} \,, \tag{5.7}$$

where $\beta = \sqrt{\frac{n_{\max} \log \varepsilon^{-1}}{d}}$ and where $C > 0$ is a constant depending only on d, the erasure probability ε and the density function f of the random vector x.

REMARK 4. *In the case of the repetition encoders treated in Sect. 5.1, we have that $n_{\max} = 1$. If we compare (5.7) with (5.3), we have thus established that among the repetition schemes ($n_{\max} = 1$), the example treated in Sect. 5.1 is optimal from the point of view of the asymptotic performance (both degree and rate of convergence).*

The bound (5.4) implies that, in order to obtain exponential convergence of the error, χ_t needs to grow at least quadratically in t or, equivalently, 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 at least linearly in t. Indeed, as we shall see, the random linear codes proposed in Sect. 6 have exactly this property. However, observe that this does not imply that linear-time encodable schemes cannot attain exponential error decays in any case, since χ_t is, as already noticed, only an upper bound to the complexity of the encoder, intended as the minimum number of operations required by any implementation of the encoder. A possibility would be to consider maps f_t which, despite being not finite-window, can still be computed with bounded complexity in some recursive way. The most obvious choice would be to consider finite-state automata schemes. Unfortunately, such schemes yield very poor performance, as it will be shown in the next subsection. A less simple choice (and which will not pursued here) would be to consider encoders obtained as serial concatenations of finite-window with finite-state automata schemes.

5.2.1. Finite-state automata encoders. Encoders which can be implemented as finite state automata yield very poor performance. In fact, the root mean squared error Δ_t in this case does not converge to 0 as $t \to +\infty$. Indeed, assume we are given a finite state alphabet A and two maps $\xi : A \times \{0,1\} \to A$, $\rho : A \times \{0,1\} \to \{0,1\}$. Moreover, fix an initial state $a^* \in A$. To the quadruple (A, ξ, ρ, a^*) we can naturally associate an encoder $\tilde{\mathcal{E}}$ as follows. Given $(w_s)_{s=1}^{\infty} \in \{0,1\}^{\mathbb{N}}$, recursively define $(y_s)_{s=1}^{\infty} = \tilde{\mathcal{E}}((w_s)_{s=1}^{\infty})$ by

$$\left\{ \begin{array}{rcll} a_{t+1} & = & \xi(a_t, w_t) & a_0 = a^* \\ y_t & = & \rho(a_t, w_t) & \end{array} \right.$$

Notice that the state updating map ξ together with the initial condition $a_0 = a^*$ yield a sequence of maps $\xi^{(t)} : \{0,1\}^t \to A$ such that $a_{t+1} = \xi^{(t)}(w_1, \dots w_t)$. If we choose $t = t_0$ in such a way that $2^{t_0} > |A|$, the map $\xi^{(t_0)}$ is necessarily 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 $A = \{w_k = w'_k,\ z_k =? \text{ for } k = 1, \dots, t_0\}$. Clearly, conditioned on A, the decoder, for any $t \geq t_0$, will decode uncorrectly at least one information bit in the first t_0 position with positive probability independent from t. Hence, $\Delta_t^2 \geq \mathbb{E}\left[||x - \hat{x}_t||^2 \,|\, A\right] \mathbb{P}(A) \geq 2^{-2t_0/d} \mathbb{P}(A)$.

6. A coding scheme with exponential error rates. The goal of this section is to show that, removing the complexity bounds, exponential convergence can be achieved. The proposed scheme will require quadratic computational complexity at the encoder and cubic complexity at the decoder.

We shall use random coding arguments employing linear tree codes over the binary field $\mathbb{Z}_2$. These arguments were first developed in the context of convolutional codes [42, 43, 15], and recently applied in the framework of anytime information theory [34, 36]. 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 [34, Th.5.1] and [36, 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 Sect. 6.3), and improving the achievable error rate for a significant range of values of ε (see Theorem 6.3 and Remark 5).

6.1. 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 $j > Ri$ (i.e. 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

$$\tilde{\mathcal{E}}_t : \mathbb{Z}_2^{m_t} \to \mathbb{Z}_2^t\,, \qquad \tilde{\mathcal{E}}_t\left((w_s)_{s=1}^{m_t}\right) := \pi_t(\phi \mathbf{w})\,, \tag{6.1}$$

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 (6.1) 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 the

dyadic expansion map introduced in Sect. 4, and define, as usual, the encoding scheme $\mathcal{E} : \mathcal{X} \to \mathbb{Z}_2^{\mathbb{N}}$ as the composition $\mathcal{E} = \tilde{\mathcal{E}} \circ \mathcal{S}$.

For the decoding part, we shall consider maximum a posteriori decoders $\tilde{\mathcal{D}}_t$. For the special case of the BEC, given the channel outputs z_t, the decoded block at time t, $(\hat{w}_s(t))_{s=1}^{m_t} = \tilde{\mathcal{D}}_t((z_s)_{s=1}^t)$, is defined to be any vector in $\{0,1\}^{m_t}$ which is compatible with the observed channel output $(z_s)_{s=1}^t$. Formally, let $\Xi_t := \{s \in \{1,\dots,t\} : z_s \neq ?\}$ be the set of non-erased positions up to time t, and $\pi_{\Xi_t} : \mathbb{Z}_2^t \to \mathbb{Z}_2^{\Xi_t}$ be the canonical projection. Then, a MAP decoder $\mathcal{D}_t : \{0,1,?\}^t \to \mathbb{Z}_2^{m_t}$ maps the channel output $(z_s)_{s=1}^t$ into any binary string $(\hat{w}_s(t))_{s=1}^{m_t}$ such that

$$\pi_{\Xi_t}\tilde{\mathcal{E}}_t((\hat{w}_s(t))_{s=1}^{m_t}) = \pi_{\Xi_t}(z_s)_{s=1}^t = \pi_{\Xi_t}\tilde{\mathcal{E}}_t((w_s)_{s=1}^{m_t})\,. \tag{6.2}$$

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

6.2. Performance analysis. We now analyze the coding scheme we have introduced. Notice first of all that, 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 $\{\delta_1, \delta_2, \dots, \delta_{m_t}\}$ be the canonical basis of $\mathbb{Z}_2^{m_t}$, and, for $0 \leq j \leq m_t$, consider the subspace[5] $K_j := \mathrm{span}(\delta_{j+1}, \dots, \delta_{m_t}) \subseteq \mathbb{Z}_2^{m_t}$. For $0 \leq j \leq m_t$, define the event $A_j := \{\ker(\pi_{\Xi_t}\tilde{\mathcal{E}}_t) \subseteq K_j\}$. Also, let us define $B_j := A_{j-1} \setminus A_j$, for $1 \leq j \leq m_t$. 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

$$\Omega = \bigcup_{1 \leq j \leq m_t} B_j \bigcup A_{m_t}\,. \tag{6.3}$$

Notice now that, from (6.2) we can deduce that $(w_s - \hat{w}_s(t))_{s=1}^{m_t} \in \ker \pi_{\Xi_t}\tilde{\mathcal{E}}_t$. 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. We immediately get from (4.2) that, if A_j occurs, then

$$||\hat{x}_t - x||^2 \leq 4d2^{-2j/d}\,, \qquad 0 \leq j \leq m_t\,. \tag{6.4}$$

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

$$\underline{\beta}'(d,\varepsilon,R) := \min\{\tfrac{1}{d}R, \tfrac{1}{2}\min_{0 \leq \eta \leq 1} D(\eta||1-\varepsilon) + \lfloor \eta - R \rfloor_+\}\,, \tag{6.5}$$

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

THEOREM 6.1. *Assume transmission over the BEC. Then, for all $0 < R < 1$, the average estimation error of the above-described random coding scheme satisfies*

$$\left(\mathbb{E}||x - \hat{x}_t||^2\right)^{1/2} \leq C\sqrt{t}\,2^{-\beta'(d,\varepsilon,R)\,t} \tag{6.6}$$

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

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

Proof. Using (6.3) and (6.4), we obtain

$$\begin{aligned}\mathbb{E}\left[||\hat{x}_t - x||^2\right] &= \sum_{j=1}^{m_t} \mathbb{E}\left[||\hat{x}_t - x||^2 |\, B_j\right] \mathbb{P}(B_j) + \mathbb{E}\left[||\hat{x}_t - x||^2 |\, A_{m_t}\right] \mathbb{P}(A_{m_t}) \\ &\leq \sum_{j=1}^{m_t} \mathbb{P}(B_j) 4d 2^{-2(j-1)/d} + 4d 2^{-2m_t/d}\,.\end{aligned} \tag{6.7}$$

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 $\pi_{\Xi_t}\tilde{\mathcal{E}}_t K_j$. Indeed, $\overline{A_j}$ implies that there exists some $v \in \mathbb{Z}_2^{m_t}$ such that $\pi_{\Xi_t}\tilde{\mathcal{E}}_t v = 0$, and $v_i \neq 0$ for some $i \geq j$. On the other hand, A_{j-1} implies that such a v has $v_i = 0$ for all $i < j$. Hence, if we define $v' \in \mathbb{Z}_2^{m_t}$ by $v'_i = 0$ for $i \leq j$, and $v'_i = v_i$ for $i > j$, we have that $v' \in K_j$ and $0 = \pi_{\Xi_t}\tilde{\mathcal{E}}_t v = \pi_{\Xi_t}\tilde{\mathcal{E}}_t\delta_j + \pi_{\Xi_t}\tilde{\mathcal{E}}_t v'$. Therefore $\pi_{\Xi_t}\tilde{\mathcal{E}}_t\delta_j = \pi_{\Xi_t}\tilde{\mathcal{E}}_t v' \in \pi_{\Xi_t}\tilde{\mathcal{E}}_t K_j$, as claimed.

Observe that $\tilde{\mathcal{E}}_t\delta_j$ is uniformly distributed over $H_j := \mathrm{span}(\delta_{\lceil j/R\rceil}, \ldots, \delta_t) \subseteq \mathbb{Z}_2^t$, and independent from $\lambda^t_{\lceil j/R\rceil}$ (the latter being defined in (3.2)). It follows that $\pi_{\Xi_t}\tilde{\mathcal{E}}_t\delta_j$ takes any value in $\pi_{\Xi_t}H_j$ with probability $2^{-\lambda^t_{\lceil j/R\rceil}}$. Since $|\pi_{\Xi_t}\tilde{\mathcal{E}}_t K_j| \leq |K_j| = 2^{m_t - j}$, we have that, for every $k = 0, \ldots, t - \lceil j/R\rceil + 1$,

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

From (3.3) it follows that

$$\begin{aligned}\mathbb{P}(B_j) &= \sum_{k=0}^{t-\lceil j/R\rceil+1} P(B_j | \lambda^t_{\lceil j/R\rceil} = k) \mathbb{P}(\lambda^t_{\lceil j/R\rceil} = k) \\ &\leq \sum_{k=0}^{t-\lceil j/R\rceil+1} 2^{-\lfloor j+k-m_t\rfloor_+} \tbinom{t-\lceil j/R\rceil+1}{k} \varepsilon^{t-\lceil j/R\rceil+1-k}(1-\varepsilon)^k \\ &\leq 2 \sum_{k=0}^{t-\lceil j/R\rceil+1} 2^{-\lfloor j+k-m_t+1\rfloor_+} 2^{-(t-\lceil j/R\rceil+1) D(\frac{k}{t-\lceil j/R\rceil+1}||1-\varepsilon)} \\ &\leq 2t 2^{-(t-j/R)\min_{0\leq\eta\leq 1} D(\eta||1-\varepsilon) + \lfloor \eta - R\rfloor_+\}}\end{aligned} \tag{6.8}$$

where the second inequality follows from standard estimations of the binomial coefficient (see e.g. [11]), and the last one by setting $\eta = \frac{k}{t-\lceil j/R\rceil+1}$, and the estimates $R \leq 1$, $m_t \leq Rt$, $j/R \leq \lceil j/R\rceil \leq j/R + 1$. Finally, (6.6) follows by substituting (6.8) into (6.7). ∎

Standard probabilistic arguments allow one to prove the following corollary of Theorem 6.1, characterizing the exponential error rate of a typical realization of the random coding scheme $(\mathcal{E}, \mathcal{D})$. Observe that the root mean squared 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, hence a random variable.

Corollary 6.2. *Assume transmission over the BEC with erasure probability ε. Then, for all $0 < R < 1$, with probability one,*

$$\left(\mathbb{E}[||x - \hat{x}_t||^2 |\phi]\right)^{1/2} \leq C t^{3/2} 2^{-\underline{\beta}'(d,\varepsilon,R)\, t}\,, \tag{6.9}$$

for a positive constant C.

Proof. For $t \in \mathbb{N}$, consider the event $A_t := \left\{\mathbb{E}[||x - \hat{x}_t||^2 |\phi] \geq t^3 2^{-2t(\underline{\beta}'(d,\varepsilon,R))}\right\}$. Form Markov's inequality and Theorem 6.1, it follows that

$$\mathbb{P}[A_t] \leq t^{-3} 2^{2t\underline{\beta}'(d,\varepsilon,R)} \mathbb{E}[||x - \hat{x}_t||^2] \leq C t^{-2}\,,$$

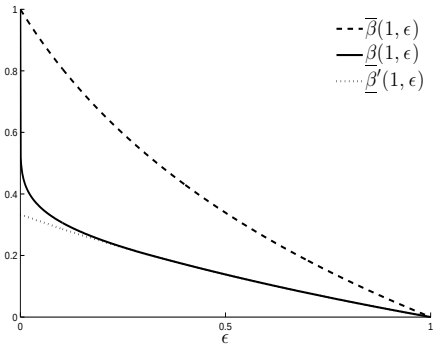

FIG. 6.1. *Upper and lower bounds to the achievable estimation error exponent achievable on the BEC (as defined in (3.5), (6.11) and Remark 5, respectively) are plotted as a function of the erasure probability ε for $d = 1$.*

so that the series $\sum_t \mathbb{P}[A_t]$ is convergent and Borel-Cantelli lemma implies that, with probability one, A_t occurs only for finitely many values of $t \in \mathbb{N}$. Then, the claim easily follows. ■

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 6.2 for certain values of R and ε. For every $0 \leq R \leq 1$, define $\gamma(R) := \min\{x \in [0,1] : \mathrm{H}(x) \geq 1 - R\}$, and

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

The following result is proved in Appendix A.

THEOREM 6.3. *Assume transmission over the BEC with erasure probability ε. Then, for all $0 < R < 1$, $\delta > 0$, with probability one*

$$\left(\mathbb{E}[||x - \hat{x}_t||^2|\phi]\right)^{1/2} \leq Kt2^{-(\underline{\beta}''(d,\varepsilon,R)+\delta)t}, \tag{6.10}$$

for a constant $K > 0$.

REMARK 5. *It follows from Corollary 6.2 and Theorem 6.3 that, for all $R < 1-\varepsilon$ random causal linear codes achieve exponential convergence rate. Optimizing over $R \in]0, 1-\varepsilon[$, this shows that the exponent*

$$\underline{\beta}(d, \varepsilon) := \max_{0\leq R\leq 1} \max\{\underline{\beta}'(d, \varepsilon, R), \underline{\beta}''(d, \varepsilon, R)\}, \tag{6.11}$$

is achievable. In Fig.6.1 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 $d = 1$. Define $\underline{\beta}'(d, \varepsilon) := \max\{\underline{\beta}'(d, \varepsilon, R) : R \in [0, 1]\}$, and $\underline{\beta}''(d, \varepsilon) := \max\{\underline{\beta}'(d, \varepsilon, R) : R \in [0, 1]\}$. Then, it is not difficult to see that $\lim_{\varepsilon\downarrow 0} \underline{\beta}'(d, \varepsilon) = 1/(d + 2)$, while $\lim_{\varepsilon\downarrow 0} \underline{\beta}''(d, \varepsilon) = 1/d$. Hence, Theorem 6.3 becomes particularly relevant for small erasure probabilities, showing that the noiseless error exponent $1/d$ (see Sect. 4) is recovered in the limit of vanishing noise: this does not follow from the average-code analysis of Theorem 3.1.

6.3. 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, $\chi_t := \sum_{s \le t} n_s$, 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 $\frac{1}{2} R^2 t^2$, while the strong law of large numbers implies that the typical encoder complexity χ_t is such that $\chi_t / \frac{1}{4} R^2 t^2$ 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

$$\pi_{\Xi_t} \tilde{\mathcal{E}}_t((w_s)_{s=1}^{m_t}) = \pi_{\Xi_t}(z_s)_{s=1}^{t} \,. \tag{6.12}$$

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 (6.12) with number of operations $o(t^3)$: see [41, pagg.247-248] for the analogous problem for linear systems over the reals. However, the system (6.12) 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 $\Theta(t^2)$.

7. 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 [21][26, 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. As usual, assume that we have a dyadic expansion $\mathcal{S}$ mapping 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, i.e. $|\Theta_t| \le n_{\max}$.

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

according 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 $(n_s)_{s\le t}$. Then, for evert $t \ge 1$, consider the random set $\Theta_t := \{\theta_{1,t}, \theta_{2,t}, \dots, \theta_{n_t,t}\}$, where $\theta_{i,t}$ are independent random variables unformly distributed according to $\nu_t(\,\cdot\,)$. Notice that in this way we have that $|\Theta_t| \le n_t \le n_{\max}$ 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 [21][26, Ch.50]. Such an algorithm works on an iteratively updated infinite hypergraph[6] $\mathcal{G}_t = (\mathcal{V}_t, \mathcal{H}_t)$ 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 $t \ge 1$, 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 $\chi_t = \sum_{s\le t} n_s$ operations up to time t, hence it has linear complexity in t. It is suboptimal with respect to the maximum a posteriori decoding: 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.

In Fig.7.1 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 soliton one [26, pag.592]

$$\mu(1) := \frac{1}{n_{\max}}\,, \qquad \mu(n) := \frac{1}{n(n-1)} \quad \forall\, 2 \le n \le n_{\max}\,. \tag{7.1}$$

The distributions ν_t have been selected as follows. We define $\rho := 2\left(d \log \varepsilon^{-1}\right)^{-1}$, $s_t := \lfloor \sqrt{2\chi_t \rho^{-1}} \rfloor$, and $\varsigma_t = \frac{\chi_t}{s_t} + \rho\frac{s_t+1}{2}$, where $\chi_t = \sum_{s\le t} n_s$. Then choose

$$\nu_t(j) := \begin{cases} \eta(\varsigma_t - \rho j) & \text{if} \quad j \le s_t \\ 0 & \text{if} \quad j > s_t\,, \end{cases} \tag{7.2}$$

Our choice was suggested by the optimization problem in the right-hand side of (5.5).

It is clear from Fig.7.1(a) that the three schemes have subexponential error decay and that increasing the degree allows one to obtain better convergence rates. Fig.7.1(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 5.2, 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] 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 $\mathcal{V}$.

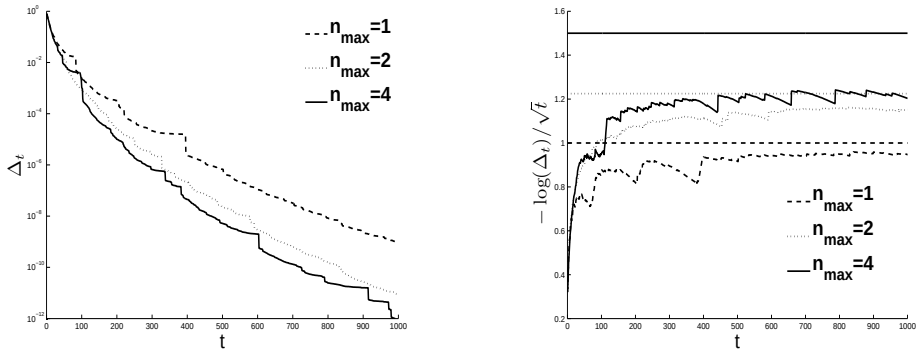

FIG. 7.1. *Monte Carlo simulations of finite-window coding schemes on the BEC, with erasure probability* $\varepsilon = 0.5$. *The performance of three coding schemes are compared: these schemes were randomly generated accordingly to (7.1) and (7.2) with* $n_{\max} = 1, 2, 4$ *respectively. In (a) the root 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 5.2. The number of samples used is 200000.*

8. Conclusion and extensions. 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 has been derived. Such a bound shows that, for the mean squared error to decrease exponentially fast in the number of channel uses, the subsequent channel inputs have depend on a linearly growing number of information bits, , i.e. bits of the dyadic expansion of the source vector. Finally, simulation results for linear-complexity coding/decoding schemes have been proposed.

Using finer information-theoretic arguments, most of the results of this paper can be extended to more general discrete memoryless channels. In particular, Theorem 3.1 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 [16, 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: see [7] for similar arguments in the more general context of distributed computation over networks. Theorem 5.2 can be extended to general discrete memoryless channels, showing that exponential error decays require that the t-th channel input depends at least on a linear number of information bits. Using arguments as in [43], Theorem 6.3 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 [16]. It is possible to extend Theorem 6.1 to arbitrary discrete memoryless channels, using a random coset approach possibly followed by a quantization as in [16, 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 [16]. 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 [15, 34, 36]. 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, to our knowledge, no result analogous to Theorem 6.3 is known to hold for non-linear random convolutional codes.

Some 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 3.1, by better exploiting the absence of feedback.

Appendix A. Proof of Theorem 6.3. We shall prove Theorem 6.3 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 $0 \le j \le m_t$, and $h \ge 0$, let us consider the random variable $\Upsilon_j^t(h) := |\{\boldsymbol{y} \in K_j \setminus K_{j+1} : w_H(\tilde{\mathcal{E}}_t\boldsymbol{y}) = h\}|$, counting 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. Observe that the causality of ϕ implies that, if $\boldsymbol{y} \in K_j$, then $\tilde{\mathcal{E}}_t\boldsymbol{y}$ belongs to $L_j := \mathrm{span}\,(\delta_s | \lceil (j+1)/R \rceil \le s \le t) \subseteq \mathbb{Z}_2^t$. 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 uniformly distributed over L_j. It follows that

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

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 $F_t := \bigcup_{j=1}^{(1-\lambda)Rt} \bigcup_{h=0}^{l_j\gamma(R+\varphi)} \{\Upsilon_j^t(h) \ge 1\}$ by $\mathbb{P}(F_t) \le \sum_{j,h} \mathbb{E}[\Upsilon_j^t(h)] \le t^2 2^{-t\lambda\varphi}$. Then, the series $\sum_n \mathbb{P}(F_n)$ 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 $\Upsilon_j^t(h) = 0$ for all $h < l_j\gamma(R+\varphi)$, for all $t \ge t_0$ and $1 \le j \le (1-\lambda)Rt$. An analogous argument shows that with probability one $\Upsilon_j^t(h) \le 2^{l_j(\mathrm{H}(\eta)-1+R+\varphi))}$, for all $l_j\gamma(R+\varphi) \le h \le l_j$, for sufficiently large t.

We are now ready to prove Theorem 6.3. 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

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

Then, for $j = 1, \ldots, \lfloor (1-\lambda)Rt \rfloor$, the union bound yields the estimation

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

Hence, (6.3) and (6.4) imply that, for $\kappa := 1-\lambda)Rt$

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

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

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

Finally, the claim 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. Inform. Theory, 24(1978), pp. 384-386.

[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, NY, 1998.

[4] S. Borade, B. Nakiboğlu, and L. Zheng, Unequal error protection: some fundamental limits, IEEE Trans. Inform. Theory, (2009), to appear.

[5] M.V. Burnashev, A new lower bound for the α-mean error of parameter transmission over the white Gaussian channel, IEEE Trans. Inform. Theory, 30(1984), pp. 23-34.

[6] R. Carli, G. Como, P. Frasca, and F. Garin, Average consensus over digital noisy networks, *Proc. of IFAC NecSys 2009*, 24-26 September, 2009, Venice (Italy).

[7] G. Como, and M. Dahleh, Lower bounds on the estimation error in problems of distributed computation, *Proc. of ITA Workshop*, Feb. 8-13, 2009, San Diego, CA, US, pp. 70-76.

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

[9] I. Csiszàr, Joint source-channel error exponent, Probl. Control Inf. Theory, 9(1980), pp. 315-328.

[10] I. Csiszàr, On the error exponent of source-channel transmission with a distortion threshold, IEEE Trans. Inform. Theory, 28(1982), pp. 823-828.

[11] I. Csiszàr, The method of types, IEEE Trans. Inform. Theory, 44(1998), pp. 2505-2523.

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

[13] R. Diekmann, A. Frommer, and B. Monien, Efficient schemes for nearest neighbor load balancing, Parallel computing, 25(1999), pp. 789-812.

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

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

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

[17] A. Gersho, and R.M. Gray, Vector quantization and signal compression, Kluver, 2001.

[18] 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, 2000, pp. 56-67.

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

[20] D. Kempe, A. Dobra, and J. Gehrke, Gossip-based computation of aggregate information, Proc. IEEE FOCS 2003, pp. 1-10.

[21] M. Luby, LT codes, in Proc. 43rd IEEE FOCS 2002, November 16-19, 2002, pp. 271-282.

[22] An analogue of Shannon information theory for networked control systems: State estimation via a noisy discrete channel, *Proc. of CDC 2004, Atlantis, Paradise Island (Bahamas), December 14-17, 2004, pp. 4485-4490.*

[23] 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, 80(2007), pp. 241-255.
[24] A.S. Matveev, State estimation via limited capacity noisy communication channels, Mathematics of Control, Signals, and Systems, 20(2008), pp. 1-35.
[25] A.S. Matveev and A.V. Savkin, *Estimation and control over communication networks*, Birkhauser, Boston, MA, 2009.
[26] D. McKay, Information theory, inference, and learning algorithms, Cambridge University Press, Cambridge, UK, 2003.
[27] B. Masnick and J. Wolf, On linear unequal error protection, IEEE Trans. Inf. Theory, 13(1967), pp. 600-607.
[28] L. Moreau, Stability of multiagent systems with time-dependent communication links, IEEE Trans. Automat. Control, 50(2005), pp. 169-182.
[29] S. Muthukrishnan, B. Ghosh, and M. Schultz, First and second order diffusive methods for rapid, coarse, distributed load balancing, Theory of computing systems, 31(1998), pp. 331-354.
[30] G.N. Nair, and N.J. Evans, Stabilizability of stochastic linear systems with finite feedback date rates, SIAM J. Control Optim., 43(2004), pp. 413–436, 2004.
[31] R. Olfati-Saber and R.M. Murray, Consensus problems in networks of agents with switching topology and time-delays, IEEE Trans. Automat. Control, 49(2004), pp. 1520-1533.
[32] W. Ren and R.W. Beard, Consensus seeking in multiagent systems under dynamically changing interaction topologies, IEEE Trans. Automat. Control, 50(2005), pp. 655-661.
[33] T.J. Richardson and R. Urbanke, Modern Coding Theory, Cambridge Univeristy Press, Cambridge, UK, 2007.
[34] A. Sahai and S. Mitter, The necessity and sufficiency of anytime capacity for stabilization of a linear system over a noisy communication link–Part I: scalar systems, IEEE Trans. Inform. Theory, 52(2006), pp. 3369-3395.
[35] A. Sahai, Why do block length and delay behave differently if feedback is present?, IEEE Trans. Inform. Theory, 54(2008), pp. 1860-1886.
[36] A. Sahai and S. Mitter, Source coding and channel requirements for unstable processes, submitted, 2006.
[37] T. Simsek, R. Jain, and P. Varaiya, Scalar estimation and control with noisy binary observation, IEEE Trans. Automat. Control, 49(2004), pp. 1598-1603.
[38] S. Tatikonda and S. Mitter, Control under communication constraints, IEEE Trans. Automat. Control, 49(2004), pp. 1056-1068.
[39] S. Tatikonda and S. Mitter, Control over noisy channels, IEEE Trans. Automat. Control, 49(2004), pp. 1196-1201.
[40] J. Tsitsiklis, Problems in decentralized decision making and computation, Ph.D. dissertation, Dep. Elec. Eng. Comput. Sci., Mass. Inst. Technol., Cambridge, MA, 1984.
[41] L.N. Treften and D. Bau, III, Numerical linear algebra, SIAM, Philadelphia, PA, 1997.
[42] A.J. Viterbi, Error bounds for convolutional codes and an asymptotically optimal decoding algorithm, IEEE Trans. Inform. Theory, 13(1967), pp. 260-269.
[43] A.J. Viterbi, Further results on optimal decoding of convolutional codes, IEEE Trans. Inform. Theory, 15(1969), pp. 732-734.
[44] J. Zhao, R. Govindan and D. Estrin, Computing aggregates for monitoring wireless sensor networks, Proc. SNPA 2003, pp. 139-148.
[45] S. Zilberstein, Using Anytime Algorithms in Intelligent Systems, AI Magazine, 17(1996), pp. 73-83.