Anytime reliable transmission of real-valued information through digital noisy channels

Giacomo Como, Fabio Fagnani and Sandro Zampieri

***Abstract*— Motivated by distributed sensor networks scenarios, we consider a problem of state estimation under communication constraints, in which a real-valued random vector needs to be reliably transmitted through a digital noisy channel. Estimations are sequentially updated by the receiver, as more and more channel outputs are observed. Assuming that no channel feedback is available at the transmitter, we study the rates at which the mean squared error of the estimation can be made to converge to zero with time. First, simple low-complexity schemes are considered, and trade-offs between performance and encoder/decoder complexity are found. Then, information-theoretic bounds on the best achievable error exponent are obtained.**

I. Introduction

The reliable transmission of information among the nodes of a network is well 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 tasks requiring information exchange. Important examples include wireless sensor networks, in which the final goal is estimation from distributed measurements, or wireless sensor and actuators networks (such as mobile multi-agent networks, in which the final goal is control. 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, one basic problem is thus to transmit a continuous quantity, namely a real number or, possibly, a real vector, through a digital noisy channel up to a certain degree of precision. An important case of channel which will be the one mostly considered in this paper is the so-called binary erasure channel (BEC), where a bit is either transmitted correctly or erased with some probability ε. This channel well models the situation of mobile agents which, depending on their position, may or may not be in the range of transmission of the other agents.

As especially pointed out in the papers by Sahai and Mitter [4], [5], [6], there is a specific feature distinguishing the problem of information transmission for control and estimation from the problem of pure information transmission. This is related to the different sensitivity to delays typically occurring in the two contexts. Indeed, while often the presence of sensible delays can be tolerated in the communication performance evaluation, this typically has disastrous effects in control, where the important question is not only where information is available, but rather where and when information is available. For this reason, the theory of digital communication for control cannot be completely reduced to the classical theory of digital communication. In fact, while the latter can be seen as the problem of efficient joint source and channel coding, in the former the time ingredient has to be considered with more attention.

More precisely, in the standard digital communication framework, data is requested to be available at the receiver only at the end of the transmission, after the completion of the decoding process. In contrast, for control we need a coding and decoding procedure able to produce a reasonable partial information transmission also in case we stop the process before the end. In other words, while in communication the performance is influenced only by the steady state behavior of the communication system, both the steady state behavior and the transient behavior play an important role in the control system performance. For all these reasons, in control we need a procedure which provides an estimate with increasing precision while the time passes by.

On the other hand, in control there is sometimes the possibility to take advantage of feedback for improving the coding and the decoding, as feedback information is in some cases naturally available to the encoder. In other cases this feedback information is difficult to be used or, as for instance in the wireless network scenario, there are situations in which the transmitter broadcasts 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 will consider the problem of estimation under communication constraints without feedback information.

A. Problem formulation

We now give the formal description of the problem we want to study. Let $\mathscr{X} \subseteq \mathbb{R}^d$ be an open subset where the vector to be transmitted is known to be confined, equipped with an a-priori probability density $f(x)$. The communication channel is assumed to have a finite input alphabet $\mathscr{Y}$ and a finite output alphabet $\mathscr{Z}$, and to be described by a family $\{p(\cdot|y)\}_{y\in\mathscr{Y}}$ of probability distributions over $\mathscr{Z}$: $p(z|y)$ is the probability that $z \in \mathscr{Z}$ is received, conditioned on the transmission of $y \in \mathscr{Y}$. We assume, for the sake of simplicity, that at every time instant t, one symbol can be

G. Como is with Laboratory for Information and Decision Systems, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, MA, USA, giacomo@mit.edu

F. Fagnani is with Dipartimento di Matematica, Politecnico di Torino, Corso Duca degli Abruzzi 24, 10129, Torino, Italy, fabio.fagnani@polito.it

S. Zampieri is with Dipartimento di Elettronica e Informatica, Università di Padova, via Gradenigo 6/A, 35131, Padova, Italy, zampi@dei.unipd.it

transmitted through the channel, and, moreover, that the channel is memoryless, namely, the output values of repeated transmissions are independent among each other. In the case of the BEC, we have that $\mathscr{Z} = \{0,1,?\}$ where ? stays for the erasure event, and

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

a) Coding scheme: Our transmission scheme consists of an encoder

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

and of a decoder

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

More precisely the decoder is defined by a family of maps

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

so that, for any $(z_s)_{s=1}^{\infty} \in \mathscr{Z}^{\mathbb{N}}$, the value at time t of $\mathscr{D}((z_s)_{s=1}^{\infty})$ is $\mathscr{D}_t((z_s)_{s=1}^{t})$. The overall sequence of maps is described by the following scheme

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

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

$$\hat{x}_t := \mathscr{D}_t(y_1^t)\,.$$

This is described by the following scheme

$$\begin{array}{ccccccc} \mathscr{X} & \xrightarrow{\ \mathscr{E}_t\ } & \mathscr{Y}^t & \xrightarrow{\text{Channel}} & \mathscr{Z}^t & \xrightarrow{\ \mathscr{D}_t\ } & \mathscr{X} \\ x & \longrightarrow & (y_s)_{s=1}^t & \longrightarrow & (z_s)_{s=1}^t & \longrightarrow & \hat{x}_t \end{array} \tag{2}$$

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

A possible particular case of the previous scheme is given below. Let $\mathscr{W}$ be a finite alphabet and consider a family of *surjective* maps

$$\mathscr{S}_t : \mathscr{X} \to \mathscr{W}^t$$

such that $\mathscr{S}_t = \pi_t \circ \mathscr{S}_{t+1}$. In other words we assume that, if $\mathscr{S}_{t+1}(x) = (w_1,\dots,w_t,w_{t+1})$, then we have that $\mathscr{S}_t(x) = (w_1,\dots,w_t)$. The family of these maps forms a map

$$\mathscr{S} : \mathscr{X} \to \mathscr{W}^{\mathbb{N}}$$

in a natural way. A simple example of this map is given by choosing $\mathscr{X} = [0,1]$, $\mathscr{W} = \{0,1,\dots,k-1\}$, and by assuming that $\mathscr{S}$ is the map which associates the k-ary expansion to a real number belonging to the interval $[0,1]$. Let

$$\mathscr{S}_t^{-1} : \mathscr{W}^t \to \mathscr{X}$$

be a left inverse of $\mathscr{S}_t$. Notice that $Q_t := \mathscr{S}_t^{-1} \circ \mathscr{S}_t$ maps $\mathscr{X}$ into $\mathscr{X}$. Since this function maps a continuous space into a finite set, it is called a quantizer.

Consider now a sequence of positive integers $m_1, m_2, \dots$, and two families of encoders and decoders

$$\tilde{\mathscr{E}}_t : \mathscr{W}^{l_t} \to \mathscr{Y}^t\,, \qquad \tilde{\mathscr{D}}_t : \mathscr{Z}^t \to \mathscr{W}^{l_t} \tag{3}$$

From $\tilde{\mathscr{E}}_t, \tilde{\mathscr{D}}_t$ we can define the encoders $\mathscr{E}_t := \tilde{\mathscr{E}}_t \circ \mathscr{S}_{m_t}$ and the decoders $\mathscr{D}_t := \mathscr{S}_{m_t}^{-1} \circ \tilde{\mathscr{D}}_t$. The overall sequence of maps is described by the following scheme

$$\begin{array}{ccccccccccc} \mathscr{X} & \stackrel{\mathscr{S}_{m_t}}{\to} & \mathscr{W}^{m_t} & \stackrel{\tilde{\mathscr{E}}_t}{\to} & \mathscr{Y}^t & \stackrel{\text{Ch.}}{\to} & \mathscr{Z}^t & \stackrel{\tilde{\mathscr{D}}_t}{\to} & \mathscr{W}^{m_t} & \stackrel{\mathscr{S}_{m_t}^{-1}}{\to} & \mathscr{X} \\ x & \mapsto & (w_s) & \mapsto & (y_s) & \mapsto & (z_s) & \mapsto & (\hat{w}_s(t)) & \mapsto & \hat{x}_t \end{array} \tag{4}$$

More specifically, in this scheme we first use a quantizer mapping x into a string of k-ary symbols $(w_1, w_2, \dots, w_{m_t})$ and then we use a block encoder. The received data are decoded by a block decoder providing an estimated version $(\hat{w}_1(t), \hat{w}_2(t), \dots, \hat{w}_{m_t}(t))$ of $(w_1, w_2, \dots, w_{m_t})$ (whose components in general depend on t) which is in turn translated to an estimate $\hat{x}_t$ of x. Notice that, since from the way we defined $\mathscr{E}_t$ we have that $\pi_{t-1} \circ \mathscr{E}_t = \mathscr{E}_{t-1}$, then we need to impose that

$$\pi_{t-1}\tilde{\mathscr{E}}_t((w_s)_{s=1}^{m_t}) = \tilde{\mathscr{E}}_{t-1}((w_s)_{s=1}^{m_{t-1}})$$

In other words, the family of encoders $\tilde{\mathscr{E}}_t$ forms in a natural way a map

$$\tilde{\mathscr{E}} : \mathscr{W}^{\mathbb{N}} \to \mathscr{Y}^{\mathbb{N}}$$

which is causal in the sense that the value a time t of $\tilde{\mathscr{E}}(w_1^{\infty})$ depends only on $w_1^{m_t}$, namely there exists a family of maps

$$E_t : \mathscr{W}^{m_t} \to \mathscr{Y}$$

such that the value of the sequence $\tilde{\mathscr{E}}(w_1^{\infty})$ at time t is given by $E_t(w_1^{m_t})$.

b) 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 $x \in \mathscr{X}$ and with respect to the possible randomness of the communication channel) at time t by

$$\Delta_t := (\mathbb{E}||x - \hat{x}_t||^2)^{1/2}\,. \tag{5}$$

In this paper we want to understand how fast Δ_t decreases as t tends to infinity. We will consider this problem first without imposing complexity constraints on the encoder and decoder algorithms, and then investigating how algorithmic implementability influences this error performance.

B. State estimation under communication constraints

The problem proposed above is related to the state estimation problem under communication constraints. Assume we have a discrete time stochastic linear system

$$x(t+1) \;=\; Ax(t) + v(t) \qquad x(0) = x_0 \tag{6}$$

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 $A \in \mathbb{R}^{n\times n}$.

We have to design a family of encoders E_t and of decoders $\mathscr{D}_t$. Each encoder E_t has $x(0), \dots, x(t)$ as input and the symbol $y(t) \in \mathscr{Y}$ as output, where $\mathscr{Y}$ is a finite alphabet.

Assume that $y(t)$ is transmitted through a channel, which provides the decoder $\mathscr{D}_t$ with the symbols $z(0),\dots,z(t)$ from which the decoder has to obtain an estimate $\hat{x}(t)$ of the current state. We distinguish to cases

1) In case the variance of $v(t)$ is big with respect to the variance of $x(0)$ or in case we are interested in the steady state performance, then the parameter
$$\limsup_{t\to\infty} \mathbb{E}[||x(t)-\hat{x}(t)||^2$$
is the most relevant parameter to be considered in designing the encoders E_t and the decoders $\mathscr{D}_t$.
2) In case the variance of $v(t)$ is small with respect to the variance of $x(0)$ and if we are interested in the transient behavior, then the prominent role is taken by the speed of convergence of $\mathbb{E}[||x(t)-\hat{x}(t)||^2$. In this case it makes sense to assume simply that $v(t)=0$.

In this paper the second case will be considered and so we will assume that $v(t)=0$. In this case, the only source of uncertainty is due to the initial condition x_0 and so the encoders/decoders task reduces to give a good estimation of x_0. On the other hand, the decoder, in order to obtain a good estimate $\hat{x}(t)$ of $x(t)$, has to obtain the best possible estimate $\hat{x}(0|t)$ of the initial condition $x(0)$ from the received data $y(0),\dots,y(t)$ and then it can define $\hat{x}(t) := A^t\hat{x}(0|t)$. In this case we have that
$$x(t)-\hat{x}(t) = A^t(x(0)-\hat{x}(0|t))$$
and the problem reduces to finding the best way of coding $x(0),\dots,x(t)$ is such a way that expansion of A^t is well dominated by the contraction of $x(0)-\hat{x}(0|t)$.

In order to clarify further this concept we present a simple example.

Example 1. *Consider the following scalar discrete time linear system*
$$x(t+1) \quad = \quad ax(t)+v(t) \qquad x(0)=x_0$$
where $a>0$ and $v(t)$ is an i.i.d. sequences of random variables with zero mean and variance σ_v^2. Assume we don't know the initial condition $x(0)$ which is assumed to be a random variable with known mean equal to zero and variance σ_x^2 which is independent of $v(t)$. Assume we run a state estimation algorithm based on the noiseless model $x(t+1)=ax(t)$ by estimating the initial condition $x(0)$ from data transmitted till time t. As before we will call this estimate $\hat{x}(0|t)$. 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 both depends on the wrong estimate of the initial condition, and on the wrong model we used. As we will see, the error $\mathbb{E}[(x(0)-\hat{x}(0|t))^2]$ is proportional to σ_x^2 so that we are allowed to write
$$\mathbb{E}[(x(0)-\hat{x}(0|t))^2] = \zeta(t)\sigma_x^2$$
where $\zeta(t)$ will be a function converging to zero whose speed of convergence will depend on the communication channel characteristics and on the coding strategy. Since $a>1$, then
$$\mathbb{E}[e(t)^2] \simeq a^{2t}\left[\zeta(t)\sigma_x^2 + \frac{\sigma_v^2}{a^2-1}\right]$$
Therefore, in case $\sigma_x^2 >> \sigma_v^2$, there will be an initial time regime in which the error will be not influenced by the model noise.

In the previous example we have seen that, if the model noise is lower than the initial state uncertainty and if we are interested in the transient behavior and so in the behavior in the first time steps, the problem consists in finding an efficient coding of the initial condition. On the other hand, if we are interested in the steady state performance (or in both transient and steady state), then it will be more efficient for the encoder to code all the state sequence $x(0),\dots,x(t)$, because in it there will be information both on the initial condition $x(0)$ and on the noise evolution $v(0),\dots,v(t)$, both affecting $x(t)$.

C. Related literature

In [4] the notion of *anytime reliability* of a communication channel was introduced, motivated by the problem of stabilizing of an unstable linear scalar plant with noisy observations. In [5] upper bounds on the anytime reliability function were obtained both with and without feedback. Techniques similar to those providing lower bounds to the anytime reliability without feedback, based on the fundamental works on infinite convolutional codes [7], [2], will be employed in Sect.IV.

The paper [6] considers the problem of encoding an unstable scalar Markov process through a DMC at the minimum possible rate, with an asymptotic constraint on the average squared-distortion reconstruction level. Our work differs from [6] in that, as explained in Sect.I-B, we concentrate on the transient behavior of the estimation error rather than on the asymptotic one. Moreover, the most original contribution of the present paper consists in analyzing the tradeoffs between performance and encoding complexity, which had not been previously considered.

D. Organization

The rest of this paper is organized as follows. In Sect.II we present some preliminary considerations: first we address the case of noiseless digital channels, in which our problem reduces to quantization; then we prove an upper bound on the achievable error exponent on the BEC; finally, we analyze some simple schemes. In Sect.III we prove a lower bound on the error probability, showing that no exponential error rates can be hoped for using finite-window encoders. Finally, in Sect.IV, we prove a lower bound on the achievable error exponent by using a random convolutional coding argument.

II. Some preliminary considerations

A. Performance of the quantizers

We start with some preliminary considerations on the errors introduced by the quantization stage. Even if in the scheme (4) the channel is noiseless, so that it is possible for the decoder to obtain $(\hat{w}_s(t))_{s=1}^t = (w_s)_{s=1}^{m_t}$, some error is nevertheless introduced by $\mathscr{S}_t$. Indeed, in this case we have that

$$\hat{x}_t = \mathscr{S}_{m_t}^{-1} \circ \mathscr{S}_{m_t}(x)$$

Let $Q_t := \mathscr{S}_t^{-1} \circ \mathscr{S}_t$. This maps $\mathscr{X}$ into itself and moreover its range has $|\mathscr{W}|^t$ elements. In the literature, such maps are called quantizers and the error (5) is called quantization error and can be rewritten as

$$(\mathbb{E}||x - Q_t(x)||^2)^{1/2}. \tag{7}$$

We refer to [3] for a comprehensive introduction to vector quantization theory. In the following we repost some basic asymptotic results which will be needed in the sequel. From [3, Th.6.2, pag.78] we can deduce the following result.

Theorem 1. *Suppose that $\mathbb{E}||x||^{2+\eta} < +\infty$ for some $\eta > 0$. Let $Q_t : \mathscr{X} \to \mathscr{X}$ be such that $|\mathbb{I}\text{m } Q_t| \le k^t$ for all $t \in \mathbb{N}$, where $k \in \mathbb{N}$. Then there exists $C_m > 0$ such that*

$$(\mathbb{E}||x - Q_t(x)||^2)^{1/2} \ge C_m k^{-t/d} \tag{8}$$

The asymptotic behavior $k^{-t/d}$ can indeed be achieved as the following example shows. For simplicity we limit to the case $k = 2$.

Example 2. *Assume that $\mathscr{X} = [-1,1]^d$. Let $q_s : [-1,1] \to \mathbb{R}$ be the uniform quantizer on $[-1,1]$ with 2^s levels, defined by*

$$q_s(x) := -1 + \frac{2k+1}{2^s} \text{ if } x \in [-1 + 2k2^{-s}, -1 + 2(k+1)2^{-s}[$$

for $0 \le k < 2^s$. Moreover, given a vector $S = (s_1, \dots, s_d) \in \mathbb{N}^d$ we define $Q_S : [-1,1]^d \to \mathbb{R}^d$ as

$$Q_S(x_1, \dots, x_d) := (q_{s_1}(x_1), \dots, q_{s_d}(x_d))$$

Notice that $x - Q_S(x) \in D^{-1}\mathscr{X}$ where $D := \text{diag}\{2^{s_1}, \dots, 2^{s_d}\}$ and so

$$||x - Q_S(x)||^2 \le \sum_{i=1}^d 2^{-2s_i}$$

Let $S(t) = (s_1(t), \dots, s_d(t)) \in \mathbb{N}^d$ be any sequence such that $\sum_i s_i(t) = t$ and $\max_i s_i(t) - \min_i s_i(t) \le 1$. This implies that

$$S(t) := \lfloor t/d \rfloor \mathbb{1} + \beta(t)$$

where $\beta(t) \in \{0,1\}^d$. It is clear the in order to encode the output of this quantizer $Q_{S(t)}(x)$ we need t bits. Moreover, notice that

$$||x - Q_{S(t)}(x)||^2 \le \sum_{i=1}^d 2^{-2s_i(t)} \le \sum_{i=1}^d 2^{-2\lfloor t/d \rfloor - 2\beta_i(t)} \le \frac{4d}{4^{t/d}}$$

This implies that

$$(\mathbb{E}||x - Q_{S(t)}(x)||^2)^{1/2} \le \left(2d^{1/2}\right) 2^{-t/d}$$

The previous theorem and the previous example show that there exist families of quantizers $Q_t : \mathscr{X} \to \mathscr{X}$ such that $|\mathbb{I}\text{m } Q_t| \le k^t$ and

$$C_m(k^{-1/d})^t \le (\mathbb{E}||x - Q_t(x)||^2)^{1/2} \le (C_M k^{-1/d})^t \tag{9}$$

where $C_m, C_M > 0$ are positive constants.

The previous result implies in particular that in the scheme (4) with noiseless channel, since $\tilde{\mathscr{E}}_t$ is injective, then $|\mathscr{W}|^{m_t} \le |\mathscr{Y}|^t$, and then

$$\mathbb{E}[||x - \hat{x}_t||^2]^{1/2} = (\mathbb{E}||x - \mathscr{S}_{m_t}^{-1} \circ \mathscr{S}_{m_t}||^2)^{1/2} \ge C(|\mathscr{Y}|^{-1/d})^t. \tag{10}$$

In this paper we are not interested in estimating the optimal constant which in general will be smaller than $2d^{1/2}$. Our interest here is for the asymptotic behavior $2^{-t/d}$ which we have now shown to be achievable. We will refer to all such schemes as dyadic quantization schemes.

Dyadic quantization schemes have an important property related to the fact that there is a natural hierarchy in the bits expansion, which is captured by the following definition. The scheme consisting of the encoder $\mathscr{E}$ and decoders $\mathscr{D}_t$ is said to be an uncoded scheme if there exists a constant $C > 0$ such that, for every $\overline{x} \in \mathscr{X}$,

$$\mathbb{E}[||x - \overline{x}|| \,|\, \mathscr{E}(x)_t \ne \mathscr{E}(\overline{x})_t] \ge C2^{-t/d}. \tag{11}$$

Dyadic schemes are always uncoded.

B. A lower bound for the BEC

Clearly, for noisy channels a degradation in performance is expected. We shall now propose a lower bound for the BEC.

Consider the general scheme (2). Given $(z_s)_{s=1}^t \in \mathscr{Z}^t$, the associated error pattern is given by a sequence $(e_s)_{s=1}^t \in \{c, ?\}^t$ defined componentwise by mapping $0,1$ into c and ? into ?. Conditioning on an infinite error pattern $(e_s)_{s=1}^\infty \in \{c, ?\}^{\mathbb{N}}$, the channel becomes a deterministic map, so that the composition of all the maps in (2) is a quantizer from $\mathscr{X}$ to itself. Notice that this quantizer has range with cardinality $|\mathscr{Y}|^{l_t}$ where l_t is the number of components of $(e_s)_{s=1}^t$ equal to c. From this fact and from Theorem 1, we can deduce that

$$\mathbb{E}\left[||x - \hat{x}_t||^2 | (e_s)_{s=1}^t\right] \ge C_m^2 |\mathscr{Y}|^{-2l_t/d}$$

Define $E_l := \left\{(e_s)_{s=1}^t \in \{c, ?\}^t : l_t = l\right\}$ and observe that $\mathbb{P}[E_l] = \binom{t}{l}\varepsilon^{t-l}(1-\varepsilon)^l$. Then, we have that

$$\begin{aligned}
\mathbb{E}\left[||x - \hat{x}_t||^2\right] &= \textstyle\sum_{(e_s) \in \{c,?\}^t} \mathbb{E}\left[||x - \hat{x}_t||^2 | (e_s)\right] \mathbb{P}\left[(e_s)\right] \\
&\ge \textstyle\sum_{l=0}^t \sum_{(e_s) \in E_l} C_m^2 |\mathscr{Y}|^{-2l/d} \mathbb{P}\left[(e_s)\right] \\
&= C_m^2 \textstyle\sum_{l=0}^t |\mathscr{Y}|^{-2l/d} \mathbb{P}\left[E_t\right] \\
&= C_m^2 \textstyle\sum_{l=0}^t |\mathscr{Y}|^{-2l/d} \binom{t}{l} \varepsilon^{t-l}(1-\varepsilon)^l \\
&= C_m^2 \left(\varepsilon + (1-\varepsilon)|\mathscr{Y}|^{-2/d}\right)^t
\end{aligned} \tag{12}$$

Hence, we have proved the following result.

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

$$\Delta_t \ge C_m 2^{-t\overline{\beta}(d,\varepsilon)}, \tag{13}$$

for all $t \geq 0$, where

$$\overline{\beta}(d,\varepsilon) := -\frac{1}{2}\log_2\left(\varepsilon + (1-\varepsilon)2^{-2/d}\right)$$

and C_m is a positive constant depending on ε and f only.

Remark 1. *The Shannon capacity of the BEC (measured in bits per channel use) is well known to equal $1-\varepsilon$, which is the average number of non-erased bits per channel use. It can be directly verified that*

$$\overline{\beta}(d,\varepsilon) < \frac{1}{d}(1-\varepsilon), \qquad \forall \varepsilon \in]0,1[. \tag{14}$$

The inequality (14), together with (13) and (9), 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 equal to t times the capacity. In other words, (14) 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 (12) shows that the reason for this is due to the fact that the exponential rate is dominated by error events E_l, $\lim_t \frac{l}{t} = \frac{\varepsilon}{\varepsilon+(1-\varepsilon)2^{-1/d}} < \varepsilon$, of asymptotically zero probability.

Remark 2. *It is not hard to see that (13) continues to hold true even if the encoder has access to noiseless causal feedback, namely if, at each time $t \geq 0$, $\mathscr{E}_t : \mathscr{X} \times \mathscr{Z}^{t-1} \to \mathscr{Y}$ is allowed to depend on the past channel outputs as well as on the observed vector x. A fortiori, (13) holds in the case of partial or noisy feedback, which is a typical situation occurring in sensor networks.*

Remark 3. *In the case of perfect causal output feedback, using dyadic schemes, the bound (13) can be achieved using the coding scheme of repeating the transmission of a bit until it is correctly received. However, it is not clear a priori whether (13) with noisy feedback, or without feedback. In Sect.II-C we will propose some simple schemes which however are not capable of reaching exponential error decay.*

C. Some simple coding schemes for the BEC

Consider the scheme (4) in which we take any $\mathscr{S}_t$ and $\mathscr{S}_t^{-1}$ such that the bound (9) holds. If we would simply use this on a BEC, decoding erasures in an arbitrary way, we would not even have the guarantee of an error converging to 0 when $t \to +\infty$: this because of (11) considering the fact that with probability ε we would have lost the first symbol of the expansion given by $\mathscr{S}_t$. The standard technique to overcome such problems is to introduce redundancy in order to cope with packet drop phenomena. Consider the encoder $\tilde{\mathscr{E}} : \{0,1\}^{\mathbb{N}} \to \{0,1\}^{\mathbb{N}}$ informally described by

$$\tilde{\mathscr{E}}((w_s)_{s=1}^{\infty}) = (w_1, w_1, w_2, w_1, w_2, w_3, w_1, w_2, w_3, w_4 \cdots).$$

More precisely, notice that for any $t \in \mathbb{N}$ there exists unique $m \in \mathbb{N}$ and $j \in \{1,2,\dots,m\}$ such that $t = (m-1)m/2 + j$. Denote these numbers by $m(t)$ and $j(t)$, respectively. If $(y_s)_{s=1}^{\infty} = \tilde{\mathscr{E}}((w_s)_{s=1}^{\infty})$, then $y_t = w_{j(t)}$. Notice that this encoder fits in the scheme (4) by taking $m_t = m(t)$. We construct the decoders $\tilde{\mathscr{D}}_t : \{0,1,?\}^t \to \{0,1\}^{m_t}$ as follows. If $(\hat{w}_j(t))_{j=1}^{m_t} = \tilde{\mathscr{D}}_t((z_s)_{s=1}^t)$, then

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

Let us estimate the performance of this scheme. Assume for simplicity that $t = m(m+1)/2$ for some $m \in \mathbb{N}$. For any $j \in \{1,2,\dots,m\}$ the number of times that w_j is transmitted is $m-j+1$ and so

$$\mathbb{P}[\hat{w}_j(t) \neq w_j] = \varepsilon^{m-j+1}$$

For $0 \leq \tau < m$, consider the event

$$A_\tau^t = \{\hat{w}_1(t) = w_1, \dots, \hat{w}_\tau(t) = w_\tau, \hat{w}_{\tau+1}(t) \neq w_{\tau+1}\},$$

and let $A_m^t := \{\hat{w}_1(t) = w_1, \dots, \hat{w}_m(t) = w_m\}$. Notice that, for all $0 \leq \tau \leq m$, we have that

$$\begin{aligned} \mathbb{P}[A_\tau^t] &= \textstyle\prod_{j=1}^{\tau} \mathbb{P}[\hat{w}_j(t) = w_j]\mathbb{P}[\hat{w}_{\tau+1}(t) \neq w_{\tau+1}] \\ &= \textstyle\prod_{j=1}^{\tau}(1-\varepsilon)^{m-j+1}\varepsilon^{m-\tau} \\ &\leq \varepsilon^{m-\tau}. \end{aligned} \tag{15}$$

Notice moreover that if $(w_s)_{s=1}^{\tau} = (\hat{w}_s)_{s=1}^{\tau}$, then $Q_\tau(x) = Q_\tau(\hat{x}_t)$. Hence

$$\begin{aligned} 2C_M^2 2^{-\tau/d} &\geq \mathbb{E}[||x - Q_\tau(x)||^2 | A_\tau^t] + \mathbb{E}[||Q_\tau(\hat{x}_t) - \hat{x}_t||^2 | A_\tau^t] \\ &\geq \mathbb{E}[||x - \hat{x}_t||^2 | A_\tau^t] \end{aligned}$$

From this it follows that

$$\Delta_t^2 = \sum_{\tau=0}^{m} \mathbb{E}\left[||x-\hat{x}_t||^2 \,|\, A_\tau^t\right]\mathbb{P}[A_\tau^t] \leq 4C_M^2 \sum_{\tau=0}^{m} \varepsilon^{m-\tau}2^{-2\tau/d}$$

If $\varepsilon 2^{2/d} > 1$, then

$$\Delta_t^2 \quad \leq \quad 4C_M^2\varepsilon^m \left[\sum_{\tau=0}^{m}(\varepsilon^{-1}2^{-2/d})^\tau\right] \leq \frac{4C_M^2}{1-(\varepsilon 2^{2/d})^{-1}}\varepsilon^m$$

Observe now that, since $t = m(m+1)/2$, we have that $m \geq \sqrt{2t} - 1$ and so

$$\Delta_t \leq C_{\varepsilon,d}\,\varepsilon^{\sqrt{t/2}} \tag{16}$$

where $C_{\varepsilon,d}$ is a positive constant depending on ε and on d. In a similar way we can prove that in the regime if $\varepsilon 2^{2/d} < 1$, we obtain

$$\Delta_t \leq C'_{\varepsilon,d}(2^{-2/d})^{\sqrt{t/2}} \tag{17}$$

In both cases we obtain a sub-exponential asymptotics of type, far away from the exponential lower bounds proved above. The first point to be discussed is if this estimation is asymptotically tight. The answer is on the positive and this can be seen going back to the expression (15), noticing that

$$\prod_{j=1}^{\tau}(1-\varepsilon^{m-j+1}) \geq \prod_{j=1}^{m}(1-\varepsilon^j).$$

This last product converges to a non zero value for $n \to \infty$ for every $\varepsilon < 1$. Hence, there exists a constant $D_\varepsilon > 0$ such that, for every $n \in \mathbb{N}$, it holds $\prod_{j=1}^{m}(1-\varepsilon)^j \geq D_\varepsilon$. Hence, $\mathbb{P}(A_\tau^t) \geq D_\varepsilon \varepsilon^{m-\tau}$. Repeating the same steps than above, we thus obtain that the estimation in (16) and (17) are tight in the sense that

$$\Delta_t \asymp \exp\left[-\beta\sqrt{t}\right]$$

where

$$\beta = \begin{cases} \sqrt{\frac{1}{2}}\ln\varepsilon^{-1} & \text{if } \varepsilon 2^{2/d} > 1 \\ \frac{\sqrt{2}\ln 2}{d} & \text{if } \varepsilon 2^{2/d} < 1 \end{cases}$$

Notice finally that ε does not even show up in the asymptotic behavior when $\varepsilon 2^{2/d} < 1$ and in particular there is no improvement for $\varepsilon \to 0$. Is it possible to improve our scheme? Is it possible to achieve exponential decays?

In this section we present a slight variation of the scheme described above which improves performance for small ε while remaining sub-exponential. The question of exponential decay will be considered later on.

Fix a positive real number m and consider the sequence of nonnegative integers

$$\tau_0 = 0,\ \ \tau_k = \lfloor m\rfloor + \lfloor 2m\rfloor + \cdots + \lfloor km\rfloor \quad \text{for } k > 0\,.$$

Consider now the mapping $\tilde{\mathscr{E}} : \{0,1\}^{\mathbb{N}} \to \{0,1\}^{\mathbb{N}}$ defined by

$$[\tilde{\mathscr{E}}(\omega)]_t = \omega_{t-\tau_k} \quad \text{if } t \in]\tau_k, \tau_{k+1}]\,.$$

Our encoder will be in this case $\mathscr{E} = \tilde{\mathscr{E}} \circ \mathscr{S}$. Define now $\tilde{\mathscr{D}}_t : \{0,1,?\}^t \to \{0,1\}^t$ by putting $[\tilde{\mathscr{D}}_t(\omega)]_s = \omega_s$ if there exist $k \in \mathbb{N}$ and $j \in]\tau_{k-1}, \tau_k] \cap [1,t]$ such that $j - \tau_{k-1} = s$ and $\omega_j \neq ?$, and $[\tilde{\mathscr{D}}_t(\omega)]_s = 0$ otherwise. Moreover, let $\mathscr{D}_t = \mathscr{S}_t^{-1} \circ \tilde{\mathscr{D}}_t$. Notice that, if $m = 1$, then $\mathscr{E}$ and $\mathscr{D}_t$ coincide with the one above.

Define now $\mu_s^n = |\{1 \le j \le n \mid \lfloor mj\rfloor \ge s\}|$, and notice that, if $s \le nm$,

$$n + 1 - \frac{s+1}{m} \le \mu_s^n \le n + 1 - \frac{s}{m}$$

At time $t = \tau_n$ we have that ω_s has been repeated exactly μ_s^n times in the first t bits of $\mathscr{S}(\omega)$.

We can now estimate the error as we have done before:

$$\begin{aligned} \Delta_t^2 &= \textstyle\sum_s \mathbb{E}\left[||X - \mathscr{D}_t\pi_t\mathscr{E}(X)||^2 \,|\, A_s\right]\mathbb{P}[A_s] \\ &\le C^2\left[\sum_{s=0}^{nm-1}\varepsilon^{\mu_{s+1}^n}2^{-2s/d} + \sum_{s=nm}^{\infty}2^{-2s/d}\right] \\ &\le C^2\left[\sum_{s=0}^{nm-1}\varepsilon^{n+1-\frac{s+1}{m}}2^{-2s/d} + \sum_{s=nm}^{\infty}2^{-2s/d}\right] \\ &= C^2\left[\varepsilon^{n+1-\frac{1}{m}}\frac{1-(\varepsilon^{1/m}2^{2/d})^{-nm}}{1-(\varepsilon^{1/m}2^{2/d})^{-1}} + \frac{2^{-2nm/d}}{1-2^{-2/d}}\right] \end{aligned}$$

Now we fix $\alpha \in]0,2[$ and we fix m in such a way that $(\varepsilon^{1/m}2^{2/d})^{-1} = 2^{-\alpha/d}$. This is equivalent to put

$$m = \frac{\ln\varepsilon^{-1}}{\ln 2^{(2-\alpha)/d}}$$

Notice that, for such value of m,

$$2^{-2nm/d} = \exp\left[-n\ln\varepsilon^{-1}\frac{\ln 2^{2/d}}{\ln 2^{(2-\alpha)/d}}\right] \le \varepsilon^n$$

Hence, we can estimate

$$\Delta_t \le C_\varepsilon \exp\left[-\frac{1}{2}n\ln\varepsilon^{-1}\right].$$

Notice that $t = \tau_n \le mn(n+1)/2$, hence, $n \ge \sqrt{\frac{2t}{m}} - 1$. We thus obtain the estimation

$$\Delta_t \le C_\varepsilon \exp\left[-\sqrt{t\frac{(1-\alpha/2)\ln 2\ln\varepsilon^{-1}}{d}}\right].$$

As before, it is not difficult to prove that this estimation is asymptotically tight which means that, for $t \to +\infty$,

$$\Delta_t^2 \asymp \exp\left[-\sqrt{t\frac{(1-\alpha/2)\ln 2\ln\varepsilon^{-1}}{d}}\right]. \tag{18}$$

In the next section we will see that, using repetition encoding schemes as the one above where bits are transmitted the way they are, such a performance can not be beated.

III. Performance of finite window coding schemes

In this section we consider a more general class of encoders encompassing previous examples but still maintaining a bounded complexity in their implementation which means that the number of computations that the encoder has to perform at every time instant remains bounded in time.

For the rest of this section, we assume we have fixed a dyadic quantization scheme for $\mathscr{X}$ consisting of the encoder $\mathscr{Q}$ and of the sequence of decoders $\mathscr{G}_t$.

We will focus on encoders $\mathscr{E} : [0,1] \to \{0,1\}^{\mathbb{N}}$ transmitting at each time t an input symbol $\mathscr{E}(x)_t$ which is a function of d_t bits of $\mathscr{Q}(x)$. More formally, we consider encoders of the form

$$\mathscr{E}(x)_t = f_t(\pi_{A_t}\mathscr{Q}(x)) \tag{19}$$

where A_t is a finite subset of $\mathbb{N}$ of cardinality $|A_t| = d_t$, and $f_t : \{0,1\}^{A_t} \to \{0,1\}$. These encoders will be called finite window encoders. Clearly each encoder of the form (19) is specified by a subset sequence (A_t) and a function sequence (f_t). To each encoder of the form (19) it is possible to associate, for every $j,t \in \mathbb{N}$, the quantity

$$z_j(t) := \sum_{s\le t}\mathbb{1}_{A_s}(j)$$

counting the number of channel inputs which have been affected by the j-th bit of the dyadic expansion $\mathscr{Q}(X)$ up to time t. Notice that

$$H_t := \sum_{j\in\mathbb{N}} z_j(t) = \sum_{s\le t} d_s\,.$$

The quantity H_t can be considered as a measure of the complexity of the encoder $\mathscr{E}$. This is our main result.

Theorem 3. *For any transmission scheme for the BEC, with erasure probability ε, consisting of a finite window encoder of the form (19) and with complexity function H_t, it holds*

$$\Delta_t \ge C\exp\left[-\sqrt{\frac{\ln 2\ln\varepsilon^{-1}}{d}H_t}\right], \tag{20}$$

for some constant $C > 0$ only depending on d and f.

Proof Assume that at time t all the $z_j(t)$ channel inputs affected by the j-th bit $\mathscr{Q}(x)_j$ have been erased. Then, there is clearly no way to reconstruct $\mathscr{Q}(x)_j$ from the output of the channel. Using (11), this gives the following lower bound to the squared estimation error, independent on the way the decoders are chosen:

$$\Delta_t^2 \geq C_1 \sup_{j\in\mathbb{N}}[2^{-2j/d}\varepsilon^{z_j(t)}],$$

for some constant $C_1 > 0$ only depending on d and f. It will be convenient to consider the looser bounds, for any $s \in \mathbb{N}$,

$$\Delta_t^2 \geq C_1 \sup_{j\leq s}[2^{-2j}\varepsilon^{z_j(t)}] \geq C_1 f_s(z(t))$$

where

$$f_s : (\mathbb{R}^+)^s \to \mathbb{R}, \quad f_s(z) := \frac{1}{s}\sum_{j=1}^{s}\exp\left(-j\frac{\ln 4}{d} + z_j \ln\varepsilon\right).$$

Hence, for every possible s,

$$\Delta_t^2 \geq C_1 \inf_{z\in M_s} f_s(z)$$

where

$$M_s := \{z \in (\mathbb{R}^+)^s, |\ \sum_j z_j = H_t\}.$$

Since the function $f_s(z)$ is strictly convex, it admits a unique minimum on the convex compact set M_s. Using Lagrange multipliers we can characterize the stationary point of $f_s(z)$ on the hyperplane $\sum_j z_j = H_t$:

$$z_j^* = \alpha - \beta j, \qquad \forall j \leq s,$$

where $\beta := -\frac{\log 4}{d\log\varepsilon} > 0$, and $\alpha = \frac{H_t}{s} + \beta\frac{s+1}{2}$. We have that $z^* \in M_s$ if and only if $z_s^* \geq 0$ which is equivalent to

$$s \leq \frac{1+\sqrt{1+\frac{8H_t}{\beta}}}{2}.$$

A possibility consists in choosing $s^* = \left\lfloor\sqrt{\frac{2H_t}{\beta}}\right\rfloor$. We get

$$\Delta_t^2 \geq C_1 \inf_{z\in M_{s^*}} f_{s^*}(z) = f_{s^*}(z^*) = Ce^{\alpha\ln\varepsilon}. \tag{21}$$

We can estimate α as follows

$$\alpha = \frac{H_t}{\left\lfloor\sqrt{2H_t/\beta}\right\rfloor} + \beta\frac{\left\lfloor\sqrt{2H_t/\beta}\right\rfloor+1}{2} \leq \beta\left(\sqrt{2H_t/\beta}+1\right),$$

(where last equality follows from straightforward algebraic computation). Inserting this last estimation inside (21), the thesis follows. ■

Remark 4. *Notice that, in the case of the repetition encoding schemes treated in Sect.II-C, we have that $H_t = t$. If we compare (20) with (18), considering the fact that α can be picked arbitrarily close to 0, we have thus established that among the repetition schemes ($H_t = t$), the example treated in Sect. II-C is optimal from the point of view of the asymptotic performance.*

IV. A CASUAL LINEAR CODING THEOREM

In this section, we shall prove a lower bound to the estimation error exponent on the BEC. We shall use random coding arguments employing anytime linear codes over the binary field $\mathbb{Z}_2$. These arguments date back to the early literature on convolutional codes [7], [2], and have been recently applied in [6] in order to lower bound the anytime reliability function without feedback.

For a rate $0 < R < 1$, we consider a random, doubly infinite, causal, binary matrix $\phi \in \mathbb{Z}_2^{\mathbb{N}\times\mathbb{N}}$ distributed as follows: $\phi_{ij} = 0$ for all $j > Ri$, while $\{\phi_{ij} | 1 \leq j \leq Ri\}$ is a family of independent r.v.s identically distributed uniformly over $\mathbb{Z}_2$. Furthermore, we shall assume ϕ to be independent both form the source vector x and from the channel, and known both at the transmitting and receiving ends. We shall naturally identify the matrix ϕ with the corresponding linear operator from $\mathbb{Z}_2^{\mathbb{N}}$ to $\mathbb{Z}_2^{\mathbb{N}}$, and, for $t \in \mathbb{N}$, consider the truncated version $\phi_t : \mathbb{Z}_2^{\lceil Rt\rceil} \to \mathbb{Z}_2^t$.

As in Sect.II-B, for a channel output sequence $(z_s)_{s\in\mathbb{N}}$, consider the error pattern $(e_s)_{s\in\mathbb{N}}$, defined by putting $e_s = ?$ if $z_s = ?$, and $e_s = c$ if $z_s \neq ?$. Also, for $t \in \mathbb{N}$, let $I_t := \{1 \leq i \leq t : z_i \neq ?\} = \{1 \leq i \leq t : e_i \neq ?\}$ be the set of non-erased positions up to time t, and let $\pi_{I_t} : \mathbb{Z}_2^t \to \mathbb{Z}_2^{I_t}$ be the canonical projection. An obvious class of decoders for ϕ is given by those $\tilde{\mathscr{D}}_t : \mathscr{Z}^t \to \mathbb{Z}_2^{\lceil Rt\rceil}$ satisfying $\tilde{\mathscr{D}}_t(\boldsymbol{z}) \in (\pi_{I_t}\circ\phi_t)^{-1}(\boldsymbol{z})$. In fact, observe that the preimage $(\pi_{I_t}\circ\phi_t)^{-1}(\boldsymbol{z})$ is never empty. Notice that, the decoding $\tilde{\mathscr{D}}_t((z_s)_{s=1}^t)$ is uniquely defined, and correct, whenever $\pi_{I_t}\circ\phi_t : \mathbb{Z}_2^{\lceil Rt\rceil} \to \mathscr{Z}^{I_t}$ is injective.

Now, let $\mathscr{S} : \mathscr{X} \to \mathbb{Z}_2^{\mathbb{N}}$ be a dyadic quantizer, and define the encoding scheme $\mathscr{E} : \mathscr{X} \to \mathbb{Z}_2^{\mathbb{N}}$ as the composition $\mathscr{E} = \phi\circ\mathscr{S}$, and the sequence of decoders $\mathscr{D}_t := (\pi_t\circ\mathscr{S})^{-1}\circ\tilde{\mathscr{D}}_t$. The following result characterizes the average mean squared error of the scheme $(\mathscr{E}, \mathscr{D}_t)$ 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.

Proposition 4. *Assume transmission over the BEC. Then, for all $0 < R < 1$, the average estimation error of the random coding scheme described above satisfies*

$$\mathbb{E}[||x - \hat{x}_t||^2] \leq Kt^2 2^{-2t\underline{\beta}(d,\varepsilon,R)} \tag{22}$$

for all $t \in \mathbb{N}$, where

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

$D(x||y) := x\log_2\frac{x}{y} + (1-x)\log_2\frac{1-x}{1-y}$ denotes the binary Kullback-Leiber distance, and $K > 0$ is a constant depending on the erasure probability ε and the density function f only.

Proof Let $\{\delta_i\}_{1\leq i\leq\lceil Rt\rceil}$ be the canonical basis of $\mathbb{Z}_2^{\lceil Rt\rceil}$. For $0 \leq j \leq \lceil Rt\rceil$, consider the subspace K_j of $\mathbb{Z}_2^{\lceil Rt\rceil}$ generated by the elements $\delta_{j+1}, \ldots, \delta_{\lceil Rt\rceil}$, and define the event $A_j := \{\ker(\pi_{I_t}\circ\phi_t) \subseteq K_j\}$. It is not difficult to verify that if A_j occurs, than $\tilde{\mathscr{D}}_t$ successfully decodes the first j bits, i.e. $\pi_j\tilde{\mathscr{D}}_t((z_s)_{s=1}^t) = \pi_j\mathscr{S}(x)$, so that (9) implies that $||\hat{x}_t - x|| \leq C_M 2^{-j/d}$. Notice that the events A_j are nested, i.e. $A_j \supseteq A_{j+1}$, and that A_0 coincides with the whole sample space. Then, by defining $B_j := A_{j-1} \cap \overline{A_j}$, we have

$$\begin{aligned}\mathbb{E}\left[||\hat{x}_t - x||^2\right] &= \sum_{j=1}^{\lceil Rt\rceil} \mathbb{E}\left[||\hat{x}_t - x||^2 \mathbb{1}_{B_j}\right] + \mathbb{E}\left[||\hat{x}_t - x||^2 \mathbb{1}_{A_{\lceil Rt\rceil}}\right] \\ &\le \sum_{j=1}^{\lceil Rt\rceil} \mathbb{P}(B_j) C_M 2^{-2(j-1)/d} + C_M 2^{-2\lceil Rt\rceil}\end{aligned} \tag{24}$$

In order to estimate the probability of the event B_j, first we claim that B_j implies that the j-th column $\pi_{I_t}\phi_t\delta_j$ is linear combination of the following ones $\{\pi_{I_t}\phi_t\delta_i\}_{j<i\le\lceil Rt\rceil}$. Indeed, A_{j-1} implies that, if $\pi_{I_t}\phi_t \boldsymbol{w} = 0$ for some $\boldsymbol{w} \in \mathbb{Z}_2^{\lceil Rt\rceil}$, than necessarily $w_i = 0$ for all $i < j$, while [1] $\overline{A_j}$ implies that $\pi_{I_t}\phi_t\boldsymbol{w} = 0$ for some $\boldsymbol{w} \in \mathbb{Z}_2^{\lceil Rt\rceil}$ such that $w_j \neq 0$.

Then, observe that the subspace of $\mathbb{Z}_2^{I_t}$ generated by the columns $\{\pi_{I_t}\phi_t\delta_i\}_{j<i\le\lceil Rt\rceil}$ has cardinality at most $2^{\lceil Rt\rceil - j}$. Since $\phi_t\delta_j$ is uniformly distributed over the subspace of $\mathbb{Z}_2^t$ generated by $\{\delta_s\}_{\lfloor j/R\rfloor\le s\le t}$, and independent from the error sequence $(e_s)_{s=1}^t$, it follows that [2]

$$\mathbb{P}(B_j|E_k^j) \le 2^{-\lceil j+k-\lceil Rt\rceil\rfloor_+},$$

$$E_k^j := \{(e_s)_{s=\lfloor j/R\rfloor}^t : |\{s : e_s = ?\}| = k\}.$$

Since $\mathbb{P}(E_k^j) = \binom{t-\lfloor j/R\rfloor}{k}\varepsilon^{t-\lfloor j/R\rfloor-k}(1-\varepsilon)^k$, we get

$$\begin{aligned}\mathbb{P}(B_j) &= \sum_{k=j+1}^{\lceil Rt\rceil} P(B_j|E_k^j)\mathbb{P}(E_k^j) \\ &\le \sum_{k=j+1}^{\lceil Rt\rceil} 2^{-\lfloor j+k-\lceil Rt\rceil\rfloor_+}\binom{t-\lfloor j/R\rfloor}{k}\varepsilon^{t-\lfloor j/R\rfloor-k}(1-\varepsilon)^k \\ &\le \sum_{k=j+1}^{\lceil Rt\rceil} 2^{-\lfloor j+k-\lceil Rt\rceil\rfloor_+} 2^{-(t-\lfloor j/R\rfloor)D(\frac{k}{t-\lfloor j/R\rfloor}||1-\varepsilon)},\end{aligned} \tag{25}$$

where the second inequality follows from standard estimations of the binomial coefficient [1].

Therefore, by combining (24) and (25), setting $j = R\lambda t$ and $k = (1-\lambda)\eta t$ for $\lambda, \eta \in [0,1]$, and observing that

$$\underline{\beta}(d,\varepsilon,R) = \min_{0\le\lambda,\eta\le1} \lambda\frac{1}{d}R + (1-\lambda)\frac{1}{2}\left(D(\eta||1-\varepsilon) + \lfloor R-\eta\rfloor_+\right)$$

we obtain the claim. ■

Let us now define

$$\underline{\beta}(d,\varepsilon) := \frac{1}{2}\max_{0\le R\le1}\underline{\beta}(d,\varepsilon,R).$$

The following corollary of Theorem 4 follows by standard probabilistic arguments.

Corollary 5. *Assume transmission over the BEC with erasure probability ε. Then, there exists a coding scheme as in (2) such that*

$$\liminf_t -\frac{1}{t}\log_2\Delta_t \ge \underline{\beta}(d,\varepsilon).$$

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 $d = 1$.

[1]For an event A, $\overline{A}$ denotes its complement.

[2]We use the notation $\lfloor x\rfloor_+ := \max\{0, x\}$

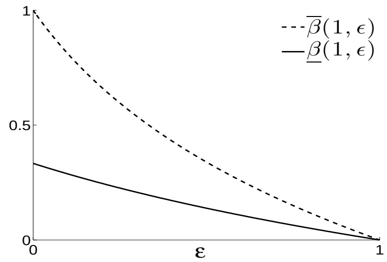

Fig. 1. Upper and lower bounds on the error exponent achievable on the BEC for $d = 1$.

V. Conclusion

We have considered the problem of anytime reliable transmission of a real-valued random vector through a digital noisy channel. 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 bounded-complexity coding schemes have been derived, showing that in this case the mean squared error cannot decrease faster than exponentially in the square root of the number of channel uses. Finally, simulation results for low-complexity coding/decoding schemes have been presented.

Current work includes extensions of the results to general discrete memoryless channels, and analysis and design of practical coding schemes. We also plan to extend the theory to deal with the case of distributed estimation over networks of agents communicating over noisy digital channels.

References

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

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

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

[4] 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(8), pp. 3369-3395, 2006.

[5] A. Sahai, "Why do block length and delay behave differently if feedback is present?", *IEEE Trans. Inf. Theory*, vol. 54(5), pp. 1860-1886, 2008.

[6] A. Sahai and S. Mitter, "Source coding and channel requirements for unstable processes", submitted to *IEEE Trans. Inf. Theory*, 2006.

[7] A. Viterbi, "Error bounds for convolutional codes and an asymptotically optimal decoding algorithm", *IEEE Trans. Inf. Theory*, vol.13 (2), pp.260-269 1967.