# On the Burnashev exponent for Markov channels

Giacomo Como
Dip. di Matem. Politecnico di Torino, Italy
visiting Yale University, New Haven, CT
giacomo.como@yale.edu

Serdar Yüksel
Electrical Engineering
Yale University, New Haven, CT
serdar.yuksel@yale.edu

Sekhar Tatikonda
Electrical Engineering
Yale University, New Haven, CT
sekhar.tatikonda@yale.edu

***Abstract*— We consider the reliability function of Markov channels with feedback and variable length channel codes. We extend Burnashev's [3] classic result to this case and present a single letter characterization for the reliability function.**

## I. Introduction

It is known that output feedback cannot increase the capacity of a discrete memoryless channel (DMC) [7]. Feedback, though, improves its reliability function. A classical result due to Burnashev [3] characterizes the reliability function of a DMC with a simple single-letter formula when one is allowed variable length block coding. It is remarkable that in this case, differently from when feedback is not available, the reliability function is exactly known at any rate below capacity, and that it has nonzero slope approaching capacity.

Recently, the problem of channel coding with feedback, and Burnashev's result in particular, have attracted renewed interest from the researchers; see [8], [9], [5], [1].

In this paper we address the problem of generalizing Burnashev's result to channels with memory. Specifically, we examine the simplest case of Markov channels with perfect state information at the transmitter and receiver and no intersymbol-interference (ISI). For this class of channels we are able to exactly characterize, in a single-letter form, the reliability function

$$E(R) = C^1\left(1 - \frac{R}{C}\right) \qquad (1)$$

for suitably defined $C, C^1$ (see (3), (4) and Theorem 1). Hence, Burnashev's exponent extends naturally to the Markov setting. However, the extension is non-trivial as its analysis involves significant technical challenges.

Our proof of the converse is based on an extension of Burnashev's original one [3], following some of the ideas of Berlin et al. [1] (see also [9]). Specifically, we provide two different bounds for the error probability, involving respectively the channel capacity $C$ and its Burnashev coefficient $C^1$, corresponding to two distinct phases which can be recognized in any sequential transmission scheme. Similarly to the memoryless case, martingale theory, and in particular Doob's optional sampling theorem, is widely used jointly with more standard information theoretic results like Fano's and log-sum inequalities. The use of standard ergodic theory for Markov chains coupled with stopping times arguments is instead peculiar of the memory case.

In order to prove the achievability of the exponent (1), we propose a simple two-phase iterative transmission scheme based on the one first considered by Yamamoto and Itoh [10] for DMCs. The performance analysis of this scheme relies on known results about the capacity of Markov channels, and the error exponent of binary hypothesis tests for irreducible Markov chains.

In Section II we formulate the problem, introduce notation, and state the main result. In Section III we prove the converse. In Section IV we prove achievability by presenting a generalization of the Yamamoto and Itoh scheme.

[1]This work was partially supported by the National Science Foundation under Grant ECCS-0547199.

## II. Problem formulation and main result

### A. Notation

Given a finite set $\mathcal{A}$, we shall denote by $\mathcal{P}(\mathcal{A})$ the space of probability measures over $\mathcal{A}$. When $\boldsymbol{\mu}$ is in $\mathcal{P}(\mathcal{A})$ and $\boldsymbol{f}$ is in $\mathbb{R}^{\mathcal{A}}$, we shall use the notation $\langle \boldsymbol{\mu}, \boldsymbol{f} \rangle = \sum_{a\in\mathcal{A}} \boldsymbol{\mu}(a)\boldsymbol{f}(a)$.

The set of all infinite $\mathcal{A}$-sequences is denoted by $\mathcal{A}^{\mathbb{N}}$, while the set of all finite $\mathcal{A}$-sequences is denoted by $\mathcal{A}^* = \vee_{n\in\mathbb{N}}\mathcal{A}^n$. For $\boldsymbol{a}$ in $\mathcal{A}^{\mathbb{N}}$, and $s \leq t$, $\boldsymbol{a}_s^t \in \mathcal{A}^{t-s+1}$ denotes the restriction of $\boldsymbol{a}$ to the discrete interval $[s,t]$. The length of an element $\boldsymbol{x}$ of $\mathcal{A}$ is denoted by $L(\boldsymbol{x})$, and the empirical frequency function is defined as

$$\boldsymbol{v}_{\mathcal{A}} : \mathcal{A}^* \to \mathcal{P}(\mathcal{A}), \quad [\boldsymbol{v}_{\mathcal{A}}(\boldsymbol{x})](a) = \frac{|\{1 \leq j \leq n \,:\, x_j = a\}|}{L(\boldsymbol{x})}.$$

When $\boldsymbol{A}$ is an $\mathcal{A}$-valued random process, we will deal with finite stopping times $T$ for $\boldsymbol{A}$ [2]. These can be represented as complete prefix-free subsets of $\mathcal{A}^*$, or $\mathcal{A}$-ary trees. With this interpretation, a $T$-measurable random variable can thus be thought as a function over the set $\mathcal{A}^T$ of the leaves of this tree.

### B. Stationary ergodic Markov channels

Throughout the paper $\mathcal{X}$, $\mathcal{Y}$, $\mathcal{S}$ will respectively denote input, output and state set, all finite.

**Definition 1** *A stationary Markov channel with no ISI is described by:*

- *a family* $\{P_Y(\,\cdot\,|x,s) \in \mathcal{P}(\mathcal{Y})\,|\, x\in\mathcal{X}, s\in\mathcal{S}\}$ *of probability measures over* $\mathcal{Y}$ *indexed by elements of* $\mathcal{X}$ *and* $\mathcal{S}$*;*
- *a stochastic matrix* $\Pi = (P_S(s|r))_{s,r}$ *over* $\mathcal{S}$*;*

- *an initial state distribution* $\boldsymbol{\mu}$ *in* $\mathcal{P}(\mathcal{S})$.

Throughout the paper we will restrict ourselves to ergodic Markov channels, satisfying the following.

**Assumption 2** $\Pi$ *is irreducible.*

We won't need any aperiodicity assumption. By Perron-Frobenius theorem we have that $\Pi$ has a unique invariant measure in $\mathcal{P}(\mathcal{S})$, whose support is all $\mathcal{S}$. Such an ergodic measure will be denoted by $\boldsymbol{\mu}_\Pi$. We will also assume that

$$\lambda := \min\left\{P_Y(y|x,s)\,|\, x\in\mathcal{X}, y\in\mathcal{Y}, s\in\mathcal{S}\right\} > 0\,. \qquad (2)$$

Assumption (2) can be relaxed but we won't deal with the general case here.

For each state $s$, the family $\{P_Y(\,\cdot\,|x,s), x\in\mathcal{X}\}$ describes a DMC: we use the notation

$$C_s := \max_{\boldsymbol{u}\in\mathcal{P}(\mathcal{X})}\left\{\sum_{x\in\mathcal{X}}\boldsymbol{u}(x)\sum_{y\in\mathcal{Y}}P_Y(y|x,s)\log\frac{P_Y(y|x,s)}{\sum_{z\in\mathcal{X}}\boldsymbol{u}(z)P_Y(y|z,s)}\right\}$$

for its capacity (notice that the optimizing distribution depends on $s$), and

$$C^1_s := \max_{x',x''\in\mathcal{X}}\left\{D\big(P_Y(\,\cdot\,|x',s)||\,P_Y(\,\cdot\,|x'',s)\big)\right\}$$

for the Kullback-Leibler divergence between the pair of its most distinguishable inputs. Notice that (2) implies that $C^1_s$ is finite for every state $s$. We will use the compact notation

$$\boldsymbol{C} = (C_s)\in[0,\log|\mathcal{X}|]^{\mathcal{S}}\,, \qquad \boldsymbol{C}^1 = (C^1_s)\in[0,+\infty)^{\mathcal{S}}\,.$$

We define

$$C := \langle\boldsymbol{\mu}_\Pi, \boldsymbol{C}\rangle\,. \qquad (3)$$

It is known that Assumpton 2 allows to conclude that $C$ defined above is actually the capacity of the Markov channel we are considering when the channel state is causally known both at the encoder and at the decoder, with and without output feedback.

We consider the quantity

$$C^1 := \langle\boldsymbol{\mu}_\Pi, \boldsymbol{C}^1\rangle\,, \qquad (4)$$

which we shall refer to as the Burnashev coefficient of the channel.

Notice that, when the state space is trivial (i.e. $|\mathcal{S}|=1$), $C$ reduces to the usual notion of capacity of a DMC, while $C^1$ reduces to the coefficient originally introduced by Burnashev in [3].

### C. Causal feedback encoders and sequential decoders

We now introduce the class of coding schemes we shall consider in this paper.

**Definition 3** *A* causal feedback encoder *is the pair of a finite message set and a sequence of maps*

$$\Phi = \left(\mathcal{W}, \{\phi_t : \mathcal{W}\times\mathcal{Y}^{t-1}\times\mathcal{S}^t\to\mathcal{X}\}_{t\in\mathbb{N}}\right)\,. \qquad (5)$$

With Def.3, we are implicitly assuming that perfect state knowledge as well as perfect output feedback are causally available at the encoder side.

Given a stationary Markov channel and a causal feedback encoder as in Def.3, we will consider a $\mathcal{W}$-valued random variable $W$ describing the message to be transmitted, a sequence $\boldsymbol{X} = (X_t)_{t\in\mathbb{N}}$ of $\mathcal{X}$-valued r.v.s (the channel input process), a sequence $\boldsymbol{Y} = (Y_t)_{t\in\mathbb{N}}$ of $\mathcal{Y}$-valued r.v.s (the channel output process), and a sequence $\boldsymbol{S} = (S_t)_{t\in\mathbb{N}}$ $\mathcal{S}$-valued r.v.s (the state process). We consider the time ordering

$$W, S_1, X_1, Y_1, S_2, X_2, Y_2, \dots\,,$$

and assume that the joint distribution of $W$, $\boldsymbol{X}$, $\boldsymbol{Y}$ and $\boldsymbol{S}$ is described by

$$\mathbb{P}(W=w) = 1/|\mathcal{W}|\,, \qquad \mathbb{P}(S_1 = s|W) = \boldsymbol{\mu}(s)\,,$$

$$\mathbb{P}(S_t = x|\,W, \boldsymbol{S}_1^{t-1}, \boldsymbol{X}_1^{t-1}, \boldsymbol{Y}_1^{t-1}) = P_S\left(s|S_{t-1}\right)\,,$$

$$\mathbb{P}(X_t = x|\,W, \boldsymbol{S}_1^{t}, \boldsymbol{X}_1^{t-1}, \boldsymbol{Y}_1^{t-1}) = \delta_{\{\phi_t(W, Y_1^{t-1}, S_1^t)\}}(x)\,,$$

$$\mathbb{P}(Y_t = y|\,W, \boldsymbol{S}_1^{t}, \boldsymbol{X}_1^{t}, \boldsymbol{Y}_1^{t-1}) = P_Y\left(y|S_t, X_t\right)\,.$$

The corresponding expectation operator will be denoted by $\mathbb{E}$.

We observe that the absence of ISI ensures that the state process $\boldsymbol{S}$ is independent of the transmitted message $W$. In particular $\boldsymbol{S}$ forms a stationary Markov chain with irreducible transition probability matrix $\Pi$. This implies that the empirical measure process $(\boldsymbol{\upsilon}_{\mathcal{S}}(\boldsymbol{S}_1^n))_{n\in\mathbb{N}}$ converges $\mathbb{P}$-almost surely to the invariant measure $\boldsymbol{\mu}_\Pi$, while satisfying a large deviations principle (see [4]) with convex, good rate function $I_\Pi(\boldsymbol{\theta})$.

**Definition 4** *A sequential decoder for a causal feedback encoder* $\Phi$ *as in (5) is a pair* $\Psi = (T, \psi)$*, where* $T$ *is a stopping time for the process* $(\boldsymbol{S}, \boldsymbol{Y})$ *and* $\psi$ *an* $\mathcal{W}$*-valued* $(\boldsymbol{S}_1^T, \boldsymbol{Y}_1^T)$*-measurable random variable.*

Notice that with Def.4 we are assuming that the state sequence is causally observable at the decoder side.

Given a causal feedback encoder $\Phi$ as in Def. 3 and a sequential decoder $\Psi$ as in Def. 4, their error probability is

$$p_e(\Phi, \Psi) := \mathbb{P}\left(\psi \neq W\right)\,.$$

Following Burnashev's approach we shall consider the expected decoding time $\mathbb{E}[T]$ as a measure of the delay of the coding scheme and accordingly define its rate by

$$R := \log|\mathcal{W}|/\mathbb{E}[T]\,.$$

### D. Main result

We are ready to state our main result. It is formulated in an asymptotic setting, considering sequences of causal encoders and sequential decoders with asymptotic average rate below capacity and vanishing error probability.

**Theorem 5** *For any* $0 < R < C$*:*

1) *any sequence* $(\Phi^n, \Psi^n)$ *of causal encoders sequential decoder pairs such that*

$$\lim_{n\in\mathbb{N}} p_e(\Phi^n, \Psi^n) = 0\,, \qquad \limsup_{n\in\mathbb{N}} \frac{\log|\mathcal{W}^n|}{\mathbb{E}[T^n]} \geq R\,, \quad (6)$$

*satisfies*

$$\limsup_{n\in\mathbb{N}} -\frac{1}{\mathbb{E}[T^n]}\log p_e\left(\Phi^n, \Psi^n\right) \leq E_B(R)\,. \quad (7)$$

2) *there exists a sequence* $(\Phi^n, \Psi^n)_{n\in\mathbb{N}}$ *of encoder-decoder pairs satisfying (6) and such that*

$$\lim_{n\in\mathbb{N}} -\frac{1}{\mathbb{E}[T^n]}\log p_e\left(\Phi^n, \Psi^n\right) = E_B(R)\,. \quad (8)$$

We observe that Burnashev's original result [3] for DMCs can be recovered as a particular case of Theorem 5 when the state space is trivial.

## III. An upper bound on the error exponent

The aim of this section is to sketch the proof of Part 1 of Theorem 5. We shall present a series of partial results whose proofs will be given in full detail elsewhere. Our results are extensions of those in [3]. In particular we have followed the approach proposed in [1] (see also [9],[5]) trying to emphasize the emergence of two distinct transmission phases, a communication one related to the channel capacity, and a binary hypothesis testing one related to the Burnashev coefficient.

### A. A generalized Fano's inequality

A first observation is that, given a stopping time $T$ for the process $(\boldsymbol{S}, \boldsymbol{Y})$, the optimal decoder $\psi : (\mathcal{S}\times\mathcal{Y})^T \to \mathcal{W}$ which can be associated to it is the maximum a posteriori (MAP) one. Thus, in order to lower bound the error probability we can restrict ourselves to MAP sequential decoders. Given channel output and state observations up to some time $t$, we denote the a posteriori error probability by

$$p_{MAP}\,(t) := 1 - \max_{w\in\mathcal{W}} \left\{\mathbb{P}\left(W = w \middle| \boldsymbol{Y}_1^t,\, \boldsymbol{S}_1^t\right)\right\}\,.$$

Notice that the quantity defined above implicitly depends on the encoder $\Phi$. Since $W$ is uniformly distributed over the message set $\mathcal{W}$, we have that $p_{MAP}(0) = (|\mathcal{W}|-1)/|\mathcal{W}|$. Moreover it is not difficult to prove the following recursive lower bound.

**Lemma 6** *Given any causal feedback encoder* $\Phi$*, for any* $t \geq 0$

$$p_{MAP}(t+1) \geq \lambda p_{MAP}(t) \qquad \mathbb{P}-a.s.$$

For every $\delta > 0$ we introduce the following stopping time for the process $(\boldsymbol{S}, \boldsymbol{Y})$

$$\tau_\delta := \inf\left\{t\in\mathbb{N} :\, t \geq T \text{ or } p_{MAP}(t) \leq \delta\right\}\,. \quad (9)$$

The following result can be proved using Fano's inequality and Doob's optional stopping theorem.

**Lemma 7** *For any* $0 < \delta < \frac{1}{2}$ *we have*

$$\mathbb{E}\left[\sum_{t=1}^{\tau_\delta} C_{S_t}\right] \geq \left(1 - \delta - \frac{p_e\left(\Phi, \Psi\right)}{\delta}\right)\log|\mathcal{W}| - \mathrm{H}(\delta)\,. \quad (10)$$

### B. A lower bound to the error probability of a composite binary hypothesis test

We now consider a particular binary hypothesis testing problem which will arise while proving the main result. Consider a nontrivial binary partition of the message set

$$\mathcal{W} = \mathcal{W}_0 \cup \mathcal{W}_1\,, \qquad \mathcal{W}_0 \cap \mathcal{W}_1 = \emptyset\,, \qquad \mathcal{W}_0, \mathcal{W}_1 \neq \emptyset\,, \quad (11)$$

and a sequential binary hypothesis test $\tilde{\Psi} = (T, \tilde{\psi})$ (where $T$ is a stopping time for $(\boldsymbol{S}, \boldsymbol{Y})$ and $\psi : (\mathcal{S}\times\mathcal{Y})^T \to \{0,1\}$) between the hypotheses $\{W \in \mathcal{W}_0\}$ and $\{W \in \mathcal{W}_1\}$.

Consider now another stooping time $\tau$ for $(\boldsymbol{S}, \boldsymbol{Y})$, satisfying

$$\tau \leq T \qquad \text{a.s.} \quad (12)$$

Suppose that $\mathcal{W}_1$ is a $(\boldsymbol{S}_1^\tau, \boldsymbol{Y}_1^\tau)$-measurable random subset of the message set $\mathcal{W}$. The following lower bound to the error probability of the binary test $\tilde{\Psi}$ conditioned on $(\boldsymbol{S}_1^\tau, \boldsymbol{Y}_1^\tau)$ can be proved using the log-sum inequality and Doob's optional sampling theorem.

**Lemma 8** *Let* $\Phi$ *be any causal encoder, and* $\tau$ *and* $T$ *stopping times for the process* $(\boldsymbol{S}, \boldsymbol{Y})$ *satisfying (12). Then, for every* $(\boldsymbol{S}_1^\tau, \boldsymbol{Y}_1^\tau)$*-measurable random message subset* $\mathcal{W}_1$

$$\mathbb{E}\left[\sum_{t=\tau+1}^{T} C^1_{S_t}\middle| \boldsymbol{S}_1^\tau, \boldsymbol{Y}_1^\tau\right] \geq -\log\frac{\mathbb{P}\left(\tilde{\psi} \neq \mathbb{1}_{\{W\in\mathcal{W}_1\}}\middle| \boldsymbol{S}_1^\tau, \boldsymbol{Y}_1^\tau\right)}{Z/4} \quad (13)$$

$\mathbb{P}$*-a.s., where* $Z := \min_{i=0,1}\left\{\mathbb{P}\left(W\in\mathcal{W}_i\middle|\, \boldsymbol{S}_1^\tau, \boldsymbol{Y}_1^\tau\right)\right\}$.

### C. Burnashev bound for Markov channels

Combining Lemmas 6, 7 and 8, it is possible to prove that the following.

**Theorem 9** *Given a causal feedback encoder* $\Phi = (\mathcal{W}, (\phi_t))$ *and a sequential decoder* $\Psi = (T, \psi)$*, for every* $\delta > 0$

$$\begin{aligned}\frac{C^1}{C}\mathbb{E}\left[\sum_{t=1}^{\tau_\delta} C_{S_t}\right] + \mathbb{E}\left[\sum_{t=\tau_\delta+1}^{T} C^1_{S_t}\right]& \\ \geq -\log p_e\left(\Phi, \Psi\right) + \frac{C^1}{C}\log|\mathcal{W}|\left(1-\alpha\right) + \beta\,,&\end{aligned} \quad (14)$$

*where* $\tau_\delta$ *is defined by (9), and*

$$\alpha := \delta + \frac{p_e(\Phi, \Psi)}{\delta}\,, \qquad \beta := \log\frac{\lambda\delta}{4} - \frac{C^1}{C}\mathrm{H}(\delta)\,.$$

When the state space is trivial, the lefthand side of (14) equals $C^1\mathbb{E}[T]$, so that (14) reduces to Burnashev's inequality (4.1) in [3]. For nontrivial state space, in order to obtain the single-letter bound of Theorem 5, it is necessary to consider a countable family of causal encoders $(\Phi^k)$ and a corresponding family of sequential decoders $(\Psi^k)$ satisfying (6).

The core idea for proving (7) consists in introducing a positive real sequence $(\delta_k)$, and showing that both the random variables

$$\tau^k := \inf\left\{t \in \mathbb{N} \middle| \, t \geq T^k \text{ or } p_{MAP}(t) \leq \delta_k\right\},$$

and $T^k - \tau^k$ 'diverge' in the probabilistic sense made precise by (16) below. The sequence $(\delta_k)$ needs to be properly chosen: we want it to be asymptotically vanishing in order to guarantee that $\tau^k$ diverges, but not too fast since otherwise $T^k - \tau^k$ would not diverge. It turns out that one possible good choice is $\delta_k := \frac{-1}{\log p_e(\Phi^k, \Psi^k)}$. From (2) it follows that

$$\lim_{k\in\mathbb{N}} \delta_k = 0, \qquad \lim_{k\in\mathbb{N}} \frac{p_e\left(\Phi^k, \Psi^k\right)}{\delta_k} = 0. \tag{15}$$

The following result can be proven using Lemma 6.

**Lemma 10** *In the previous setting, for every $M$ in $\mathbb{N}$, we have*

$$\lim_{k\in\mathbb{N}} \mathbb{P}\left(\tau^k \leq M\right) = 0, \qquad \lim_{k\in\mathbb{N}} \mathbb{P}\left(T^k - \tau^k \leq M\right) = 0. \tag{16}$$

We have already noticed that the irreducibility of the stochastic matrix $\Pi$ implies that the empirical measure process associated to the state sequence $\boldsymbol{S}$ converges $\mathbb{P}$-almost surely to the invariant measure $\boldsymbol{\mu}_\Pi$. This fact, together with (16), allows us to prove the following.

**Lemma 11** *In the previous setting*

$$\lim_{k\in\mathbb{N}} \frac{1}{\mathbb{E}[\tau^k]} \mathbb{E}\left[\sum_{t=1}^{\tau^k} C_{S_t}\right] = C, \tag{17}$$

*and*

$$\lim_{k\in\mathbb{N}} \frac{1}{\mathbb{E}[T^k - \tau^k]} \mathbb{E}\left[\sum_{t=\tau^k+1}^{T^k} C^1_{S_t}\right] = C^1. \tag{18}$$

By taking the limit of both sides of (14) and substituting (17) and (18), we get (7).

## IV. An asymptotically optimal scheme

In order to prove Part 2 of Theorem 5, thus showing the achievability of the Burnashev's exponent $E_B(R)$, we propose an iterative transmission scheme consisting in a generalization of the one introduced by Yamamoto and Itoh [10] for DMCs. This scheme consists of a sequence of epochs. Each epoch is made up of two distinct transmission phases, respectively named communication and confirmation phase. In the communication phase the message to be sent is encoded in a block code and transmitted over the channel. At the end of this phase the decoder makes a tentative decision about the message sent based on the observation of the channel outputs and of the state sequence. As perfect feedback is available, the result of this decision is known at the encoder. In the confirmation phase a binary acknowledge message, confirming the decoder's estimation if it is correct, or denying it when it is wrong, is sent by the encoder through a fixed-length repetition code. The decoder performs a binary hypothesis test in order to decide whether a deny or a confirmation message has been sent. If a confirmation is detected the transmission halts, while if a deny is detected the system restarts transmitting the same message with the same protocol. Again because of perfect feedback availability at the encoder, there are no synchronization problems.

More precisely we design our scheme as follows. Given a design rate $R$ in $(0, C)$, let us fix an arbitrary $\gamma$ in $(\frac{R}{C}, 1)$. For every $n$ in $\mathbb{N}$, consider a message set $\mathcal{W}_n$ of cardinality $|\mathcal{W}_n| = \exp(\lfloor nR \rfloor)$ and two blocklengths $\hat{n}$ and $\tilde{n}$ respectively defined as $\hat{n} = \lceil n\gamma \rceil$, $\tilde{n} := n - \hat{n}$.

### A. Fixed-length coding for the transmission phase

It is known that $C$ equals the capacity of the stationary Markov channel we are considering. This implies that, since $\lim_{n\in\mathbb{N}} \frac{\log|\mathcal{W}_n|}{\hat{n}} = \frac{R}{\gamma} < C$, there exists a sequence of causal encoders with no output feedback $\{\hat{\phi}^n_t : \mathcal{W}_n \times \mathcal{S}^t \to \mathcal{X}\}$, and a sequence of decoders of fixed length $\hat{n}$ $\{\hat{\psi}^n : \mathcal{S}^{\hat{n}} \times \mathcal{Y}^{\hat{n}} \to \mathcal{W}_n\}$ with error probability asymptotically vanishing with the blocklength. More precisely, the error probability of the pair $(\hat{\Phi}^n, \hat{\Psi}^n)$ goes to zero uniformly with respect to initial state distribution. Thus, denoting by $p(n)$ the maximum over all initial state distributions $\boldsymbol{\mu}$ of the error probability of the pair $(\hat{\phi}^n, \hat{\Psi}^n)$, we have that $\lim_{n\in\mathbb{N}} p(n) = 0$. The pair $(\hat{\Phi}^n, \hat{\Psi}^n)$ will be used in the first phase of each epoch of our iterative transmission scheme.

### B. Binary hypothesis test for the confirmation phase

For the second phase we consider a causal repetition encoder $\tilde{\Phi}^n$ of length $\tilde{n}$, defined by $\tilde{\phi}^n_t(m) = x^m_{s_t}$ for $m = 0, 1$, where for every state $s$ we denote by $x^0_s$ and $x^1_s$ one of the most distinguishable input pairs for the $s$-th channel, i.e. such that $C^1_s = D(P(\cdot|x^0_s)||P(\cdot|x^1_s))$.

Suppose that an acknowledge message $m = 0$ is sent. Then it is easy to verify that the pair sequence $(S_t, Y_t)^{\tilde{n}}_{t=1}$ forms a Markov chain with state space $\mathcal{S} \times \mathcal{Y}$ and transition probabilities $P_0(v, y|s, z) := P_S(v|s) P_Y(y|s, x^0_s)$. Analogously, if a deny message $m = 1$ has been sent, with transition probabilities $P_1(v, y|s, z) := P_S(v|s) P_Y(y|s, x^1_s)$. It can be checked that Assumption 2 and (2) guarantee that both the stochastic matrices $\Pi_i := (P_i(v, y|s, z))$, $i = 0, 1$, are irreducible, with invariant measures given by

$$\boldsymbol{\mu}_i(s, y) = \boldsymbol{\mu}(s) P_Y(y|x^i_s, s).$$

Notice that we can rewrite

$$C^1 = \sum_{s\in\mathcal{S}} \boldsymbol{\mu}(s) D\left(P_Y(\cdot|x^0_s, s)||P_Y(\cdot|x^0_s, s)\right) = D(\boldsymbol{\mu}_0||\boldsymbol{\mu}_1).$$

Using binary hypothesis test results generalizing Stein lemma to irreducible Markov chains (see [6], [4]), it is possible to show that $C^1$ is achievable as type one error exponent, while satisfying the constraint of vanishing type zero error probability. In fact, for every $n$ in $\mathbb{N}$ define the decoder $\tilde{\Psi}^n : \mathcal{S}^{\tilde{n}} \times \mathcal{Y}^{\tilde{n}} \to \{0, 1\}$,

$$\tilde{\Psi}^n(\boldsymbol{y}, \boldsymbol{s}) = \begin{cases} 1 & \text{if} \quad I_{\Pi_0}(\boldsymbol{v}_{\mathcal{S}\times\mathcal{Y}}(\boldsymbol{s}, \boldsymbol{y})) > \alpha_n \\ 0 & \text{if} \quad I_{\Pi_0}(\boldsymbol{v}_{\mathcal{S}\times\mathcal{Y}}(\boldsymbol{s}, \boldsymbol{y})) \leq \alpha_n, \end{cases}$$

where $I_{\Pi_0}$ is the large deviations rate function associated to the stochastic matrix $\Pi_0$ ([4]), and $(\alpha_n)$ is a real positive sequence satisfying

$$\lim_{n\in\mathbb{N}} \alpha_n = 0\,, \qquad \lim_{n\in\mathbb{N}} \frac{1}{n}\log\alpha_n = 0\,.$$

Defining $p_0(n)$ (respectively $p_1(n)$) as the maximum over all possible initial state distributions $\boldsymbol{\mu}$ of the error probability of the pair $(\tilde{\Phi}^n, \tilde{\Psi}^n)$ conditioned on the transmission of a $'0'$ $('1')$ message, it is possible to show that

$$\lim_{n\in\mathbb{N}} p_0(n) = 0\,, \qquad \lim_{n\in\mathbb{N}} -\frac{1}{\tilde{n}}\log p_1(n) = C^1\,.$$

### C. Performance of the proposed scheme

Once chosen the encoder decoder pairs $(\hat{\Phi}^n, \hat{\Psi}^n)$ and $(\tilde{\Phi}^n, \tilde{\Psi}^n)$ for the communication and the confirmation phase respectively, the iterative protocol described at the beginning of this section defines a causal encoder $\Phi^n = (\mathcal{W}_n, (\phi_t^n))$ and a sequential decoder $\Psi^n = (T^n, \psi^n)$. It can be verified that the error probability of such a scheme satisfies

$$p_e(\Phi^n, \Psi^n) \leq p_1(n)\,,$$

while its decoding time is dominated by a scaled geometric random variable:

$$\mathbb{P}(T^n > kn) \leq (p_0(n) + p(n))^k\,, \qquad k \geq 0\,. \tag{19}$$

As a consequence we have that

$$\lim_{n\in\mathbb{N}} \frac{\log|\mathcal{W}_n|}{\mathbb{E}[T^n]} = R\,,$$

and

$$\lim_{n\in\mathbb{N}} \frac{-\log p_e(\Phi^n, \Psi^n)}{\mathbb{E}[T^n]} = C^1(1-\gamma)\,.$$

Thus (8) can be deduced from the arbitrariness of $\gamma$ in $(\frac{R}{C}, 1)$.

We emphasize the fact that the two phases in each epoch of the scheme are of fixed length, while the number of epochs is possibly variable. Hence, the overall transmission length of the scheme is variable. However, (19) guarantees that with high probability the transmission halts after the first epoch, a property making this scheme appealing for practical implementation, as already noticed in [5]. Moreover we observe that the first transmission phase only requires an asymptotically vanishing error probability, not necessarily at an exponential rate.

## V. Conclusions and open problems

In this paper we have determined the reliability function for Markov channels with feedback and variable length channel codes. In the future we will relax the assumption that there is no ISI, and that both the transmitter and the receiver have access to the state. In the former case controlled Markov chain theory will be used. In the latter case output feedback will have a dual role: to estimate the state and to increase the reliability function. Extensions to noisy feedback will also be considered.

## References


[1] P. Berlin, B. Nakiboglu, B. Rimoldi, E. Telatar, "A Simple Derivation of Burnashev's Reliability function", http://arxiv.org/abs/cs/0610145, 2006.

[2] V. S. Borkar, *Probability Theory: An Advanced Course*, Springer, 1995.

[3] M. V. Burnashev, "Data Transmission over a Discrete Channel with Feedback. Random Transmission Time", *Prob. of Inform. Transm.*, vol. 12 (4), pp. 1030, 1976.

[4] A. Dembo, O. Zeitouni, *Large Deviations Techniques and Applications*, 2nd ed., Springer, 1998.

[5] B. Nakiboglu, R. G. Gallager, "Error exponents for variable-length block codes with feedback and cost constraints", preprint, av. at http://arxiv.org/abs/cs.IT/0612097, 2006.

[6] S. Natarajan, Large deviations, hypotheses testing, and source coding for finite Markov chains, *IEEE Trans. Inform. Theory*, vol. 31, pp. 360-365, 1985.

[7] C. E. Shannon "The zero error capacity of a noisy channel", *IEEE Trans. Inform. Theory*, vol. 2, pp. 8-19, 1956.

[8] A. Tchamkerten, I. E. Telatar, "On the Universality of Burnashev's Error Exponent", *IEEE Trans. Inform. Theory*, vol.51, pp. 2940-2944, 2005.

[9] A. Tchamkerten, I. E. Telatar, "Variable Length Coding over an Unknown Channel", *IEEE Trans. Inform. Theory*, vol.52, pp. 2126-2145, 2006.

[10] H. Yamamoto, K. Itoh, "Asymptotic performance of a modified Schalkwijk-Barron scheme for channels with noiseless feedback, *IEEE Trans. Inf. Theory*, vol. 25, pp. 729-733, 1979.