On the Universality of Burnashev's Error Exponent

Aslan Tchamkerten and İ. Emre Telatar
Laboratoire de Théorie de l'Information
Ecole Polytechnique Fédérale de Lausanne
CH-1015, Switzerland
Email: {firstname.lastname}@epfl.ch

Abstract— **We consider communication over a time invariant discrete memoryless channel with noiseless and instantaneous feedback. We assume that the communicating parties are not aware of the underlying channel, however they know that it belongs to some specific family of discrete memoryless channels. Recent results [4] show that for certain families (e.g., binary symmetric channels and Z channels) there exists coding schemes that universally achieve any rate below capacity while attaining Burnashev's error exponent. We show that this is not the case in general by deriving an upper bound to the universally achievable error exponent.**

I. Introduction

Burnashev [1] proved that, given a discrete memoryless channel (DMC) Q with noiseless and instantaneous (causal) feedback, and with finite input and output alphabets $\mathcal{X}$ and $\mathcal{Y}$, the maximum achievable error exponent is given by

$$E_B(R,Q) \triangleq \left(\max_{(x,x')\in\mathcal{X}\times\mathcal{X}} D(Q(\cdot|x)\|Q(\cdot|x'))\right)\left(1-\frac{R}{C(Q)}\right) \quad (1)$$

where

$$D(Q(\cdot|x)\|Q(\cdot|x')) \triangleq \sum_{y\in\mathcal{Y}} Q(y|x)\ln\frac{Q(y|x)}{Q(y|x')}$$

is the Kullback-Liebler distance[1] between the output distributions induced by the input letters x and x', and where R and $C(Q)$ denote the rate and the channel capacity. From now on $E_B(R,Q)$ will be referred as the Burnashev's error exponent.

Suppose now that the DMC under use is revealed neither to the transmitter nor to the receiver but that it is known that the channel belongs to some specific set $\mathcal{Q}$ of DMCs. Does Burnashev's result still hold? In other words can one design a coding scheme that asymptotically (as the decoding delay tends to infinity) yields the error exponent (1) simultaneously on all channels in $\mathcal{Q}$? A partial answer is provided in [4] for the family of Binary Symmetric Channels (BSCs) with crossover probability $\varepsilon \in [0, L]$ and with $L \in [0, 1/2)$. Given any $\gamma \in [0, 1)$ there exists coding schemes that achieve simultaneously over that family a rate guaranteed to be at least γ times the channel capacity, and with a corresponding maximum error exponent, i.e., equal to (1). Similarly, if one now is interested in having a low error probability instead of a high communication rate, there exists coding schemes that universally achieve a rate guaranteed to be at most γ times the channel capacity, and with a corresponding error exponent that is also maximum. A similar results holds for the class of Z channels with crossover probability $\varepsilon \in [0, L]$ and with $L \in [0, 1)$. In [4] it is shown that, given any $\gamma \in [0, 1)$, there exists coding schemes that simultaneously reach the maximum error exponent at a rate equal to γ times the channel capacity. In other words, for BSCs and Z channels it is possible to achieve Burnashev's error exponent universally while having a certain control on the rate.

In this paper we consider the possibility of extending the results in [4] to arbitrary family of channels, such as for instance the set of all binary input channels with some finite output alphabet. We show that, under some conditions on a pair of channels Q_1 and Q_2, no zero-rate coding scheme achieves the Burnashev's exponent simultaneously on both Q_1 and Q_2. Therefore the results obtained in [4] cannot be generalized to arbitrary families of channels: in general, given a family of DMCs, Burnashev's error exponent is not universally achievable at all rates below capacity.

II. Preliminaries and Statement of Result

We first remind the definitions of coding schemes, rate, error probability and error exponent for a DMC with perfect feedback. Then we state a theorem of Burnashev, present our result, and a sketch of its proof.

Definition 1 (Coding Scheme): Given two finite alphabets $\mathcal{X}$ and $\mathcal{Y}$ and a message set $\mathcal{M}$ of size $M \geq 1$, an encoder (or codebook) is a sequence of functions

$$\mathcal{C}^M = \{X_n : \mathcal{M}\times\mathcal{Y}^{n-1} \longrightarrow \mathcal{X}\}_{n\geq 1}\,. \quad (2)$$

The symbol x_n to be sent at time n is obtained by evaluating X_n for the message and the feedback sequence received so far, i.e., $x_n \triangleq X_n(m, y^{n-1})$ where $y^{n-1} \triangleq y_1, y_2, \ldots, y_{n-1}$. A codeword for message m is the sequence of functions

A journal version of the material presented in this proceeding has been accepted for publication in the IEEE Transactions on Information Theory. This work was supported (in part) by the National Competence Center in Research on Mobile Information and Communication Systems (NCCR-MICS), a center supported by the Swiss National Science Foundation under grant number 5005-67322.

[1] ln denotes the logarithm to the base e.

$\{X_n(m,\,\cdot\,)\}_{n\geq 1}$. A decoder (Ψ^M, T) consists of a sequence of functions

$$\Psi^M = \{\psi_n^M : \mathcal{Y}^n \longrightarrow \mathcal{M}\}_{n\geq 1}\,, \tag{3}$$

and a stopping time $T(M)$, with respect to to the received symbols $Y_1, Y_2, \ldots$,[2]that represents the decision time. The decoded message is given by $\psi^M_{T(M)}(y^{T(M)})$. A coding scheme is a tuple $\mathcal{S}^M = (\mathcal{C}^M, \Psi^M, T(M))$.

Definition 2 (Rate): For a channel Q, an integer $M \geq 1$ and a coding scheme $\mathcal{S}^M = (\mathcal{C}^M, \Psi^M, T(M))$, the average rate is

$$R(\mathcal{S}^M, Q) \triangleq \frac{\ln M}{\mathbb{E}T(M)} \text{ nats per symbol}\,, \tag{4}$$

where $\mathbb{E}T(M)$ denotes the expected decision time over uniformly chosen messages, i.e.,

$$\mathbb{E}T(M) \triangleq \frac{1}{M}\sum_{m\in\mathcal{M}} \mathbb{E}\left(T(M) \;\middle|\; \text{message } m \text{ is sent}\right)\,. \tag{5}$$

The asymptotic rate for a sequence of coding schemes $\theta = \{\mathcal{S}^M\}_{M\geq 1}$ and a given channel Q is

$$R(\theta, Q) \triangleq \lim_{M\to\infty} R(\mathcal{S}^M, Q) \tag{6}$$

whenever the limit exists.

Definition 3 (Error Probability): The average (over uniformly chosen messages) error probability given a coding scheme $\mathcal{S}^M$ and a channel Q is defined as

$$\mathbb{P}(\mathcal{E}|Q, \mathcal{S}^M) = \frac{1}{M}\sum_{m\in\mathcal{M}} \mathbb{P}\left(\psi^M_{T(M)}(Y^{T(M)}) \neq m \;\middle|\; \text{message } m \text{ is sent}\right)\,. \tag{7}$$

Let us denote by θ a particular sequence of coding schemes $\{\mathcal{S}^M\}_{M\geq 1}$, and by Θ the set of all sequences of coding schemes.

Definition 4 (Error Exponent): Given a channel Q and a sequence of coding schemes $\theta = \{\mathcal{S}^M\}_{M\geq 2} = \{(\mathcal{C}^M, \Psi^M, T(M))\}_{M\geq 1}$ such that $\mathbb{P}(\mathcal{E}|Q, \mathcal{S}^M) \to 0$ as $M \to \infty$, the error exponent is

$$E(\theta, Q) \triangleq \liminf_{M\to\infty} -\frac{1}{\mathbb{E}T(M)} \ln \mathbb{P}(\mathcal{E}|Q, \mathcal{S}^M)\,. \tag{8}$$

We now state an important result related to the error exponent of DMCs with perfect feedback:

Theorem 1 (Burnashev 1976 [1]): Let Q be a DMC with capacity $C(Q)$. For any $R \in [0, C(Q)]$ and any $\theta = \{\mathcal{S}^M\}_{M\geq 1} \in \Theta$ such that $R(\theta, Q) = R$,

$$\limsup_{M\to\infty} -\frac{1}{\mathbb{E}T(M)} \ln \mathbb{P}(\mathcal{E}|Q, \mathcal{S}^M) \leq E_B(R, Q)\,. \tag{9}$$

Moreover there exists $\theta \in \Theta$ such that $R(\theta, Q) = R$ and $E(\theta, Q) = E_B(R, Q)$.

[2]An integer-valued random variable U is said to be a stopping time with respect to $Y_1, Y_2, \ldots$ if, given $Y_1, Y_2, \ldots, Y_n$, the event $\{T = n\}$ is independent of $Y_{n+1}, Y_{n+2}, \ldots$ for all $n \geq 1$.

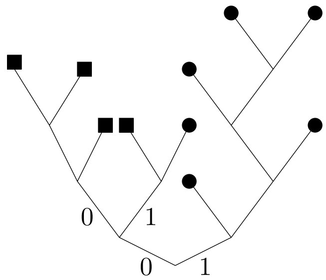

Fig. 1. Example of a "decision tree" for a binary output channel. The square leaves correspond to message A whereas the round leaves correspond to message B.

Before we state our result, let us define the quantity $K(Q_i, Q_j)$ for any two channels Q_i and Q_j with same input alphabet $\mathcal{X}$:

$$K(Q_i, Q_j) \triangleq \max_{(x,x')\in\mathcal{X}\times\mathcal{X}} [D(Q_i(\cdot|x)\|Q_i(\cdot|x')) + D(Q_i(\cdot|x)\|Q_j(\cdot|x'))]\,. \tag{10}$$

Theorem 2: Let Q_1 and Q_2 be two DMCs on $\mathcal{X}\times\mathcal{Y}$ such that for $(i,j) \in \{(1,2),(2,1)\}$

$$K(Q_i, Q_j) < 2 \max_{(x,x')\in\mathcal{X}\times\mathcal{X}} D(Q_i(\cdot|x)\|Q_i(\cdot|x'))\,. \tag{11}$$

For any $\theta \in \Theta$, either $E(\theta, Q_1) < E_B(0, Q_1)$ or $E(\theta, Q_2) < E_B(0, Q_2)$.

A simple example of channels Q_1 and Q_2 that satisfy the assumptions of Theorem 2 is given by $Q_1 = \text{BSC}(\varepsilon)$ and $Q_2 = \text{BSC}(1-\varepsilon)$ where $0 < \varepsilon < 1/2$. In this case we have

$$\begin{aligned} K(Q_1, Q_2) &= \max_{(x,x')\in\mathcal{X}\times\mathcal{X}} D(Q_1(\cdot|x)\|Q_1(\cdot|x')) \\ &= \max_{(x,x')\in\mathcal{X}\times\mathcal{X}} D(Q_2(\cdot|x\|Q_2(\cdot|x')) \\ &= K(Q_2, Q_1)\,. \end{aligned} \tag{12}$$

From Theorem 2 we conclude that, given a family of DMCs $\mathcal{Q}$, in general no zero-rate coding scheme achieves Burnashev's error exponent universally over $\mathcal{Q}$. Therefore the property of the families of BSCs and Z channels that was shown in [4] is not true for an arbitrary class of channels: even with perfect feedback, the fact that the channel is unknown may result in an error exponent smaller than the best error exponent that could be obtained if the channel were revealed to both the transmitter and the receiver [1].

Sketch of the Proof of Theorem 2:[3] The theorem is proved by deriving an upper bound on the maximum achievable error exponent for two-message coding schemes, and using the fact that the zero-rate error exponent is upper bounded by the error exponent for a fixed number of messages.

Given a decoder of a two-message coding scheme for a channel Q with output alphabet $\mathcal{Y}$, the set of all output sequences for which a decision is made can be represented by the leaves of a complete $|\mathcal{Y}|$-ary tree. The set of leaves is divided into two sets that correspond to declaring message

[3]The proof of Theorem 2 can be found in [5].

A and B. The decoder starts climbing the tree from the root. At each time it chooses the branch that corresponds to the received symbol. When a leaf is reached the decoder makes a decision as indicated by the label of the leaf (see figure 1 for an example).

From a probabilistic point of view, given a particular coding scheme, the decision time of the decoder determines the probability space of the output sequences, equivalently the set of leaves. On this probability space, each set of encoding functions $\{X_n(m,\cdot)\}_{n\geq 1}, m \in \{A,B\}$, together with the transition probability matrix of the channel Q induces a probability measure that we denote by P_m. In other words, associated to a channel and a two-message coding scheme, there is a natural probability space with two probability measures P_A and P_B that correspond to the sending of message A and B.

Assume now that the transmitter and the receiver still communicate using a particular two-message coding scheme but neither the transmitter nor the receiver know which channel will be used: it might be either Q_1 or Q_2, both defined on the same common input and output alphabets $\mathcal{X}$ and $\mathcal{Y}$. Let $P_{m,i}$ denote the probability of the output sequence when message $m \in \{A,B\}$ is being sent through channel Q_i, $i \in \{1,2\}$. In order to decode, the receiver has to perform a statistical test for two composite hypothesis $\{P_{A,1},P_{A,2}\}$ and $\{P_{B,1},P_{B,2}\}$. From classical results in hypothesis testing [3] it is well known that the error probabilities of such a test essentially depend on "how close" the hypotheses are. More precisely, given that message m is sent through channel Q_i, one can show that the error probability behaves as $e^{-E_{m,i}}$ where $E_{m,i}$ equals to the smallest divergence term between $D(P_{m',i}||P_{m,i})$ and $D(P_{m',j}||P_{m,i})$, with $m' \neq m$ and $j \neq i$. Using a martingale argument we show that, whenever Q_1 and Q_2 satisfy the hypothesis of Theorem 2, the eight terms $D(P_{m',j}||P_{m,i})$, where $i,j \in \{1,2\}$, $m,m' \in \{A,B\}$, and with $m \neq m'$, cannot be simultaneously large. We then deduce that, under the hypothesis of Theorem 2, no two-message coding scheme yields Burnashev's exponent simultaneously on Q_1 and Q_2.

To that end, consider two probability measures P_1 and P_2 on a probability space $(\Omega,\mathcal{F})$. It is well known that unless P_1 and P_2 are singular,[4] the quantities $P_1(E)$ and $P_2(E^c)$ cannot be both rendered arbitrary small by an appropriate choice of $E \in \mathcal{F}$.[5] More specifically, from the data processing inequality for divergence,[6] we have the following lower bounds on $P_1(E)$ in terms of $P_2(E^c)$

$$P_1(E) \geq \exp\left[\frac{-D(P_2\|P_1) - H(P_2(E))}{1-P_2(E^c)}\right] \tag{15}$$

where $H(\alpha) \triangleq -\alpha\ln\alpha - (1-\alpha)\ln(1-\alpha)$.

Suppose the communicating parties use a particular two-message coding scheme $(\mathcal{C}^2,\Psi^2,T(2))$ on some known channel Q. Since we will deal only with two-message coding schemes, from now on we simply write $(\mathcal{C},\Psi,T)$ instead of $(\mathcal{C}^2,\Psi^2,T(2))$. Letting E be the set of leaves for which message A is declared, respectively the set of leaves for which message B is declared, from (15) we obtain

$$P_B(A) \geq \exp\left[\frac{-D(P_A\|P_B) - H(P_A(A))}{1-P_A(B)}\right]$$
$$\text{and} \quad P_A(B) \geq \exp\left[\frac{-D(P_B\|P_A) - H(P_B(B))}{1-P_B(A)}\right] \tag{16}$$

where $P_m(m')$ denotes the probability under P_m of the set of leaves for which message m' is declared. Note that since one is normally interested in the case where $P_B(A)$ and $P_A(B)$ are small, the terms on the right hand side of (16) are essentially $\exp[-D(P_A\|P_B)]$ and $\exp[-D(P_B\|P_A)]$.

Assume now that the transmitter and the receiver still use the two-message coding scheme $(\mathcal{C},\Psi,T)$, but that they don't know which channel will be used, it might be either Q_1 or Q_2, both defined on the same common input and output alphabets $\mathcal{X}$ and $\mathcal{Y}$. We now have four distributions on the set of leaves, namely, $P_{m,i}$ with $m \in \{A,B\}$, $i \in \{1,2\}$. There are also four error probabilities $P_{A,1}(B)$, $P_{A,2}(B)$, $P_{B,1}(A)$ and $P_{B,2}(A)$. Using (15) with $E = B$, and $(P_1,P_2) = (P_{A,1},P_{B,1}),(P_{A,1},P_{B,2}),\ldots$ we get the following inequalities:

$$P_{A,1}(B) \geq \exp\left[\frac{-D(P_{B,1}\|P_{A,1}) - H(P_{B,1}(B))}{1-P_{B,1}(A)}\right] \tag{17}$$

$$P_{A,1}(B) \geq \exp\left[\frac{-D(P_{B,2}\|P_{A,1}) - H(P_{B,2}(B))}{1-P_{B,2}(A)}\right] \tag{18}$$

$$P_{A,2}(B) \geq \exp\left[\frac{-D(P_{B,1}\|P_{A,2}) - H(P_{B,1}(B))}{1-P_{B,1}(A)}\right] \tag{19}$$

$$P_{A,2}(B) \geq \exp\left[\frac{-D(P_{B,2}\|P_{A,2}) - H(P_{B,2}(B))}{1-P_{B,2}(A)}\right]. \tag{20}$$

In similar fashion one also obtains

$$P_{B,1}(A) \geq \exp\left[\frac{-D(P_{A,1}\|P_{B,1}) - H(P_{A,1}(A))}{1-P_{A,1}(B)}\right] \tag{21}$$

$$P_{B,1}(A) \geq \exp\left[\frac{-D(P_{A,2}\|P_{B,1}) - H(P_{A,2}(A))}{1-P_{A,2}(B)}\right] \tag{22}$$

$$P_{B,2}(A) \geq \exp\left[\frac{-D(P_{A,1}\|P_{B,2}) - H(P_{A,1}(A))}{1-P_{A,1}(B)}\right] \tag{23}$$

$$P_{B,2}(A) \geq \exp\left[\frac{-D(P_{A,2}\|P_{B,2}) - H(P_{A,2}(A))}{1-P_{A,2}(B)}\right]. \tag{24}$$

A martingale argument yields

$$D(P_{B,1}\|P_{A,1}) + D(P_{B,2}\|P_{A,1})$$

[4] P_1 and P_2 are said singular if there exists $E \in \mathcal{F}$ such that $P_1(E) = 1$ and $P_2(E) = 0$.

[5] E^c denotes the complementary set of E in Ω.

[6] Let $(\Omega,\mathcal{F})$ be a probability space, let P_1 and P_2 be two probability measures on $(\Omega,\mathcal{F})$ and let $E \in \mathcal{F}$. From the data processing inequality for divergence [2, p. 55], we have

$$D(P_2\|P_1) \geq D(P_2(E)\|P_1(E)) \tag{13}$$

where

$$D(P_2(E)\|P_1(E)) \triangleq$$
$$P_2(E)\ln\frac{P_2(E)}{P_1(E)} + (1-P_2(E))\ln\frac{(1-P_2(E))}{(1-P_1(E))}. \tag{14}$$

Expanding (13) we deduce that

$$P_1(E) \geq \exp\left[\frac{-D(P_2\|P_1) - H(P_2(E))}{1-P_2(E^c)}\right]$$

where $H(P_2(E)) \triangleq -P_2(E)\ln P_2(E) - P_2(E^c)\ln P_2(E^c)$ and where E^c denotes the complementary set of E in Ω.

$$
\begin{aligned}
&+ D(P_{B,1}\|P_{A,2}) + D(P_{B,2}\|P_{A,2}) \\
&+ D(P_{A,1}\|P_{B,1}) + D(P_{A,2}\|P_{B,1}) \\
&+ D(P_{A,1}\|P_{B,2}) + D(P_{A,2}\|P_{B,2}) \\
&\quad \leq 2K(Q_1, Q_2)\mathbb{E}_1 T + 2K(Q_2, Q_1)\mathbb{E}_2 T \qquad (25)
\end{aligned}
$$

where $\mathbb{E}_i T$ denotes the expected decoding when channel Q_i is used. Now pick any sequence of two-message coding schemes that yields vanishing error probabilities $P_{A,1}(B)$, $P_{A,2}(B)$, $P_{B,1}(A)$ and $P_{B,2}(A)$. From (25), under the hypothesis that for $(i,j) \in \{(1,2),(2,1)\}$

$$
K(Q_i, Q_j) < 2 \max_{(x,x') \in \mathcal{X} \times \mathcal{X}} D(Q_i(\cdot|x)\|Q_i(\cdot|x')) \ , \qquad (26)
$$

we derive that the error exponent of this sequence of two-message coding schemes cannot be made simultaneously equal to $E_B(0, Q_1)$ on channel Q_1 and equal to $E_B(0, Q_2)$ on channel Q_2, yielding the desired result.

REFERENCES

[1] M. V. Burnashev, *Data transmission over a discrete channel with feedback: random transmission time*, Problems of Information Transmission, vol. 12, number 4, p. 250–265, 1976.

[2] I. Csiszar and J. Körner, *Information Theory*, Budapest: Akademiai Kiado, 1981.

[3] D. Siegmund, *Sequential Analysis*, New York: Springer-Verlag, 1985.

[4] A. Tchamkerten and İ. E. Telatar, *Variable length coding over an unknown channel*, submitted to the IEEE Trans. on Info. Th.

[5] A. Tchamkerten and İ. E. Telatar, *On the universality of Burnashev's error exponent*, to appear in the IEEE Trans. on Info. Th.