REFERENCES

[1] D. Falconer, S. L. Ariyavisitakul, A. Benyamin-Seeyar, and B. Eidson, "Frequency domain equalization for single-carrier broadband wireless systems," *IEEE Commun. Mag.*, vol. 40, no. 4, pp. 58–66, Apr. 2002.
[2] J. G. Proakis, *Digital Communications*, 4th ed. New York: McGraw-Hill, 2001.
[3] *Air Interface for Fixed Broadband Wireless Access Systems Part A: Systems between 2 and 11 GHz*, IEEE Std. 802.16ab-01/01, 2001.
[4] H. Sari, G. Karam, and I. Jeanclaude, "Transmission techniques for digital terrestrial TV broadcasting," *IEEE Commun. Mag.*, vol. 33, no. 2, pp. 100–109, Feb. 1995.
[5] M. O. Polley, W. F. Schreiber, and S. J. Wee, "Comments on transmission techniques for digital terrestrial TV broadcasting," *IEEE Commun. Mag.*, vol. 33, no. 11, pp. 22–26, Nov. 1995.
[6] V. Tarokh, N. Seshadri, and A. R. Calderbank, "Space-time codes for high data rate wireless communication: Performance criterion and code construction," *IEEE Trans. Inf. Theory*, vol. 44, no. 2, pp. 744–765, Mar. 1998.
[7] H. Lu and P. V. Kumar, "Rate-diversity tradeoff of space-time codes with fixed alphbat and optimal construction for PSK modulation," *IEEE Trans. Inf. Theory*, vol. 49, no. 10, pp. 2747–2751, Oct. 2003.
[8] Z. Liu, Y. Xin, and G. B. Giannakis, "Linear constellation precoding for OFDM with maximum multipath diversity and coding gains," *IEEE Trans. Commun.*, vol. 51, no. 3, pp. 416–427, Mar. 2003.
[9] Z. Wang and G. B. Giannakis, "Linearly precoded or coded OFDM against wireless channel fades?," in *Proc. 3rd IEEE Workshop on Signal Processing Advances in Wireless Communications*, Taoyuan, Taiwan, ROC, Mar. 2001, pp. 267–270.
[10] ——, "Complex-field coding for OFDM over fading wireless channels," *IEEE Trans. Inf. Theory*, vol. 49, no. 3, pp. 707–720, Mar. 2003.
[11] Z. Wang, X. Ma, and G. B. Giannakis, "OFDM or single-carrier block transmissions?," *IEEE Trans. Commun.*, vol. 52, no. 3, pp. 380–394, Mar. 2004.

On the Universality of Burnashev's Error Exponent

Aslan Tchamkerten and İ. Emre Telatar, *Member, IEEE*

***Abstract*—We consider communication over a time-invariant discrete memoryless channel (DMC) with noiseless and instantaneous feedback. We assume that the transmitter and the receiver are not aware of the underlying channel, however, they know that it belongs to some specific family of DMCs. Recent results show that for certain families (e.g., binary-symmetric channels and Z channels) there exist 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.**

***Index Terms*—Burnashev's error exponent, discrete memoryless channels (DMCs), feedback, composite hypothesis testing, two-message communication, unknown channel, zero-rate communication.**

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

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

where[1]

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

is the Kullback–Liebler distance 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. We will refer to $E_B(R,Q)$ 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 feedback 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 to this question is provided in [5] 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 result holds for the class of Z channels with crossover probability $\varepsilon\in[0,L]$ and with $L\in[0,1)$. In [5] it is shown that, given any $\gamma\in[0,1)$, there exist 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 correspondence, we consider the possibility of extending the results in [5] to arbitrary families 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 [5] cannot be extended 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 MAIN RESULT

Throughout this correspondence, we shall be concerned with feedback communication and assume that there are two possible messages to be conveyed, either message A or message N.

Assume that the channel is a DMC Q, revealed to both the transmitter and the receiver, and with finite input and output alphabet $\mathcal{X}$ and $\mathcal{Y}$. In the presence of perfect feedback, the encoder is aware of what has been previously received by the decoder. This allows to have a variable time delivery per message and also allows the encoder to adapt the codewords on the run, based on the available feedback information. Hence, the following definition of coding scheme for feedback communication is natural.[2]

Manuscript received September 14, 2004; revised January 26, 2005. 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 5005-67322.

The authors are with the Laboratoire de Théorie de l'Information (LTHI), ISC-I&C, Ecole Polytechnique Fédérale de Lausanne, CH-1015 Lausanne, Switzerland (e-mail: aslan.tchamkerten@epfl.ch; emre.telatar@epfl.ch).

Communicated by K. Kobayashi, Associate Editor for Shannon Theory.
Digital Object Identifier 10.1109/TIT.2005.850229

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

[2] Definition 1 is standard (see, e.g., [3]).

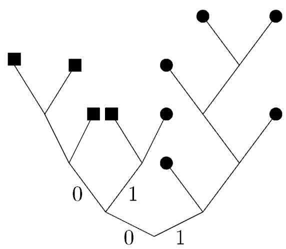

Fig. 1. Given a coding scheme (Ξ, Γ, T) for a binary-output channel, the set of all received sequences for which a decision is made is represented by the leaves of a complete binary tree. Message A is declared at the square leaves whereas message N is declared at the round leaves. The decoder climbs the tree by moving left or right depending whether it receives a zero or a one, until it reaches a leaf and makes a decision accordingly.

Definition 1 (Two-Message Coding Scheme): A codebook (or encoder) is a sequence of functions

$$\Xi \triangleq \{X_n : \{A, N\} \times \mathcal{Y}^{n-1} \longrightarrow \mathcal{X}\}_{n \geq 1} \tag{2}$$

where $X_n(m, y^{n-1})$ represents the symbol sent at time n, given that the message to be conveyed is m, and that the received symbols up to time $n-1$ are $y^{n-1} \triangleq y_1, y_2, \ldots, y_{n-1}$.

A decoder consists of a sequence of functions

$$\Gamma \triangleq \{\gamma_n : \mathcal{Y}^n \longrightarrow \{A, N\}\}_{n \geq 1} \tag{3}$$

and a stopping time (decision time) T with respect to the received symbols $Y_1, Y_2, \ldots$[3] The decoder declares the message $\gamma_T(y^T)$. A two-message coding scheme is a tuple (Ξ, Γ, T).

Consider a two-message coding scheme (Ξ, Γ, T) used over the channel Q. Given the decoder (Γ, T), 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.[4] The set of leaves is divided into two sets that correspond to declaring either message A or message N (see Fig. 1 for an example). 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.

From a probabilistic point of view, the decision time T determines the probability space of the output sequences, or, equivalently, the set of leaves. On this probability space, each sequence of encoding functions $\{X_n(m, \cdot)\}_{n \geq 1}$, $m \in \{A, N\}$, 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 any channel Q and two-message coding scheme (Ξ, Γ, T), there is a natural probability space with two probability measures P_A and P_N that correspond to the sending of message A or N. It will be important in the sequel to have this perspective in mind, namely, to consider the messages as inducers of probabilities on the probability space defined by the decision time. In the sequel, we shall often be concerned with the relative entropy between P_A and P_N that we denote by $D(P_A \,\|\, P_N)$. This quantity is naturally defined on the probability space set by the decision time.

Given a coding scheme (Ξ, Γ, T), the probability of declaring message N while A is sent is denoted $P_A(N)$. In other words, $P_A(N)$ denotes the probability under P_A of the set of all leaves of the decision tree for which message N is declared. Similarly, let $P_N(A)$ be the probability of declaring message A while N is sent. With these conventions, the average error probability $\mathbb{P}(\mathcal{E})$ is given by $(1/2)(P_A(N) + P_N(A))$ and the average decoding time $\mathbb{E}T$ is given by $(1/2)(\mathbb{E}_A T + \mathbb{E}_N T)$, where the subscripts indicate to which message the expectations refer to.

Given a DMC Q and a sequence of two-message coding schemes $\omega = \{(\Xi_i, \Gamma_i, T_i)\}_{i \geq 1}$, let $P_A^i(N)$ and $P_N^i(A)$ denote the error probabilities with respect to (Ξ_i, Γ_i, T_i) and Q.

Definition 2 (Error Exponents): Given a DMC Q let $\omega = \{(\Xi_i, \Gamma_i, T_i)\}_{i \geq 1}$ be a sequence of two-message coding schemes such that $P_A^i(N) \to 0$ and $P_N^i(A) \to 0$ as $i \to \infty$.[5] The error exponents with respect to messages A and N are defined as

$$E_A(\omega, Q) \triangleq \liminf_{i \to \infty} -\frac{1}{\mathbb{E}_A T_i} \ln P_A^i(N) \tag{4}$$

and

$$E_N(\omega, Q) \triangleq \liminf_{i \to \infty} -\frac{1}{\mathbb{E}_N T_i} \ln P_N^i(A) \tag{5}$$

and the average error exponent is defined as

$$E(\omega, Q) \triangleq \liminf_{i \to \infty} -\frac{1}{\mathbb{E} T_i} \ln \mathbb{P}(\mathcal{E}_i) \tag{6}$$

where $\mathbb{P}(\mathcal{E}_i)$ and $\mathbb{E}T_i$ denote the average error probability and the average decoding time with respect to (Ξ_i, Γ_i, T_i) and Q.

We now give a precise formulation of our problem. Given a family of DMCs $\mathcal{Q}$, which elements have the same input and output alphabets $\mathcal{X}$ and $\mathcal{Y}$, does a sequence of two-message coding schemes ω exist such that

$$E(\omega, Q) = E_B(0, Q)$$

for all $Q \in \mathcal{Q}$? The main result of this correspondence is a sufficient condition on a pair of channels Q_1 and Q_2 under which the answer is negative. First, let us define[6]

$$K(Q_1, Q_2) \triangleq \max_{(x, x') \in \mathcal{X} \times \mathcal{X}} [D(Q_1(\cdot \,|\, x) \,\|\, Q_1(\cdot \,|\, x')) + D(Q_1(\cdot \,|\, x) \,\|\, Q_2(\cdot \,|\, x'))]. \tag{7}$$

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

$$K(Q_i, Q_j) < 2 \max_{(x, x') \in \mathcal{X} \times \mathcal{X}} D(Q_i(\cdot \,|\, x) \,\|\, Q_i(\cdot \,|\, x')). \tag{8}$$

[3]An integer-valued random variable T is called a stopping time with respect to a sequence of random variables $Y_1, Y_2, \ldots$ if, conditioned on $Y_1, \ldots, Y_n$, the event $\{T = n\}$ is independent of $Y_{n+1}, Y_{n+2}, \ldots$ for all $n \geq 1$, and $\mathbb{P}(T < \infty) = 1$. Throughout this correspondance we will assume without loss of generality that the decision time T is not randomized. Indeed, we could assume that, at each time n, the decision of the decoder whether to stop depends not only on the received symbols y^n, but also on the outcome of an auxiliary random experiment. This random experiment can be specified, in complete generality, by a binary random variable $R(y^n)$. If at time n the decoder has received y^n, the decoder makes a decision if $R(y^n) = 0$ and transmission continues if $R(y^n) = 1$. By means of feedback the encoder is informed about the outcome of the random experiment. Now suppose that, before any message is transmitted, the decoder collects a family of samples $\{r(y^n)\}$, for all y^n and $n \geq 1$ (actually it only needs to collect sample $r(y^n)$ only if $r(y^{n-1}) = 1$). It then sends these samples to the encoder through the feedback link. Then transmission starts and the decoder decides whether to stop on the basis of the family of samples $\{r(y^n)\}$, i.e., in a nonrandomized way.

[4]A tree is said to be a complete $|\mathcal{Y}|$-ary tree if any vertex is either a leaf or has $|\mathcal{Y}|$ immediate descendents.

[5]Clearly, such a sequence exists if the channel capacity $C(Q)$ is strictly positive.

[6]Notice that $E_B(0, Q_i) \leq K(Q_i, Q_j)$ for $i, j \in \{1, 2\}$.

For any sequence of two-message coding schemes ω, either $E(\omega, Q_1) < E_B(0, Q_1)$, or $E(\omega, Q_2) < E_B(0, Q_2)$, or both.

Since the zero-rate error exponent is upper-bounded by the error exponent for a fixed number of messages, whenever Q_1 and Q_2 satisfy the hypothesis of the theorem, no zero-rate coding scheme achieves an error exponent equal to $E_B(0, Q_1)$ on Q_1 and an error exponent equal to $E_B(0, Q_2)$ on Q_2. Stated otherwise, if Q_1 and Q_2 satisfy the hypothesis of the theorem, then no zero-rate coding scheme achieves on both channels the maximum error exponent that could be obtained if the channels were revealed to both the transmitter and the receiver. A simple example of channels Q_1 and Q_2 that satisfy the assumptions of the theorem is given by $Q_1 =$ BSC (ε) and $Q_2 =$ BSC $(1-\varepsilon)$ where $0 < \varepsilon < 1/2$. In this case, we have

$$\begin{aligned} K(Q_1, Q_2) &= \max_{(x,x') \in \{0,1\} \times \{0,1\}} D(Q_1(\cdot \,|\, x) \,\|\, Q_1(\cdot \,|\, x'))) \\ &= \max_{(x,x') \in \{0,1\} \times \{0,1\}} D(Q_2(\cdot \,|\, x \,\|\, Q_2(\cdot \,|\, x')) \\ &= K(Q_2, Q_1) \end{aligned} \tag{9}$$

and (8) holds.[7]

III. Two-Message Coding for Two Channels

In this section, we will prove the theorem.

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,[8] the quantities $P_1(B)$ and $P_2(B^c)$ cannot both be rendered arbitrary small by an appropriate choice of $B \in \mathcal{F}$.[9] More precisely, from the data processing inequality for divergence,[10] we have the following lower bound on $P_1(B)$:

$$P_1(B) \geq \exp\left[\frac{-D(P_2 \,\|\, P_1) - H(P_2(B))}{1 - P_2(B^c)}\right] \tag{12}$$

where $H(\alpha) \triangleq -\alpha \ln \alpha - (1-\alpha)\ln(1-\alpha)$. In the sequel, we shall use (12) in order to derive bounds on the maximum error exponents that can simultaneously be achieved over two channels.

[7]Note that there are pairs of BSCs, with crossover probabilities ε_1 and ε_2, such that $\varepsilon_1 + \varepsilon_2 \neq 1$ and that also satisfy (8), e.g., $\varepsilon_1 = 0.1$ and $\varepsilon_2 = 0.77$.

[8]P_1 and P_2 are said to be singular if there exists some $B \in \mathcal{F}$ such that $P_1(B) = 0$ and $P_2(B) = 1$.

[9]B^c denotes the complementary set of B in Ω.

[10]Let $(\Omega, \mathcal{F})$ be a probability space, let P_1 and P_2 be two probability measures on $(\Omega, \mathcal{F})$, and let $B \in \mathcal{F}$. From the data processing inequality for divergence [3, p. 55], we have

$$D(P_2 \,\|\, P_1) \geq D(P_2(B) \,\|\, P_1(B)) \tag{10}$$

where

$$D(P_2(B) \,\|\, P_1(B)) \triangleq P_2(B) \ln \frac{P_2(B)}{P_1(B)} + (1 - P_2(B)) \ln \frac{(1-P_2(B))}{(1-P_1(B))}. \tag{11}$$

Expanding (10) we deduce that

$$P_1(B) \geq \exp\left[\frac{-D(P_2 \,\|\, P_1) - H(P_2(B))}{P_2(B)}\right]$$

where $H(P_2(B)) \triangleq -P_2(B) \ln P_2(B) - (1-P_2(B)) \ln(1-P_2(B))$.

Suppose we use some coding scheme (Ξ, Γ, T) on a known channel Q. Letting B be the set of leaves for which message A is declared, respectively, the set of leaves for which message N is declared, from (12) we obtain

$$\begin{aligned} P_N(A) &\geq \exp\left[\frac{-D(P_A \,\|\, P_N) - H(P_A(A))}{1 - P_A(N)}\right] \\ P_A(N) &\geq \exp\left[\frac{-D(P_N \,\|\, P_A) - H(P_N(N))}{1 - P_N(A)}\right]. \end{aligned} \tag{13}$$

Note that since one is usually interested in the case where $P_N(A)$ and $P_A(N)$ are small, the terms on the right-hand side of the two inequalities in (13) are essentially $\exp[-D(P_A \,\|\, P_N)]$ and $\exp[-D(P_N \,\|\, P_A)]$.

Assume now that the transmitter and the receiver still want to communicate using (Ξ, Γ, T), but that 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 on the output sequences when message $m \in \{A, N\}$ is being sent through channel $Q_i, i \in \{1,2\}$. We now have four distributions defined on the probability space set by the decision time T, namely, $P_{m,i}$ with $m \in \{A, N\}$ and $i \in \{1,2\}$. There are also four error probabilities $P_{A,1}(N)$, $P_{A,2}(N)$, $P_{N,1}(A)$, and $P_{N,2}(A)$. Using (12) with $B = N$, and $(P_1, P_2) = (P_{A,1}, P_{N,1}), (P_{A,1}, P_{N,2}), \ldots$ we get the following inequalities:

$$P_{A,1}(N) \geq \exp\left[\frac{-D(P_{N,1} \,\|\, P_{A,1}) - H(P_{N,1}(N))}{1 - P_{N,1}(A)}\right] \tag{14}$$

$$P_{A,1}(N) \geq \exp\left[\frac{-D(P_{N,2} \,\|\, P_{A,1}) - H(P_{N,2}(N))}{1 - P_{N,2}(A)}\right] \tag{15}$$

$$P_{A,2}(N) \geq \exp\left[\frac{-D(P_{N,1} \,\|\, P_{A,2}) - H(P_{N,1}(N))}{1 - P_{N,1}(A)}\right] \tag{16}$$

$$P_{A,2}(N) \geq \exp\left[\frac{-D(P_{N,2} \,\|\, P_{A,2}) - H(P_{N,2}(N))}{1 - P_{N,2}(A)}\right]. \tag{17}$$

In a similar fashion one also obtains

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

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

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

$$P_{N,2}(A) \geq \exp\left[\frac{-D(P_{A,2} \,\|\, P_{N,2}) - H(P_{A,2}(A))}{1 - P_{A,2}(N)}\right]. \tag{21}$$

These equations can be interpreted in terms of the error probabilities of a hypothesis test that distinguishes the two composite hypothesis "message A" $= \{P_{A,1}, P_{A,2}\}$ and "message N" $= \{P_{N,1}, P_{N,2}\}$. The following proposition will be the key ingredient in the proof of the theorem.

Proposition: Let Q_1 and Q_2 be two DMCs on $\mathcal{X} \times \mathcal{Y}$. For any coding scheme (Ξ, Γ, T)

$$D(P_{N,1} \,\|\, P_{A,1}) + D(P_{N,1} \,\|\, P_{A,2}) \leq K(Q_1, Q_2)\mathbb{E}_{N,1}T \tag{22}$$

$$D(P_{N,2} \,\|\, P_{A,2}) + D(P_{N,2} \,\|\, P_{A,1}) \leq K(Q_2, Q_1)\mathbb{E}_{N,2}T \tag{23}$$

where $K(Q_i, Q_j)$ is defined in (7). If $K(Q_i, Q_j) = 0$ and $\mathbb{E}_{N,i}T = \infty$ we set $K(Q_i, Q_j)\mathbb{E}_{N,i}T = 0$.

Proof of the Proposition: We only prove inequality (22). Inequality (23) can then be easily derived from (22) by exchanging the roles of Q_1 and Q_2.

We have the following cases:

a. $K(Q_1, Q_2) = \infty$,
b. $0 < K(Q_1, Q_2) < \infty$ and $\mathbb{E}_{N,1}T = \infty$,
c. $K(Q_1, Q_2) < \infty$ and $\mathbb{E}_{N,1}T < \infty$,
d. $K(Q_1, Q_2) = 0$ and $\mathbb{E}_{N,1}T = \infty$.

In the cases a and b the inequality (22) trivially holds.

Case c: Let us define, for all $n \geq 1$, the random variables

$$Z_n \triangleq \ln \frac{P_{N,1}(Y^n)}{\sqrt{P_{A,1}(Y^n)P_{A,2}(Y^n)}}$$

and

$$S_n \triangleq Z_n - \frac{n}{2}K(Q_1, Q_2). \tag{24}$$

We first prove that the sequence $\{S_n\}_{n\geq 1}$ forms a supermartingale with respect to the output symbols $Y_1, Y_2, \ldots$ when this sequence is generated according to $P_{N,1}$. By applying the Stopping Theorem for Supermartingales (see, e.g., [4, Theorem 6.4.1]) we will obtain

$$0 \geq \mathbb{E}_{N,1} S_T \tag{25}$$

which is equivalent to the desired result

$$D(P_{N,1} \,\|\, P_{A,1}) + D(P_{N,1} \,\|\, P_{A,2}) \leq K(Q_1, Q_2)\mathbb{E}_{N,1}T. \tag{26}$$

Since $K(Q_1, Q_2) < \infty$, we have for all $y \in \mathcal{Y}$ and $x, x' \in \mathcal{X}$ the following implications:

$$Q_1(y \,|\, x') > 0 \Leftrightarrow Q_1(y \,|\, x) > 0 \Rightarrow Q_2(y|x') > 0 \tag{27}$$

implying that

$$\mathbb{E}_{N,1}|S_n| < \infty, \qquad \text{for all} \quad n \geq 1. \tag{28}$$

To show that $\{S_n\}_{n\geq 1}$ is a supermartingale, first one can easily check that

$$\begin{aligned}\mathbb{E}(Z_{n+1}|y^n, Q_1, N) &\\ = z_n + \mathbb{E}\Bigg(\ln \frac{\mathbb{P}(Y_{n+1} \,|\, y^n, Q_1, N)}{\sqrt{\mathbb{P}(Y_{n+1} \,|\, y^n, Q_1, A)\mathbb{P}(Y_{n+1} \,|\, y^n, Q_2, A)}} &\\ \Bigg| \, y^n, Q_1, N\Bigg)\end{aligned} \tag{29}$$

where the conditioning is made on the received sequence y^n, the underlying channel Q_1, and the sent message N. Denote by β_j^m the probability that $X_{n+1} = j$ given y^n and that the sent message is m.[11] It follows that

$$\mathbb{P}(Y_{n+1} = k \,|\, y^n, Q_i, m) = \sum_j Q_i(k \,|\, j)\beta_j^m \tag{30}$$

and hence we have the following identities:

$$\begin{aligned}&\mathbb{E}\left(\ln \frac{\mathbb{P}(Y_{n+1}|y^n, Q_1, N)}{\sqrt{\mathbb{P}(Y_{n+1} \,|\, y^n, Q_1, A)\mathbb{P}(Y_{n+1} \,|\, y^n, Q_2, A)}} \,\Bigg|\, y^n, Q_1, N\right)\\ &\quad = \frac{1}{2}\sum_k \left(\sum_j Q_1(k \,|\, j)\beta_j^N\right) \ln \frac{\sum_j Q_1(k|j)\beta_j^N}{\sum_j Q_1(k|j)\beta_j^A}\\ &\quad + \frac{1}{2}\sum_k \left(\sum_j Q_1(k \,|\, j)\beta_j^N\right) \ln \frac{\sum_j Q_1(k|j)\beta_j^N}{\sum_j Q_2(k|j)\beta_j^A}\\ &\quad = \frac{D(P_1(\beta^N) \,\|\, P_1(\beta^A)) + D(P_1(\beta^N) \,\|\, P_2(\beta^A))}{2}\end{aligned} \tag{31}$$

with $\beta^m \triangleq (\beta_1^m, \beta_2^m, \ldots, \beta_{|\mathcal{X}|}^m)$, and where $P_i(\beta^m)$ denotes the distribution $\sum_j Q_i(\cdot|j)\beta_j^m$. Now since $P_i(\beta^m)$ is linear in β^m, by the convexity of the Kullback–Liebler distance in both of its arguments (see, e.g., [2, Theorem 2.7.2]) the function

$$(\beta^A, \beta^N) \mapsto D(P_1(\beta^N) \,\|\, P_1(\beta^A)) + D(P_1(\beta^N) \,\|\, P_2(\beta^A)) \tag{32}$$

is convex and its maximum occurs at some (β^A, β^N) where β^A and β^N have all but one coordinate equal to zero. Therefore we have

$$\begin{aligned}&\max_{\beta^A, \beta^N}[D(P_1(\beta^N) \,\|\, P_1(\beta^A)) + D(P_1(\beta^N) \,\|\, P_2(\beta^A))]\\ &\quad = \max_{(x,x')\in\mathcal{X}\times\mathcal{X}}[D(Q_1(\cdot \,|\, x) \,\|\, Q_1(\cdot \,|\, x'))\\ &\qquad + D(Q_1(\cdot \,|\, x) \,\|\, Q_2(\cdot \,|\, x'))]\\ &\quad = K(Q_1, Q_2).\end{aligned} \tag{33}$$

From (29), (31), and (33) we deduce that

$$\mathbb{E}_{N,1} S_1 \leq 0 \tag{34}$$

and that, for all $n \geq 1$ and $y^n \in \mathcal{Y}^n$

$$\mathbb{E}\left(S_{n+1} \Big| y^n, Q_1, N\right) \leq s_n. \tag{35}$$

From (28) and (35), the sequence $\{S_n\}_{n\geq 1}$ forms a supermartingale with respect to $Y_1, Y_2, \ldots$ when this sequence is generated according to $P_{N,1}$.

We now check that the Stopping Theorem for Supermartingales can be applied, i.e., we verify that for all $n \geq 1$

$$\mathbb{E}\left(|S_{n+1} - S_n| \Big| S^n, Q_1, N\right) < M$$

for some constant $M < \infty$. If we consider the conditioning on y^n instead of s^n, from (29) we have

$$\begin{aligned}&\mathbb{E}\left(|S_{n+1} - S_n| \Big| y^n, Q_1, N\right) \leq \frac{K(Q_1, Q_2)}{2}\\ &\quad + \mathbb{E}\left(\left|\ln \frac{\mathbb{P}(Y_{n+1}|y^n, Q_1, N)}{\sqrt{\mathbb{P}(Y_{n+1}|y^n, Q_1, A)\mathbb{P}(Y_{n+1}|y^n, Q_2, A)}}\right| \,\Bigg|\, y^n, Q_1, N\right).\end{aligned} \tag{36}$$

From (27) we deduce that the expectation on the right-hand side of (36) can be upper-bounded by some finite constant for all $n \geq 1$. Hence, there exists some $M < \infty$ such that

$$\mathbb{E}\left(|S_{n+1} - S_n| \Big| S^n, Q_1, N\right) < M \tag{37}$$

for all $n \geq 1$. Since by assumption $\mathbb{E}_{N,1}T < \infty$, the Stopping Theorem for Supermartingales yields

$$0 \geq \mathbb{E}_{N,1}S_1 \geq \mathbb{E}_{N,1}S_T. \tag{38}$$

Case d: If $K(Q_i, Q_j) = 0$ the channels Q_i and Q_j are the same and with zero capacity. In particular, $Z_n = 0$ for all $n \geq 1$ and hence $D(P_{N,1} \,\|\, P_{A,1}) + D(P_{N,1} \,\|\, P_{A,2}) = 0$. □

Proof of the Theorem: The main idea that underlies the proof is the following. Informally, from the proposition we will first deduce an upper bound on the sum of the error exponents that can be obtained by any sequence of two-message coding schemes ω on two channels Q_1 and Q_2. Under the assumption (8), this upper bound is smaller than $E_B(0, Q_1) + E_B(0, Q_2)$, which yields the desired result.

Pick a coding scheme (Ξ, Γ, T). From the proposition we have

$$D(P_{N,1} \,\|\, P_{A,1}) + D(P_{N,1} \,\|\, P_{A,2}) \leq K(Q_1, Q_2)\mathbb{E}_{N,1}T \tag{39}$$

[11]Note that β_j^m is not a function of the channel, it depends only on the coding scheme. In particular, as shall be clear from the proof, the proposition remains valid if one considers coding schemes with randomized encoding procedures, which is captured by vectors $\beta^m \triangleq (\beta_1^m, \beta_2^m, \ldots, \beta_{|\mathcal{X}|}^m)$ having at least two nonzero components.

and

$$D(P_{N,2} \,\|\, P_{A,2}) + D(P_{N,2} \,\|\, P_{A,1}) \le K(Q_2, Q_1)\mathbb{E}_{N,2}T. \quad (40)$$

By exchanging the roles of A and N we also obtain

$$D(P_{A,1} \,\|\, P_{N,1}) + D(P_{A,1} \,\|\, P_{N,2}) \le K(Q_1, Q_2)\mathbb{E}_{A,1}T \quad (41)$$

and

$$D(P_{A,2} \,\|\, P_{N,2}) + D(P_{A,2} \,\|\, P_{N,1}) \le K(Q_2, Q_1)\mathbb{E}_{A,2}T. \quad (42)$$

From (39)–(42) we get

$$\begin{aligned} &D(P_{N,1} \,\|\, P_{A,1}) + D(P_{N,2} \,\|\, P_{A,1}) \\ &\quad + D(P_{A,1} \,\|\, P_{N,1}) + D(P_{A,2} \,\|\, P_{N,1}) \\ &\quad + D(P_{N,1} \,\|\, P_{A,2}) + D(P_{N,2} \,\|\, P_{A,2}) \\ &\quad + D(P_{A,1} \,\|\, P_{N,2}) + D(P_{A,2} \,\|\, P_{N,2}) \\ &\qquad \le 2K(Q_1, Q_2)\mathbb{E}_1 T + 2K(Q_2, Q_1)\mathbb{E}_2 T \end{aligned} \quad (43)$$

where $\mathbb{E}_i T$ denotes the average decoding time when channel Q_i is used, i.e., $\mathbb{E}_i T \triangleq (1/2)(\mathbb{E}_{A,i}T + \mathbb{E}_{N,i}T)$. From (43), we infer that at least one of the two following inequalities holds:

$$\begin{aligned} &\min\Big\{ D(P_{N,1} \,\|\, P_{A,1}) + D(P_{N,2} \,\|\, P_{A,1}), \\ &\qquad D(P_{A,1} \,\|\, P_{N,1}) + D(P_{A,2} \,\|\, P_{N,1}) \Big\} \\ &\qquad \le K(Q_1, Q_2)\mathbb{E}_1 T \end{aligned} \quad (44)$$

$$\begin{aligned} &\min\Big\{ D(P_{N,1} \,\|\, P_{A,2}) + D(P_{N,2} \,\|\, P_{A,2}), \\ &\qquad D(P_{A,1} \,\|\, P_{N,2}) + D(P_{A,2} \,\|\, P_{N,2}) \Big\} \\ &\qquad \le K(Q_2, Q_1)\mathbb{E}_2 T. \end{aligned} \quad (45)$$

Suppose now there exists a sequence of two-message coding schemes $\omega = \{(\Xi_i, \Gamma_i, T_i)\}_{i \ge 1}$ such that the error probabilities $P^i_{A,1}(N)$, $P^i_{A,2}(N)$, $P^i_{N,1}(A)$, and $P^i_{N,1}(A)$ vanish as $i \to \infty$. It follows that at least one of the two following inequalities holds for infinitely many i:

$$\begin{aligned} &\min\left\{ D\left(P^i_{N,1} \middle\| P^i_{A,1}\right) + D\left(P^i_{N,2} \middle\| P^i_{A,1}\right), \right. \\ &\qquad \left. D\left(P^i_{A,1} \middle\| P^i_{N,1}\right) + D\left(P^i_{A,2} \middle\| P^i_{N,1}\right) \right\} \\ &\qquad \le K(Q_1, Q_2)\mathbb{E}^i_1 T \end{aligned} \quad (46)$$

$$\begin{aligned} &\min\left\{ D\left(P^i_{N,1} \middle\| P^i_{A,2}\right) + D\left(P^i_{N,2} \middle\| P^i_{A,2}\right), \right. \\ &\qquad \left. D\left(P^i_{A,1} \middle\| P^i_{N,2}\right) + D\left(P^i_{A,2} \middle\| P^i_{N,2}\right) \right\} \\ &\qquad \le K(Q_2, Q_1)\mathbb{E}^i_2 T. \end{aligned} \quad (47)$$

Suppose that (46) holds for infinitely many i. Since by assumption

$$K(Q_1, Q_2) < 2 \max_{x,x'} D(Q_1(\cdot \,|\, x) \,\|\, Q_1(\cdot \,|\, x')), \quad (48)$$

from the inequalities (14)–(15) and (18)–(19) we deduce that at least one of the following two inequalities holds:

$$E_A(\omega, Q_1) < \max_{x,x'} D(Q_1(\cdot \,|\, x) \,\|\, Q_1(\cdot \,|\, x')) \quad (49)$$

$$E_N(\omega, Q_1) < \max_{x,x'} D(Q_1(\cdot \,|\, x) \,\|\, Q_1(\cdot \,|\, x')) \quad (50)$$

and, therefore,

$$E(\omega, Q_1) < \max_{x,x'} D(Q_1(\cdot \,|\, x) \,\|\, Q_1(\cdot \,|\, x')). \quad (51)$$

Similarly, if (47) holds for infinitely many i

$$E(\omega, Q_2) < \max_{x,x'} D(Q_2(\cdot \,|\, x) \,\|\, Q_2(\cdot \,|\, x')). \quad (52)$$

Hence, whenever Q_1 and Q_2 satisfy the hypothesis of the theorem, for any sequence of coding schemes ω, either $E(\omega, Q_1) < E_B(0, Q_1)$ or $E(\omega, Q_2) < E_B(0, Q_2)$. □

IV. Conclusion

Given a family of DMCs $\mathcal{Q}$, in general, no zero-rate coding scheme achieves the maximum error exponent universally over $\mathcal{Q}$. Hence, the property of the families of BSCs and Z channels that was shown in [5] does not hold 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 have been obtained if the channel were revealed to the transmitter and the receiver [1]. If we look at the problem of two-message coding over two channels from a hypothesis testing perspective, as already mentioned previously, the goal of the decoder is to discriminate between two composite hypothesis "$\text{message } A$" $= \{P_{A,1}, P_{A,2}\}$ and "$\text{message } N$"$=$ $\{P_{N,1}, P_{N,2}\}$. The encoder, with the help of feedback has a certain control on the output of the channel so that it may help the decoder to better distinguish between the two hypothesis. From our result, we may conclude in a certain sense that, in spite of the help provided by feedback, it alone may not be enough if the underlying channel is unknown to both the transmitter and the receiver.

Acknowledgment

The authors wish to thank I. Csiszár and C.-E. Pfister for stimulating discussions, and M. V. Burnashev, J. L. Massey, and B. Rimoldi for valuable comments on the manuscript.

References

[1] M. V. Burnashev, "Data transmission over a discrete channel with feedback: Random transmission time," *Probl. Inf. Transm.*, vol. 12, no. 4, pp. 250–265, 1976.

[2] T. Cover and J. A. Thomas, *Elements of Information Theory*. New York: Wiley, 1991.

[3] I. Csiszár and J. Körner, *Information Theory*. Budapest, Hungary: Akademiai Kiado, 1986.

[4] S. Ross, *Stochastic Processes*. New York: Wiley, 1996.

[5] A. Tchamkerten and İ. E. Telatar, "Variable length coding over an unknown channel," *IEEE Trans. Inf. Theory*, to be published.