On the Use of Training Sequences for Channel Estimation

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*— Suppose $\mathcal{Q}$ is a family of discrete memoryless channels. An unknown member of $\mathcal{Q}$ will be available with perfect (causal) feedback for communication. A recent result [9] shows the existence, for certain families of channels (e.g. Binary Symmetric Channels and Z channels), of coding schemes that achieve Burnashev's exponent universally over these families. In other words, in certain cases, there is no loss in the error exponent by ignoring the channel: transmitter and receiver can design optimal blind coding schemes that perform as well as the best feedback coding schemes tuned for the channel under use. Here we study the situation where communication is carried by first testing the channel by means of a training sequence, then coding the information according to the channel estimate. We provide an upper bound on the maximum achievable error exponent of any such scheme. If we consider Binary Symmetric Channels and Z channels this bound is much lower than Burnashev's exponent. This suggests that in terms of error exponent, a good universal feedback scheme entangles channel estimation with information delivery, rather than separating them.**

I. Introduction

When considering information transmission over a channel that is partially known to either the transmitter or the receiver or both, it is common to employ a *training sequence*. This sequence is sent prior to the data to be conveyed and its purpose is to help the decoder (for channels without feedback) or both the encoder and the decoder (for channels with feedback) to adjust its/their parameters for the upcoming communication. For example, in slow fading channels without feedback, a training sequence can be sent at the beginning of each coherence interval, so that the receiver can estimate the channel characteristics, and then communicate with these parameters (see, e.g., [1], [5], [6], [10]).

Here we study feedback communication over a time invariant discrete memoryless channel (DMC) with perfect feedback, i.e, noiseless and instantaneous (causal). We assume that the transmitter and the receiver are not aware of the transition probability matrix Q of the channel, however, both know that Q belongs to some subset $\mathcal{Q}$ of DMCs.

In principle, the sending of a training sequence need not affect the rates achievable by the communication system: the test sequence length can be made negligible compared to the length of the subsequent information sequence. However, and this is the main concern of this paper, the separation of the channel estimation from the information coding may result in itself in a penalty in terms of error exponent.

At this point we would like to mention the work of Feder and Lapidoth [4] in which universal decoders for families of channels without feedback are considered. It is shown that there exists universal decoders that are optimal in the sense that they perform (asymptotically) as well as the Maximum Likelihood decoder tuned for the channel over which transmission is carried out. In particular it is shown that the combination of a training sequence and a Maximum Likelihood decoder designed for the estimated channel is not optimal. The results presented in this paper, while concerning feedback channels, have the same flavor.

A journal version of the material presented in this proceeding has been submitted 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. .

II. Preliminaries

In this section we remind the definitions of coding schemes, rate, error probability, and error exponent for a DMC with perfect feedback, then we state a Theorem due to Burnashev.

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

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

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 $y^{n-1} = y_1, y_2, \ldots, y_{n-1}$. A codeword for message m is the sequence of functions $\{X_n(m, \;\cdot\;)\}_{n\geq 1}$. A decoder $(\Psi^M, T(M))$ consists of a sequence of functions

$$\Psi^M = \{\psi_n^M : \mathcal{Y}^n \longrightarrow \mathcal{M}\}_{n\geq 1} \,, \tag{2}$$

and a stopping time $T(M)$, relative to the received symbols $Y_1, Y_2, \ldots$,[1] that represents the decision time. The decoded message is given by $\psi^M_{T(M)}(y^{T(M)})$. A coding scheme is a tuple $\mathcal{S}^M = (\mathcal{C}^M, \Psi^M, T(M))$.

[1] 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 $n \geq 1$.

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[2]

$$R(\mathcal{S}^M, Q) \triangleq \frac{\ln M}{\mathbb{E}T(M)} \text{ nats per symbol }, \quad (3)$$

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

$$\mathbb{E}T(M) \triangleq \frac{1}{M} \sum_{m \in \mathcal{M}} \mathbb{E}\left(T(M) \mid \text{ message } m \text{ is sent}\right) . \quad (4)$$

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

$$R(\theta, Q) \triangleq \lim_{M \to \infty} R(\mathcal{S}^M, Q) \quad (5)$$

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

$$\mathbb{P}(\mathcal{E}|Q, \mathcal{S}^M) = \frac{1}{M} \sum_{m \in \mathcal{M}} \mathbb{P}\left(\psi^M_{T(M)}(Y^{T(M)}) \neq m \;\middle|\; \text{ message } m \text{ is sent}\right) . \quad (6)$$

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

$$E(\theta, Q) \triangleq \liminf_{M \to \infty} -\frac{1}{\mathbb{E}T(M)} \ln \mathbb{P}(\mathcal{E}|Q, \mathcal{S}^M) . \quad (7)$$

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

Theorem 1 (Burnashev 1976 [2]): 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 $R(\theta, Q) = R$,

$$\limsup_{M \to \infty} -\frac{1}{\mathbb{E}T(M)} \ln \mathbb{P}(\mathcal{E}|Q, \mathcal{S}^M) \leq E_B(R, Q) \quad (8)$$

where

$$E_B(R, Q) \triangleq \left(\max_{(x,x') \in \mathcal{X} \times \mathcal{X}} D(Q(\cdot|x) \| Q(\cdot|x')) \right) \left(1 - \frac{R}{C(Q)}\right) , \quad (9)$$

and 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 between the output distributions induced by the input letters x and x'. Moreover there exists $\theta \in \Theta$ such that $R(\theta, Q) = R$ and $E(\theta, Q) = E_B(R, Q)$.

From now on $E_B(R, Q)$ will be referred as the Burnashev's exponent.

[2] ln denotes the natural logarithm.

III. Statement of Result

In this section we propose a general definition of a training based scheme and provide an upper bound on the error exponent of any such coding schemes. Then we draw a few comparisons between training based schemes and universal schemes studied in [9] that do not separate channel estimation and information delivery. Finally we give a sketch of the prove of our result.

Let $\mathcal{Q}$ be a family of DMCs. We suppose that communication is carried over some element $Q \in \mathcal{Q}$ that is revealed neither to the transmitter nor to the receiver. The communication schemes we shall focus on are referred as "training based schemes" and admit two phases: a first phase of fixed length t, the "training period" (or "test period"), during which the channel parameter is estimated, and a second phase used to carry information. The choice of the encoder/decoder pair used for the second phase is based upon the channel estimate that results from the first phase. Formally we define training based schemes as being coding schemes that satisfy the following two requirements:

I. *Given a set of M messages, a training based scheme $\mathcal{S}^M = (\mathcal{C}^M, \Psi^M, T(M))$ admits a* rate function N_t : $\mathcal{Y}^t \longrightarrow \mathbb{R}_+$ *that associates to each output y^t of the training sequence, the (average) length of the second phase. During the test period, each input symbol is trained a fixed number of times.*

II. *A sequence of training based schemes $\{\mathcal{S}^M = (\mathcal{C}^M, \Psi^M, T(M))\}_{M \geq 1}$ satisfies for some $A < \infty$ and $\gamma \in [0, 1)$, the conditions $T(M) \leq A \ln M$ for all $M \geq 1$ and*

$$\lim_{M \to \infty} \mathbb{P}\left(\frac{\ln M}{T(M)} = \gamma C(Q) \middle| Q \right) = 1$$

for all $Q \in \mathcal{Q}$ with capacity $C(Q)$.

A few comments are in order. Condition I requires to employ for the second phase a coding scheme whose rate depends only upon the output of the test sequence. One cannot use, as a second phase, a coding scheme with a rate that adapts itself according to the channel under use, implicitly estimating the channel (see, e.g., [7], [9]). Also notice that without condition I, it may be possible to first train, then use a variable length coding scheme that simply ignores the result of the testing part while adapting its rate on the run. Hence, condition I implies that, at least from the rate point of view, training based schemes do not estimate the channel during the second phase. Also notice that the requirement I imposes neither a restriction on the channel estimation itself nor on the decision that results from it. Moreover, variable length codes can be used for the second phase provided that, once the training period is over, the average decoding time is set. In particular, the decoding time $T(M)$ equals $t + N_t$ where the average value of N_t depends only on the outcome of the training period.

We impose condition II essentially in order to have some control on the rate, through the "normalized rate" γ, and also to compare training based schemes with universal coding strategies that are proposed in [9] and that have the property that the channel estimation and the coding part are not separated. Finally the restriction that $T(M) \leq A \ln M$ for all $M \geq 1$ may be considered as a mild technical requirement provided that $\inf_{Q\in\mathcal{Q}} C(Q) > 0$.

Our result stands in the following theorem:

Theorem 2: Let $\mathcal{Q}$ be a family of DMCs that have the same input alphabet $\mathcal{X}$ and same output alphabet $\mathcal{Y}$, and let $\theta = \{\mathcal{S}^M\}_{M\geq 1}$ be a sequence of training based schemes for $\mathcal{Q}$ and with parameter $\gamma \in [0,1)$. For any $Q \in \mathcal{Q}$,

$$\limsup_{M\to\infty} -\frac{1}{\mathbb{E}T(M)} \ln \mathbb{P}(\mathcal{E}|Q, \mathcal{S}^M) \leq E_{tbs}(\gamma, Q), \qquad (10)$$

where

$$E_{tbs}(\gamma, Q) \triangleq \min_{V\in A(Q)} \frac{1}{C(V)} \max\left\{\max_{x\in\mathcal{X}} D(V(\cdot|x)||Q(\cdot|x)), E_B(\gamma C(V), Q)\right\} \qquad (11)$$

with $A(Q) \triangleq \{W \in \mathcal{Q} : C(W) \geq C(Q)\}$.

IV. Examples

Given a particular family of channels Theorem 2 gives an upper bound on the error exponent that can be achieved by any training based scheme. We may want to compare this bound with the maximum error exponent that can be universally achieved. Unfortunately few families of channels exists for which we know the maximum error exponent that can be universally achieved. Among them the Binary Symmetric Channels (BSCs) and the Z Channels families [9].

A. Binary Symmetric Channels

Assume that $\mathcal{Q} = \text{BSC}_L$ where BSC_L denotes the set of BSCs with crossover probability $\varepsilon \in [0, L]$ with $L \in [0, 1/2)$.[3] For conciseness, from now on ε denotes both the crossover probability and the BSC with this crossover probability, and $C(\varepsilon)$ its capacity, i.e., $C(\varepsilon) \triangleq \ln 2 + \varepsilon \ln \varepsilon + (1-\varepsilon)\ln(1-\varepsilon)$. It is easy to see that (11) now becomes

$$E_{tbs}(\gamma, \varepsilon) = \min_{\delta\in[0,\varepsilon]} \frac{1}{C(\delta)} \max\{D(\delta||\varepsilon), E_B(\gamma C(\delta), \varepsilon)\} \qquad (12)$$

where $D(\delta\|\varepsilon) \triangleq \delta \ln \frac{\delta}{\varepsilon} + (1-\delta)\frac{1-\delta}{1-\varepsilon}$. Moreover, one can show that the function $E_{tbs}(\gamma, \varepsilon)$ has a slope that vanishes at capacity, more precisely we have for all $\varepsilon \in (0, L]$

$$\lim_{\gamma\uparrow 1} \frac{E_{tbs}(\gamma, \varepsilon)}{1-\gamma} = 0\ . \qquad (13)$$

As one may notice, $E_{tbs}(\gamma, \varepsilon)$ is the same function for any value of $L \in [0, 1/2)$. In figure 1, we plot for two channels ($\varepsilon = 0.1$ and $\varepsilon = 0.4$) the function $R \mapsto E_{tbs}(R/C(\varepsilon), \varepsilon)$ (lower curve) and Burnashev's exponent given by (9) (upper line).

In order to discuss the result we obtain for BSCs, let us first briefly refer to recent results obtained in [9]. Theorem 1 [9] claims that, given any constant $\gamma \in [0,1)$, and the family BSC_L with $L \in [0, 1/2)$, there exists coding schemes that achieve, for every channel $\varepsilon \in \text{BSC}_L$ a rate at least equal to $\gamma C(\varepsilon)$ and a corresponding maximum error exponent, i.e., equal to (9). Suppose now one is interested in having a low error probability instead of a high communication rate. Similarly there exists coding schemes that universally achieve a rate that is guaranteed to be at most γ times the channel capacity and with a corresponding error exponent that is also maximum.

In contrast with these results training based schemes cannot achieve Burnashev's exponent for BSCs. While feedback does not help to increase capacity Burnashev's result tells us that feedback is of particular help at rates close to capacity: a little drop in the rate results in a linear augmentation in the error exponent. Training based schemes fail precisely in having this property: the slope of their error exponent equals zero at capacity. Hence an important feature of feedback is lost and the situation becomes essentially the same as if the channel were revealed to both the transmitter and the receiver and no feedback were available (since the sphere packing bound has a slope equal to zero at capacity).

We may also draw a parallel between feedback communication over a known BSC and an unknown BSC. In the first case, Dobrushin [3] showed that the restriction to fixed length block codes results in an error exponent upper bounded by the sphere packing bound,[4] hence having zero slope at capacity. In the second case, the restriction to training based schemes also results in an error exponent that has zero slope at capacity, even though training based schemes allow variable length codes.

Note however that the comparison between training based schemes and the optimal coding schemes derived in [9] is not completely fair since for training based schemes we require an exact control on the rate through the parameter γ, whereas for the optimal coding schemes γ only yields an upper or a lower bound.

B. Z Channels

Assume that $\mathcal{Q} = \text{Z}_L$ where Z_L denotes the set of Z channels with crossover probability $\varepsilon \in [0, L]$ with $L \in [0, 1)$.[5] Pick a particular channel $Q \in \text{Z}_L$ with nonzero crossover probability. One can find a $\gamma \in [0,1)$ sufficiently close to 1 as well as a channel $W \in \text{Z}_L$ such that $\gamma C(W) > C(Q)$. Therefore we have

$$\begin{aligned} &E_{tbs}(\gamma, Q) \\ &\triangleq \min_{V\in A(Q)} \frac{1}{C(V)} \max\left\{\max_{x\in\mathcal{X}} D(V(\cdot|x)||Q(\cdot|x)), E_B(\gamma C(V), Q)\right\} \\ &\leq \frac{1}{C(W)} \max\left\{\max_{x\in\mathcal{X}} D(W(\cdot|x)||Q(\cdot|x)), E_B(\gamma C(W), Q)\right\} \\ &= \frac{1}{C(W)} \max_{x\in\mathcal{X}} D(W(\cdot|x)||Q(\cdot|x)) \\ &< \infty \end{aligned} \qquad (14)$$

[3] As mentioned in the paragraph preceding the Theorem 2, the restriction that $T(M) \leq A \ln M$ of requirement II is a mild condition provided that $\inf_{Q\in\mathcal{Q}} C(Q) > 0$. For this reason we restrict L to be strictly less than $1/2$.

[4] The result remains true for symmetric channels.

[5] As mentioned in the paragraph preceding the Theorem 2, the restriction that $T(M) \leq A \ln M$ of requirement II is a mild condition provided that $\inf_{Q\in\mathcal{Q}} C(Q) > 0$. For this reason we restrict L to be strictly less than 1.

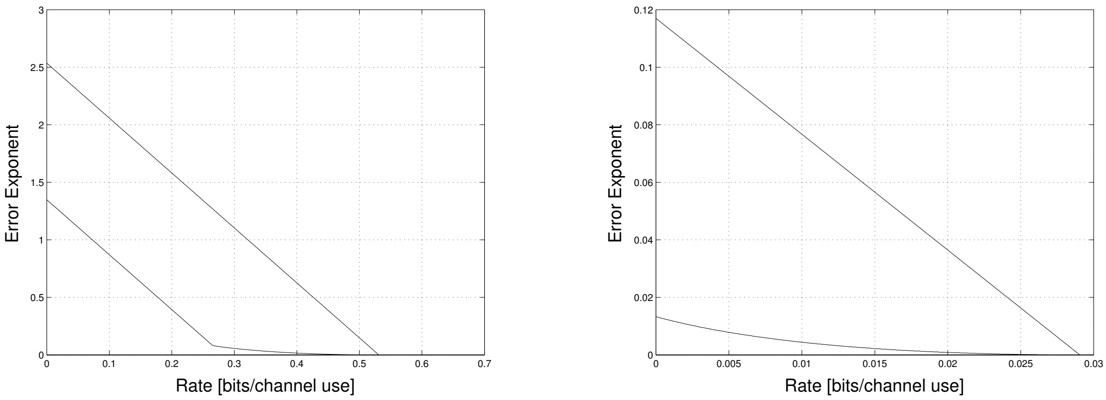

Fig. 1. Upper bound on the error exponent of training based schemes (lower curve) and Burnashev's error exponent (upper line).

where the second equality holds because $E_B\left(\gamma C(W), Q\right) = 0$ since Burnashev's exponent vanishes above capacity, and where the last inequality holds since Q has a nonzero crossover probability. Hence, training based schemes for the Z_L family have a finite error exponent for any $Q \in \mathrm{Z}_L$ with nonzero crossover probability, and for γ sufficiently close to 1. This is in contrast with a result obtained in [9]. Theorem 2 [9] claims that given the family Z_L for some $L \in [0, 1)$ and any constant $\gamma \in [0, 1)$, there exists coding schemes that universally achieve a rate equal to $\gamma C(Q)$ and a corresponding error exponent equal to Burnashev's, in this case infinite. These coding schemes do not separate the estimation of the channel from the information delivery, hence are not training based schemes.

Finally, in the light of the present result and the results in [9], it appears from the previous examples that, at least for BSCs and Z channels, a necessary condition for a universal coding scheme to reach Burnashev's exponent is not to separate channel estimation and the information delivery.

Sketch of the Proof of Theorem 2:[6] We restrict ourselves to the case where $\mathcal{Q} = \mathrm{BSC}_L$ with $L \in [0, 1/2)$. The general case is essentially a straightforward extension of this case.

Without loss of generality we assume the training sequence to be the all–zero sequence of length t. First, one can show that in order to fulfill the requirement II the rate function has to "strongly" rely on the empirical channel that results from the training period. In other words, if during the training period the channel behaves like BSC(δ), from the condition II one deduces that the length of the second phase must be approximately $\frac{\ln M}{\gamma C(\delta)} - t$. Using the fact that the rate function's decision is essentially based on the empirical channel, one shows that a large probability of error occurs because of the atypical behavior of the channel during the training, more specifically a large probability of error occurs whenever during the training the channel behaves as a channel with a higher capacity.

Suppose the true channel is ε, with $\varepsilon \neq 0$, and let δ be such that $C(\delta) \geq C(\varepsilon)$, i.e., $\delta \in [0, \varepsilon]$. We lower bound the error probability of training based schemes as

$$\begin{aligned}
&\mathbb{P}(\text{error}) \\
&\geq \mathbb{P}(\text{error and during the training the channel behaves like } \delta) \\
&= \mathbb{P}\left(\text{error}\middle|\text{during the training the channel behaves like } \delta\right) \\
&\qquad \times \mathbb{P}(\text{during the training the channel behaves like } \delta)\ .
\end{aligned} \tag{15}$$

By a principle of large deviations we have

$$\mathbb{P}(\text{during the training the channel behaves like } \delta) \approx e^{-tD(\delta||\varepsilon)}\ . \tag{16}$$

where $D(\delta||\varepsilon) \triangleq \delta \ln \frac{\delta}{\varepsilon} + (1-\delta)\ln\frac{(1-\delta)}{(1-\varepsilon)}$.

Since the average length of the second phase is approximately equal to $\frac{\ln M}{\gamma C(\delta)} - t$, and since Burnashev's exponent

[6]The proof of Theorem 2 can be found in [8].

yields a lower bound to the error probability we have

$$\mathbb{P}\left(\text{error}\middle|\text{during the training the channel behaves like } \delta\right)$$
$$\gtrapprox e^{-\left(\frac{\ln M}{\gamma C(\delta)}-t\right)E_B\left(\frac{\ln M}{\frac{\ln M}{\gamma C(\delta)}-t},\varepsilon\right)}. \quad (17)$$

From (15), (16), and (17) one deduces the desired result by optimizing the fraction of the communication time dedicated to the training and noticing that δ is arbitrary in $[0, \varepsilon]$.

REFERENCES

[1] M. Brehler, M. K. Varanasi and P. Dayal *Leveraging coherent space-time codes for noncoherent channels via training*, to appear in the IEEE trans. Inform. Theory.

[2] 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.

[3] Dobrushin, *Asymptotic bounds on the probability of error for the transmission of messages over a memoryless channel using feedback*, (in Russian) Probl. Kibern., vol. 8, p. 161–168, 1962.

[4] M. Feder and A. Lapidoth, *Universal decoding for channels with memory*, IEEE trans. Inform. Theory, vol. 44, number 5, p. 1726–1745, 1998.

[5] B. Hassibi and B. M. Hochwald, *How much training is needed in multiple-antenna wireless links?*, IEEE trans. Inform. Theory, vol. 49, number 4, p. 951–963, 2003.

[6] Q. Sun, D. C. Cox, H.C. Huang and A. Lonzano *Estimation of continuous flat fading MIMO channels*, IEEE trans. Inform. Theory, vol. 1, number 4, p. 549–553, 2002.

[7] N. Shulman, *Communication over an unknown channel via common broadcasting*, Ph.D. Thesis, Tel Aviv University, 2003.

[8] A. Tchamkerten and İ. E. Telatar, *On the use of training sequences for channel estimation*, submitted to the IEEE trans. inform. Theory.

[9] A. Tchamkerten and İ. E. Telatar, *Variable length coding over unknown channels*, submitted to the IEEE trans. inform. Theory.

[10] T. F. Wong and B. Park, *Training sequence optimization in MIMO systems with colored interference*, IEEE Transactions on Communications, Nov. 2004.