Message communications and channel state estimation for state dependent channels

Sviatoslav Voloshynovskiy, Oleksiy Koval, Emre Topak, Thierry Pun
CUI, University of Geneva
24 rue du Général-Dufour
1211 Geneva 4
Switzerland
{svolos,koval,topak,pun}@cui.unige.ch

Abstract

In this paper, we consider the problem of pure information transmission and channel state estimation via state dependent channels. We show that the knowledge of auxiliary random variable, used in the codebook construction of random binning techniques, is sufficient to perform the optimal channel state estimation. We compare the obtained results with optimal rate-distortion region obtained using more involved coding strategies based on hybrid random binning and uncoded transmission. This analysis is performed for the generalized Gel'fand-Pinsker formulation and Gaussian Costa setup.

1 Introduction

We consider the problem of pure information transmission and channel state estimation via state dependent channels shown in Fig. 1. This problem has numerous practical applications including digital data-hiding, authentication, wireless communications etc. [1], [2], [3]. In the information-theoretic setup, this problem is dedicated to the trade-off analysis between the amount of reliably communicated independent information given by the message M and the accuracy of channel or host X^N state estimation.

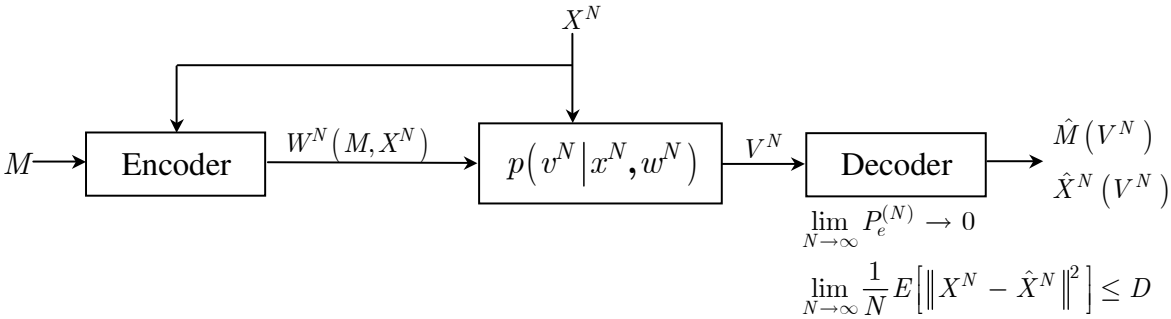


Figure 1: Pure information and channel state transmission over state dependent channels

Notations We use capital letters to denote scalar random variables X, X^N to denote vector random variables, corresponding small letters x and x^N to denote the realizations of scalar and vector random variables, respectively. The superscript N is

used to designate length-N vectors $x^N = [x[1], x[2], ..., x[N]]^T$ with k^{th} element $x[k]$. We use $X \sim p_X(x)$ or simply $X \sim p(x)$ to indicate that a random variable X is distributed according to $p_X(x)$. The mathematical expectation of a random variable $X \sim p_X(x)$ is denoted by $E_{p_X}[X]$ or simply by $E[X]$ and $Var[X]$ denotes the variance of X. Calligraphic fonts $\mathcal{X}$ denote sets $X \in \mathcal{X}$ and $|\mathcal{X}|$ denotes the cardinality of set $\mathcal{X}$. $\mathbf{I}_N$ denotes the $N \times N$ identity matrix. We also define the WIR as WIR=$10\log_{10} \frac{\sigma_W^2}{\sigma_X^2}$ and the WNR as WNR= $10 \log_{10} \frac{\sigma_W^2}{\sigma_Z^2}$, where σ_X^2, σ_W^2, σ_Z^2 represent the variances of channel interference, input to the channel and noise, respectively. We use the definition of the so-called *strongly typical set* [4] p. 358 $A_\delta^{*(N)}(X)$ with respect to $p_X(.)$ that is the set of N-tuples x^N satisfying:

$$A_\delta^{*(N)}(X) = \begin{cases} x^N : \text{ for all } a \in \mathcal{X} \\ \mathbf{N}(a|x^N) = 0, \text{ if } p_X(a) = 0, \\ \left| \frac{1}{N} \mathbf{N}(a|x^N) - p_X(a) \right| < \frac{\delta}{|\mathcal{X}|}, \end{cases} \tag{1}$$

where $\mathbf{N}(a|x^N)$ is the number of a occurrences in the sequence x^N and δ is an arbitrary small positive constant.

The *strongly jointly typical* sequences (x^N, y^N) with respect to the joint distribution $p_{XY}(.)$ on $\mathcal{X} \times \mathcal{Y}$ satisfy:

$$A_\delta^{*(N)}(X,Y) = \begin{cases} (x^N, y^N) : \text{ for all } a \in \mathcal{X}, b \in \mathcal{Y}, \\ \mathbf{N}(a,b|x^N, y^N) = 0, \text{ if } p_{XY}(a,b) = 0, \\ \left| \frac{1}{N} \mathbf{N}(a,b|x^N, y^N) - p_{XY}(a,b) \right| < \frac{\delta}{|\mathcal{X}||\mathcal{Y}|}, \end{cases} \tag{2}$$

where $\mathbf{N}(a,b|x^N, y^N)$ is the number of the pair (a,b) occurrences in the pair of sequences (x^N, y^N).

2 Problem Formulation

The encoder in order to send a message $M \in \mathcal{M}$, $\mathcal{M} = \{1, 2, \ldots, 2^{NR}\}$, where R denotes the rate of communications and N stands for the length of communicated sequences, generates $W^N(M, X^N)$ and transmits it over the state dependent channel $p_{V^N|W^N,X^N} = \prod_{i=1}^N p_{V|W,X}(v_i|w_i, x_i)$. The decoder decodes the send message $\hat{M}$ as well as estimates the channel state $\widehat{X}^N$.

Code construction: Introduce an auxiliary random variable U with an alphabet $\mathcal{U}$ via $p_{U|X}(.)$. Generate $|\mathcal{J}||\mathcal{M}|$ codewords $u^N(m, j)$, $m \in \mathcal{M}$, $j = \{1, 2, \cdots, |\mathcal{J}|\}$ with $|\mathcal{J}| = 2^{NR'}$ independently at random according to the marginal distribution $p_U(.)$ and distribute them into corresponding bins with $|\mathcal{J}|$ in each bin.

Definition 1: (Channel code): A $(2^{NR}, N)$ code for this channel consists of an input message index set $\mathcal{M}$, encoder mapping:

$$\phi^N : \{1, 2, \ldots, 2^{NR}\} \times \mathcal{X}^N \to \mathcal{W}^N, \tag{3}$$

and decoding mappings:

$$\begin{aligned} g^N : \mathcal{V}^N &\to \{1, 2, \ldots, 2^{NR}\}, \\ \psi^N : \mathcal{V}^N &\to \hat{\mathcal{X}}^N. \end{aligned} \tag{4}$$

Definition 2: (Performance measures): The performance is measured by the average probability of error:

$$P_e^{(N)} = \frac{1}{|\mathcal{M}|} \sum_{m \in \mathcal{M}} Pr(g^N(V^N) \neq m | M = m) \tag{5}$$

and by the minimum mean squarred error (MMSE) of estimation:

$$E[d^N(X^N, \hat{X}^N(V^N))] = \frac{1}{N} E[\| X^N - \hat{X}^N(V^N) \|^2], \tag{6}$$

where $d^N(x^N, y^N) = \frac{1}{N} \sum_{i=1}^N d(x_i, y_i)$ and $d(x_i, y_i) : \mathcal{X} \times \mathcal{Y} \to R^+$ and expectation is with respect to the joint probability density function (pdf) $p_{X^N, Y^N}(x^N, y^N)$.

Definition 3: (Rate-distortion pair):
A rate-distortion pair (R, D) is said to be achievable if, and only if, $P_e^{(N)} \to 0$ and $E[d^N(X^N, \hat{X}^N(V^N))] \leq D$ for $N \to \infty$.

For the additive Gaussian channel $V^N = W^N + X^N + Z^N$ with noise $Z^N \sim \mathcal{N}(0, \sigma_Z^2 \mathbf{I}_N)$ and input $W^N \sim \mathcal{N}(0, \sigma_W^2 \mathbf{I}_N)$ and the MMSE distortion measure (6) Sutivong *et al* [1] showed that the optimal (R, D) trade-off is a closure of the convex hull of (R, D) pair satisfying:

$$R \leq \frac{1}{2} \log_2 \left(1 + \frac{\gamma \sigma_W^2}{\sigma_Z^2} \right), \tag{7}$$

$$D \geq \frac{\sigma_X^2 (\gamma \sigma_W^2 + \sigma_Z^2)}{\left(\sigma_X + \sqrt{(1-\gamma)\sigma_W^2} \right)^2 + \gamma \sigma_W^2 + \sigma_Z^2}, \tag{8}$$

where $0 \leq \gamma \leq 1$ is a power sharing factor.

Decoding of the sent message is performed using jointly typical principle, while the best in the MMSE sense host estimate is obtained in the following form: $\hat{X}^N(V^N) = \frac{\sigma_X^2 + \sqrt{(1-\gamma)\sigma_W^2 \sigma_X^2}}{\left(\sqrt{\sigma_X^2} + \sqrt{(1-\gamma)\sigma_W^2} \right)^2 + \gamma \sigma_W^2 + \sigma_Z^2} V^N$ with the estimation distortion given by the right part of (8). The way in which the host-related information is transmitted is known as *uncoded transmission* that consists of host scaling at the encoder to increase the host energy and corresponding MMSE estimator at the decoder.

Evidently, varying γ, it is possible to compromise the rate of pure information transmission and channel state estimation accuracy. In particular, two limiting cases where analyzed in [1]. The first one ($\gamma = 1$) corresponds to the pure information transmission similarly to the Costa setup, while in the second one ($\gamma = 0$) all the available power σ_W^2 is spent for the uncoded channel state communications via state dependent channels.

The rate distortion pairs for these setups are obtained from (7) and (8) for the corresponding values of γ:

$$\gamma = 1 : (R, D) = \left(\frac{1}{2} \log_2 \left(1 + \frac{\sigma_W^2}{\sigma_Z^2} \right), \frac{\sigma_X^2 (\sigma_W^2 + \sigma_Z^2)}{\sigma_X^2 + \sigma_W^2 + \sigma_Z^2} \right), \tag{9}$$

$$\gamma = 0 : (R, D) = \left(0, \frac{\sigma_X^2 \sigma_Z^2}{(\sigma_X + \sigma_W)^2 + \sigma_Z^2} \right). \tag{10}$$

It is important to note that for $\gamma = 1$ (Costa setup), the estimation of the host signal $\hat{X}^N$ is just the MMSE estimation solely based on the received signal V^N. No

side information from the results of message decoding and particularly from U^N, which might be known at the decoder as well as at the encoder, is used to assist estimation. It should be also pointed out that optimal $\alpha = \alpha_{opt}$ was explicitly used in the derivations. Therefore, it is interesting to establish the impact of this side information on the estimation accuracy for the general case of any α.

Contrarily to the previous work, we analyse a coding setup where the encoder is optimized for pure information transmission similar to Gel'fand-Pinsker setup [5], while the channel estimation is considered as a granted option, i.e., as a side product of message communications.

3 Main Results

We consider the codebook construction based on random binning principle. The codebook consists of $2^{N(R+R')}$ auxiliary codewords $U^N \in \mathcal{U}^N$ generated via $p_{U^N|X^N}$, with $2^{NR'}$ codewords in each bin.

In this paper, we want to demonstrate the possibility of using the sequence U^N at the decoder to assist better channel state estimation. We will show that since the codeword u^N selected at the encoder based on the host state x^N in the bin of the communicated message m can be also recovered at the decoder under the condition of reliable pure information communications, this knowledge can be used to assist $\hat{X}^N$ estimation according to:

$$\psi^N : \mathcal{V}^N \times \mathcal{U}^N \to \mathcal{X}^N; \quad \hat{X}^N = E[X^N | V^N, \hat{U}^N]. \tag{11}$$

In the analysis of Gel'fand-Pinsker setup, it is assumed that conditions of reliable message communications are satisfied and $\hat{m} = m$ with $P_e^{(N)} \to 0$ as $N \to \infty$. This implies that given the distorted data v^N, the decoder can uniquely find a jointly typical pair $(u^N(m,j), v^N) \in A_\delta^{*(N)}(U,V)$, and, thus, it can declare that $\hat{m} = m$ and $\hat{u}^N = u^N(m,j)$, where $\hat{u}^N$ corresponds to the estimate of $u^N(m,j)$ used at the encoder. It is possible to show that that upper bound for the reliable message decoding P_e^{Dec} and the corresponding bound for the reliable finding U^N, i.e., P_e^U, coincide thus confirming the possibility of reliable estimate of $\hat{U}^N$ under perfect recovery of message m. In particular, P_e^{Dec} is upper bounded by:

$$\begin{aligned} P_e^{Dec} &< (2^{NR} - 1)2^{NR'}2^{-N[I(U;V)-\delta]}, & (12)\\ &< 2^{N[R+R']}2^{-N[I(U;V)-\delta]}. & (13) \end{aligned}$$

Thus, we require that δ is small, N is large and $R + R' < I(U;V)$ to have $P_e^{Dec} \to 0$.

The probability of error at the decoder in finding u^N used at the encoder can be upper bounded by:

$$\begin{aligned} P_e^U &< (2^{N[R+R']} - 1)2^{-N[I(U;V)-\delta]}, & (14)\\ &< 2^{N[R+R']}2^{-N[I(U;V)-\delta]}. & (15) \end{aligned}$$

Thus, one can assume without loss of generality that the above bound coincides with the upper bound (12). Consequentially, if one selects such $R + R'$ that these bounds asymptotically go to zero, it would be sufficient to consider only generalized case when $\hat{u}^N = u^N$.

Once the estimate $\hat{U}^N$ is found, one can derive the estimate of $\hat{X}^N$ according to (11).

For the Gaussian assumptions, i.e., $X^N \sim \mathcal{N}(0, \sigma_X^2 \mathbf{I}_N)$, $W^N \sim \mathcal{N}(0, \sigma_W^2 \mathbf{I}_N)$ and $Z^N \sim \mathcal{N}(0, \sigma_Z^2 \mathbf{I}_N)$, we have obtained the following estimate $\hat{X}^N$:

$$\hat{X}^N = aV^N + bU^N, \tag{16}$$

where $a = \sigma_X^2 \sigma_W^2 (1-\alpha)(-2\alpha\sigma_W^2\sigma_X^2 + \sigma_X^2\sigma_W^2 + \alpha^2\sigma_W^2\sigma_X^2 + \alpha^2\sigma_Z^2\sigma_X^2 + \sigma_Z^2\sigma_W^2)^{-1}$, $b = \sigma_X^2(\sigma_W^2\alpha + \alpha\sigma_Z^2 - \sigma_W^2)(-2\alpha\sigma_W^2\sigma_X^2 + \sigma_X^2\sigma_W^2 + \alpha^2\sigma_W^2\sigma_X^2 + \alpha^2\sigma_Z^2\sigma_X^2 + \sigma_Z^2\sigma_W^2)^{-1}$ and α stands for Costa optimization parameter [6], selected to maximize the achievable rate $R(\alpha)$:

$$R(\alpha) = \frac{1}{2}\log_2 \frac{\sigma_W^2(\sigma_W^2 + \sigma_X^2 + \sigma_Z^2)}{\sigma_W^2\sigma_X^2(1-\alpha)^2 + \sigma_Z^2(\sigma_W^2 + \alpha^2\sigma_X^2)}, \tag{17}$$

is found to be optimum $\alpha_{opt} = \frac{\sigma_W^2}{\sigma_W^2 + \sigma_Z^2}$.

The variance of this estimator is:

$$D^r(\alpha) = E[d^N(\hat{X}^N, X^N)] = \frac{\sigma_X^2\sigma_W^2\sigma_Z^2}{\alpha^2\sigma_X^2\sigma_Z^2 + \sigma_W^2(\sigma_X^2(1-\alpha)^2 + \sigma_Z^2)}. \tag{18}$$

Therefore, the achievable rate-distortion pair $(R(\alpha), D^r(\alpha))$ is:

$$(R(\alpha), D^r(\alpha)) = (\tfrac{1}{2}\log_2 \tfrac{\sigma_W^2(\sigma_W^2+\sigma_X^2+\sigma_Z^2)}{\sigma_W^2\sigma_X^2(1-\alpha)^2+\sigma_Z^2(\sigma_W^2+\alpha^2\sigma_X^2)}, \tfrac{\sigma_X^2\sigma_W^2\sigma_Z^2}{\alpha^2\sigma_X^2\sigma_Z^2+\sigma_W^2(\sigma_X^2(1-\alpha)^2+\sigma_Z^2)}). \tag{19}$$

It is important to note that when $\alpha = \alpha_{opt}$, the rate distortion pair (19) coincides with (9) for $\gamma = 1$ that supports the results obtained in [1]. However, in the specific cases when $\alpha \neq \alpha_{opt}$, one can obtain quite interesting for practice results.

If $\alpha = 0$, that corresponds for example to the so-called *spread spectrum* data-hiding or communications, (19) reduces to the $(R(\alpha), D^r(\alpha))$ pair:

$$(R(\alpha=0), D^r(\alpha=0)) = \left(\frac{1}{2}\log_2\left(1 + \frac{\sigma_W^2}{\sigma_W^2+\sigma_X^2}\right), \frac{\sigma_X^2\sigma_Z^2}{\sigma_X^2+\sigma_Z^2}\right). \tag{20}$$

If $\alpha = 1$, that corresponds by analogy to the so-called *quantization index modulation*, (19) is:

$$(R(\alpha=1), D^r(\alpha=1)) = \left(\frac{1}{2}\log_2 \frac{\sigma_W^2(\sigma_W^2+\sigma_X^2+\sigma_Z^2)}{\sigma_Z^2(\sigma_W^2+\sigma_X^2)}, \frac{\sigma_X^2\sigma_W^2}{\sigma_X^2+\sigma_W^2}\right), \tag{21}$$

where the achieved rate is higher in comparison to (20) but the distortion depends on the variance of the channel input σ_W^2 and is not asymptotically decreasing with $\sigma_Z^2 \to 0$.

4 Results of computer modeling and conclusions

To confirm the theoretical findings, we have performed the experimental validation of different message communication and channel state estimation scenarios for the Gaussian setup. Figure 2 summarizes the known results for the achievable rates of the Costa setup (17) with different values of optimization parameter α for the WIR equals to -6 dB and -16 dB. In particular, for two asymptotic cases, when $\alpha = 0$ one obtains the performance of spread spectrum communications and when $\alpha = 1$ it corresponds to the quantization index modulation. These results are demonstrated to underline

the critical dependence of the achieved rates on the selection of α. Obviously, the capacity of the AWGN channel is achieved for $\alpha = \alpha_{opt}$ that provides interference-free communications. It should be again pointed out here that the Costa design of α_{opt} aims at maximizing the achievable rate and does not assume any constraints on the channel state estimation accuracy.

In the above analysis we have referred to the generic selection of parameter α. However, actually it depends on the channel input constraint and the noise. Normally in the practice of the wireless communications and data hiding, the actual value of the noise variance is rarely known in advance at the encoder. Thus, α is selected keeping in mind some critical, the least favorable, or average conditions of system applications. This definitely provides the mismatch between the optimal parameter and the actual one that leads to some decrease in the system performance in terms of maximum achievable rate that will be shown by the results of our simulation.

Nevertheless, it was interesting to investigate the hypothetical system performance in terms of channel state estimation accuracy, if one assumes the perfect knowledge of the operational scenario at the encoder that makes possible to choose the optimal parameter according to the Costa result.

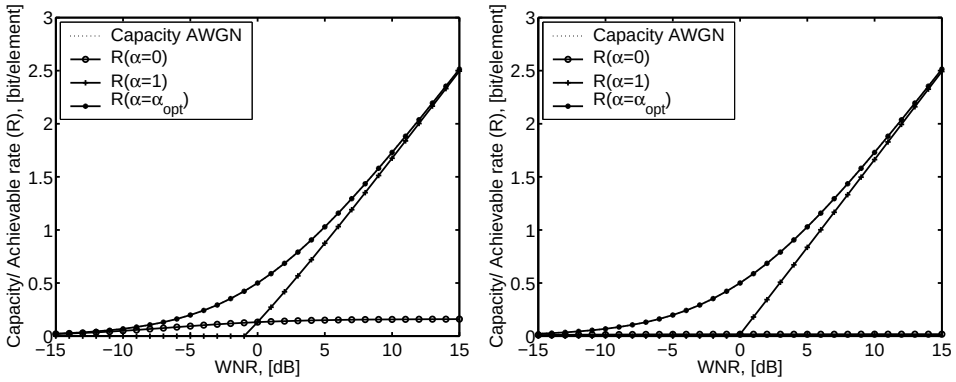


Figure 2: Achievable rate: (a) WIR=-6 dB and (b) WIR=-16 dB.

To investigate the impact of α on the channel state estimation accuracy, we have performed a number of simulations for different assumptions. First, we have applied the direct MMSE estimation without taking into account knowledge of U^N at the decoder. It is interesting to point out that, the variance of this estimate D^r_{MMSE} equals to the variance $D^r(\alpha = \alpha_{opt})$ of the corresponding estimate obtained with the knowledge of U^N that is plotted in Figure 3 for both WIRs. Secondly, assuming the knowledge of U^N, we have computed the variance of the channel state estimation $D^r(\alpha)$ according to (18) for the above asymptotic values of α (Figure 3).

The obtained results confirm the non-optimality of the optimal Costa $\alpha = \alpha_{opt}$ selection for channel state estimation. This behavior is justified by the fact that for $\alpha = 0$ (spread spectrum communications) $U^N = W^N$ and it represents additional interference source for channel state estimation. Therefore, $D^r(\alpha = 0) = \frac{\sigma_X^2 \sigma_Z^2}{\sigma_X^2 + \sigma_Z^2}$ and asymptotically perfect host or channel state recovery at high WNRs ($\sigma_Z^2 \to 0$) is possible that is shown in Figure 3. Contrarily, for $\alpha = 1$, the result is equal to $\frac{\sigma_X^2 \sigma_W^2}{\sigma_X^2 + \sigma_W^2}$ that is independent from σ_Z^2. The result for $\alpha = \alpha_{opt}$ asymptotically converges to $\alpha = 1$

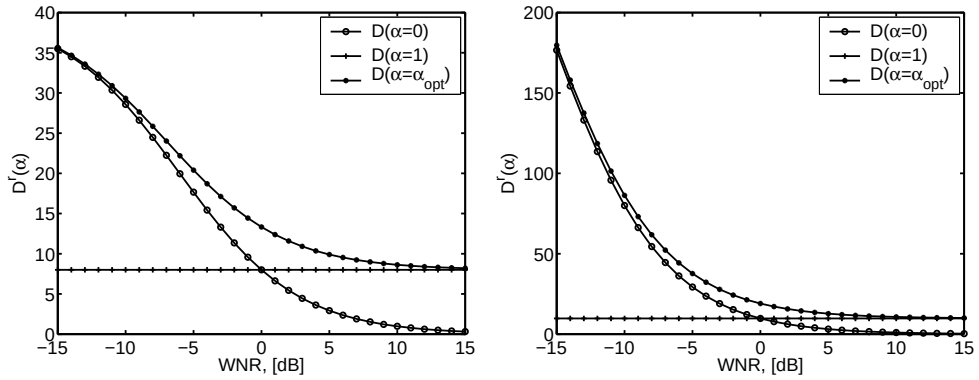


Figure 3: Distortion: (a) WIR=-6 dB and (b) WIR=-16 dB.

as WIR$\rightarrow \infty$.

Finally, we computed the corresponding achievable rate-distortion pairs in the Costa setup and compared it with the optimal results (7) and (8) when power and time(space)-sharing are used. The obtained results are presented in Figure 4. These results demonstrate that depending on maximization of the communication rate or of the estimation distortion is more important in certain application, different scenarios are possible. Evidently, the maximum rate of reliable communications for the whole range of WNRs might be achieved only when $\alpha = \alpha_{opt}$. Oppositely, when more accurate channel state estimation is necessary for the fixed communication rate, deviation from the rate maximization conditions are required.

It is interesting to note that the optimal rate-distortion pair (R, D) obtained using power-sharing can be asymptotically achieved based on the considered Costa scheme with U^N estimate as WIR$\rightarrow \infty$. This result also outperforms the time-sharing setup under considered conditions. Therefore, the knowledge of U^N at the estimator of X^N can help reduce the estimation variance contrarily to a particular case considered in [1] when the case $\alpha = \alpha_{opt}$ was only analyzed using only random binning coding.

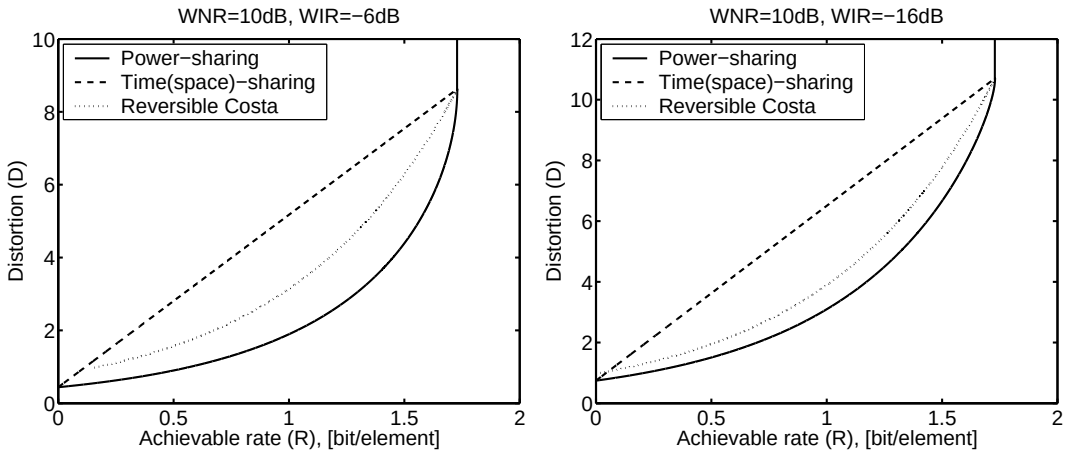
Acknowledgment

This paper was partially supported by SNF Professeur Boursier grant PP002–68653, by the European Commission through the IST Programme under contract IST-2002-507932-ECRYPT, FP6-507609-SIMILAR and Swiss IM2 projects.

The information in this document reflects only the authors views, is provided as is and no guarantee or warranty is given that the it is fit for any particular purpose. The user thereof uses the information at its sole risk and liability.

References

[1] A. Sutivong, M. Chiang, T.M. Cover, and Y.-H. Kim. Channel capacity and state estimation for state-dependent gaussian channels. *IEEE Trans. on Information Theory*, 51(4):1486–1495, April 2005.



(a) (b)

Figure 4: Optimal (R, D) trade-off regions for the Gaussian setup.

[2] E. Martinian, G. W. Wornell, and B. Chen. Authentication with distortion criteria. *IEEE Trans. on Information Theory*, pages 2523–2542, July 2005.

[3] F. M. J. Willems and T. Kalker. Coding theorems for reversible embedding. In *Proc. DIMACS Series in Discrete Mathematics and Theoretical Computer Science; American Mathematical Society*, volume 66, pages 61–76, 2004.

[4] T. Cover and J. Thomas. *Elements of Information Theory.* Wiley and Sons, New York, 1991.

[5] S.I. Gel'fand and M.S. Pinsker. Coding for channel with random parameters. *Probl. Control and Inf. Theory*, 9(1):19–31, 1980.

[6] M. Costa. Writing on dirty paper. *IEEE Trans. on Information Theory*, 29(3):439–441, May 1983.