DIRTY PAPER CODING WITH PARTIAL STATE INFORMATION

Çağatay Dikici, Khalid Idrissi, Atilla Baskurt

LIRIS, INSA Lyon, France

Christine Guillemot, Caroline Fontaine

TEMICS, IRISA Rennes, France

ABSTRACT

A generalization of the problem of dirty paper coding is considered in which (possibly different) noisy versions of the state information, assumed to be i.i.d. Gaussian random variables, are available at the encoder and at the decoder. This paper derives the maximum achievable rate formula for this general problem. This general setup encompasses the cases where the state information is perfectly known either at the encoder or decoder or at both. Moreover, it generalizes the analysis to cases where the state information is known only partially. In addition, this paper shows that in realistic situations where the AWGN noise power is not known at the encoder, partial information at the decoder can increase the maximum achievable rate with respect to Costa's coding setup. Our setup is relevant for practical applications such as watermarking under desynchronization attacks.

1. INTRODUCTION

The problem of coding for communication over a channel whose conditional probability distribution is controlled by a random state parameter finds applications in diverse areas ranging from coding for information storage in a memory with defective cells, data hiding, and coding for multiple input multiple output communication. The particular case where the state is known causally at the encoder only (no channel state information at the receiver) has been first considered by Shannon in 1958 [1]. In [2], Gel'fand and Pinsker consider the channel coding problem with non-causal state information available at the transmitter. In their setup, the transmitter wishes to send a message $M \in \{1, ..., |\mathcal{M}|\}$ over a memoryless channel defined by the transition probabilities $p(y|x, s)$, where X and Y are the channel input and output and S is an i.i.d. random variable representing the sequence of states $\{S_1, \dots S_N\}$ of the channel known non-causally at the encoder but unknown at the decoder. In [3], Costa has shown that there is no loss in capacity if the channel state is additive white Gaussian interference ("dirt"). The design of codes for approaching Costa's capacity is known as the dirty paper coding problem. The capacity loss is derived in [4] for additive white Gaussian channel state S partially available at the encoder but not to the decoder. The impact of the dirty paper coding with no state information at the encoder and partial information at the decoder is studied in [5]. The capacity for information storage in a memory where the channel state is perfectly available at the decoder but not to the encoder is derived in [6]. The authors in [7] consider a generalized Gel'fand-Pinsker coding problem and derive capacity formulas, as well as random coding and sphere packing exponents.

In this paper, we focus on the particular problem of dirty paper coding with correlated partial state information at the encoder and at the decoder. The Gel'fand-Pinsker coding problem where (possibly different) noisy versions of the channel sequence are available at both sides has actually been first considered in [8] for a binary input - binary output channel. The targeted application was information storage in a memory with defecting cells. Here, the problem we focus on can be regarded as a special case of a coding problem with two-sided state information examined in [9]. The state information, the channel input and output are assumed to be i.i.d. Gaussian random variables. The maximum achievable rates formulas are derived for this general problem as a function of α by expressing as in [3] $U = X + \alpha S$ where U is an auxiliary random variable. This will give us the general capacity formula for the cases where there is only a partial or a null side information at the encoder side, while there is a perfect one at the decoder side. The analytic expressions of capacity/maximum achievable rate gains and losses with respect to Costa's set up are given for six particular cases with optimum and non optimum values of the α parameter. It is shown that in the general case, a capacity gain or loss can be obtained in a realistic situation where the optimum α is not known.

2. PROBLEM STATEMENT

Consider the communication problem shown in Figure 1. We use the same notation as [3] throughout this paper. An index $M \in \{1, ..., |\mathcal{M}|\}$ will be sent to the receiver in n uses of the channel, where $|\mathcal{M}|$ is the greatest integer smaller than or equal to e^{nR}, and R is the rate in nats per transmission. Let $S = (S_1, S_2, ..., S_n)$ be the sequence of noncausal state of the channel for n transmissions, assumed to be a sequence of independent identically distributed (i.i.d.) $\mathcal{N}(0, Q)$ random variables. We consider the cases where this sequence of states is partially known to the encoder $S_1 = (S_{1,1}, S_{1,2}, ..., S_{1,n})$ and to the decoder $S_2 = (S_{2,1}, S_{2,2}, ..., S_{2,n})$ noncausally and expressed as S_1 and S_2 throughout this paper. This problem can be cast into a two-sided state information set-up close to the one considered in [9], where S is defined by a pair of

independent and identically distributed (i.i.d.) correlated state information (S_1, S_2) available at the sender and at the receiver respectively. The state information available at the encoder and at the decoder is expressed in terms of the channel state as $S_1 = S + \theta$ and $S_2 = S + T$, where θ and T are i.i.d. random variables according to $\mathcal{N}(0, L)$ and $\mathcal{N}(0, K)$.

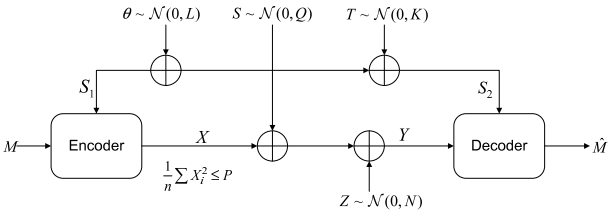


Figure 1. Channel Coding with state informations.

Based on M and S_1, the encoder sends a codeword X, which must satisfy the power constraint $(1/n)\sum_{i=1}^{n} X_i^2 \leq P$. The channel output is given by $Y = X + S + Z$, where the channel noise Z is i.i.d. according to $\mathcal{N}(0, N)$. Upon receipt of Y and S_2, the decoder creates an estimate $\hat{M}(Y, S_2)$ of the index M. Under the assumption that the index M is uniformly distributed over $\{1, .., M\}$, the probability of error P_e, is given by

$$P_e = \frac{1}{M}\sum_{k=1}^{M} Pr\left\{\hat{M}(Y, S_2) \neq k | M = k\right\}. \quad (1)$$

The general formula for the capacity of this set-up in the case of finite alphabets is given by [9]:

$$C = \max_{p(x,u|s_1)} [I(U; Y, S_2) - I(U; S_1)], \quad (2)$$

where the maximum is over all joint distributions of $p(x, u|s_1)$, where U is an auxiliary random variable with finite cardinality. But, in our case the alphabets are continuous and the only general capacity expression that has been stated is [7]:

$$C = \sup_{p(x,u|s_1)} \min_{p(y|x,s)} [I(U; Y, S_2) - I(U; S_1)], \quad (3)$$

So, here we will be interested in the estimation of the maximum achievable rate for particular distributions and constructions, and see that in some cases it can be identified with the capacity.

The perfect codes can be created as in [9] using the random binning argument. First, $e^{n(I(U;Y,S_2)-2\epsilon)}$ i.i.d. sequences of U are generated according to distribution $p(u)$ and each of them is indexed as $U(i)$ where $i \in \{1, 2, ..., e^{n(I(U;Y,S_2)-2\epsilon)}\}$. Then these sequences are randomly distributed into $e^{n(R-4\epsilon)}$ bins where R corresponds to the rate of the system. Given the state $S_1 = S + \theta$ and the message $M \in \{1, ..., |\mathcal{M}|\}$, the encoder searches the codeword $U(i)$ within the bin indexed by M such that the pair $(U(i), S_1)$ is jointly typical. Then it sends the corresponding X which is jointly typical with $(U(i); S_1)$. During the transmission, the signal is exposed to the additive interference S and Z. The receiver receives $Y = X + S + Z$ from the channel and observes the noncausal state information $S_2 = S + T$. The decoder searches for the sequence $U(i)$ such that $(U(i), Y, S_2)$ is strongly jointly typical and assigns $\hat{M}$ as the index of the bin containing the sequence $U(i)$. All possible error events go to 0 as $n \to \infty$ [9].

3. ACHIEVABLE RATE

We assume that X, S, Z, θ and T are random variables with respective Gaussian distributions $\mathcal{N}(0, P)$, $\mathcal{N}(0, Q)$, $\mathcal{N}(0, NI)$, $\mathcal{N}(0, LI)$, and $\mathcal{N}(0, K)$. Hence, the joint distribution $f(X, S, Z, \theta, T)$ is a multivariate Gaussian $\sim \mathcal{N}(0, \Sigma)$ where the covariance matrix is $\mathbf{\Sigma} = \text{diag}(P, Q, N, L, K)$. We consider $U = X + \alpha S_1 = X + \alpha S + \alpha\theta$ where α is a parameter to be determined.

The achievable rate is then function of the parameter α and is given by $R(\alpha) = I(U; Y, S_2) - I(U; S_1)$, where

$$\begin{bmatrix} U \\ Y \\ S_2 \end{bmatrix} = \begin{pmatrix} 1 & \alpha & 0 & \alpha & 0 \\ 1 & 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 \end{pmatrix} \cdot \begin{bmatrix} X \\ S \\ Z \\ \theta \\ T \end{bmatrix}. \quad (4)$$

Then,

$$\mathbf{B}\cdot\mathbf{\Sigma}\cdot\mathbf{B}^{\mathrm{t}} = \begin{pmatrix} (P + \alpha^2(Q + L)) & (P + \alpha Q) & \alpha Q \\ (P + \alpha Q) & (P + Q + N) & Q \\ \alpha Q & Q & (Q + K) \end{pmatrix}. \quad (5)$$

Hence, the joint entropy of the random variables $(U; Y; S_2)$ is

$$h(U; Y; S_2) = 1/2 \ln\left((2\pi e)^3 |\mathbf{B}\mathbf{\Sigma}\mathbf{B}^{\mathrm{t}}|\right). \quad (6)$$

The relevant mutual informations can be calculated to yield

$$\begin{aligned} I(U; Y, S_2) &= h(U) + h(Y; S_2) - h(U; Y; S_2) \\ &= h(X + \alpha S + \alpha\theta)) + h(X + S + Z; S + T) \\ &\quad - h(U; Y; S_2) \\ &= \frac{1}{2}\ln\left((2\pi e)(P + \alpha^2(Q + L))\right) \\ &\quad + \frac{1}{2}\ln\left((2\pi e)^2((P + Q + N)(Q + K) - Q^2)\right) \\ &\quad - \frac{1}{2}\ln\left((2\pi e)^3(PQK(1-\alpha)^2 + NK(P + \alpha^2(Q + L)) + \alpha^2 L(PQ + PK + QK + NQ) + PNQ)\right) \end{aligned} \quad (7)$$

and similarly

$$I(U; S_1) = h(U) + h(S+\theta) - h(U; S+\theta) = \frac{1}{2}\ln\left(\frac{P + \alpha^2(Q + L)}{P}\right). \quad (8)$$

Then $R(\alpha)$ yields

$$R(\alpha) = \frac{1}{2}\ln\left(\frac{P((P+Q+N)(Q+K) - Q^2)}{PQK(1-\alpha)^2 + NK(P + \alpha^2(Q+L)) + \alpha^2 L(PQ + PK + QK + NQ) + PNQ}\right). \quad (9)$$

Similarly to [3], the graphs of $R(\alpha)$ versus α are presented in Figure 2 where $P = Q = N = 1$ and for several {L,K} pairs such as $\{0,0\}$, $\{0,1\}$, $\{1,0\}$, $\{1,1\}$, $\{0,\infty\}$ and $\{1,\infty\}$. The graph $R(\alpha)$ is tangent to the line of the capacity term where $K = L = 0$ if and only if at least one of the state information accessible to the decoder or to the encoder is perfect. Otherwise there exists a capacity loss.

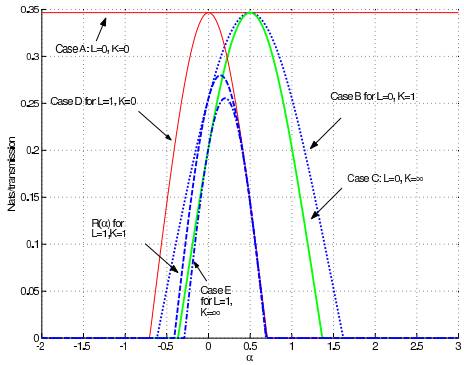


Figure 2. $P = Q = N = 1$, Graphs of $R(\alpha)$ for {L,K} pairs $\{0,0\}$, $\{0,1\}$, $\{1,0\}$, $\{1,1\}$, $\{0,\infty\}$ and $\{1,\infty\}$. The rate of transmission $R(\alpha)$ is calculated in nats per unit transmission (Maximum value 0.3466 nats/transmission corresponds to 1 bit/transmission).

Maximizing $R(\alpha)$ over α, we get

$$\max_{\alpha} R(\alpha) = \frac{1}{2}\ln\left(1+\frac{P(QK+QL+KL)}{N(QK+QL+KL)+QLK}\right), \tag{10}$$

which is obtained for $\alpha^* = PQK/(PQK + QNK + L(PQ + PK + QK + NQ + NK))$.

4. REMARKS

Our general general setup can be reduced into six known different cases as listed in below by modifying the variance of the correlation noises θ and T.

Case A: $S_1 = S,\ S_2 = S$. This corresponds to the encoder-decoder state pair [perfect, perfect] where $K \to 0$ and $L \to 0$. Then the achievable rate is [1]

$$R_{\text{Case-A}} = \lim_{K\to 0,\ L\to 0} R(\alpha) = \frac{1}{2}\ln\left(1+\frac{P}{N}\right) \tag{11}$$

which is independent of α and reaches C^*, hence showing that it is in fact the capacity, and that the capacity is achieved by this construction.

Case B: $S_1 = S,\ S_2 = S + T$. This corresponds to the encoder-decoder state pair [perfect, partial] where $L \to 0$. The achievable rate of the system is given by

$$R_{\text{Case-B}}(\alpha)=\lim_{L\to 0} R(\alpha)=\frac{1}{2}\ln\left(\frac{P(K(P+Q+N)+Q(P+N))}{PQK(1-\alpha)^2+NK(P+\alpha^2Q)+PNQ}\right). \tag{12}$$

$R_{\text{Case-B}}(\alpha)$ is maximized for $\alpha^\diamond = P/(P+N)$ which corresponds to a rate of $R_{\text{Case-B}}(\alpha^\diamond) = \frac{1}{2}\ln\left(1+\frac{P}{N}\right) = C^*$.

Case C: $S_1 = S,\ S_2 = S + T$ which corresponds to the encoder-decoder state pair [perfect, $\emptyset$] where $L \to 0$ and $K \to \infty$.

The achievable rate becomes

$$R_{\text{Case-C}}(\alpha)=\lim_{K\to\infty,\ L\to 0} R(\alpha)=\frac{1}{2}\ln\left(\frac{P(P+Q+N)}{PQ(1-\alpha)^2+N(P+\alpha^2Q)}\right) \tag{13}$$

This rate is maximized for $\alpha^\diamond = P/(P+N)$ then giving $R_{\text{Case C}}(\alpha^\diamond) = \frac{1}{2}\ln\left(1+\frac{P}{N}\right) = C^*$. As Costa showed, the capacity is then reached.

Case D: $S_1 = S + \theta,\ S_2 = S$. The encoder-decoder state pair is [partial, perfect] where $K \to 0$. The achievable rate in this case is

$$R_{\text{Case-D}} = \lim_{K\to 0} R(\alpha) = \frac{1}{2}\ln\left(\frac{P(P+N)}{\alpha^2L(P+N)+PN}\right). \tag{14}$$

The rate $R_{\text{Case-D}}$ is independent of the state power Q. It is maximized for $\alpha^\nabla = 0$ which corresponds to a maximum achievable rate of $R_{\text{Case-D}}(\alpha^\nabla) = \frac{1}{2}\ln\left(1+\frac{P}{N}\right) = C^*$.

Case E: $S_1 = S + \theta,\ S_2 = S + T$. The encoder-decoder state pair is [partial, $\emptyset$] where $K \to \infty$. For this setup the rate is

$$R_{\text{Case-E}}(\alpha)=\lim_{K\to\infty} R(\alpha)=\frac{1}{2}\ln\left(\frac{P(P+Q+N)}{PQ(1-\alpha)^2+N(P+\alpha^2(Q+L))+\alpha^2L(P+Q)}\right), \tag{15}$$

It is maximized for $\alpha^\dagger = PQ/(PQ + QN + LP + LQ + LN)$ which corresponds to a rate of

$$R_{\text{Case-E}}(\alpha^\dagger) = \frac{1}{2}\ln\left(1+\frac{P(Q+L)}{N(Q+L)+QL}\right). \tag{16}$$

Please note that there exists a loss in Case E with respect to Case A ($R_{\text{Case-E}}(\alpha^\dagger) < R_{\text{Case-A}}$). Here, we cannot state that $R_{\text{Case-E}}(\alpha^\dagger)$ corresponds to a capacity: it is the maximum achievable rate for our construction. Zaidi et. al. [4] analyze the capacity loss of a setup similar to the Case E such that the channel state S is not perfectly available to the encoder and is defined by $S = S1 + \theta$ where in our case $S_1 = S + \theta$.

Case F: $S_1 = \emptyset,\ S_2 = S$. The encoder-decoder state pair is [$\emptyset$, perfect] where $K \to 0$ and $\alpha = 0$. For this setup the rate is

$$R_{\text{Case-F}} = \lim_{K\to 0} R(0) = \frac{1}{2}\ln\left(1+\frac{P}{N}\right). \tag{17}$$

Since there is no state information available at the encoder, the auxiliary variable U is $U = X$. Please remark that the capacity is reached, stating its value for this case, and showing that this construction enables to achieve it. The graphs of cases A, B, C and D are given in Figure 2 for $P = Q = N = L = 1$.

5. CAPACITY/RATE GAIN / LOSS ANALYSIS

In this section, we analyze the capacity analysis of Dirty paper codes with partial state information at the encoder and decoder sides given in Equation 9, and the special cases of this setup which are given in Section 3. Moreover the rate gain/loss between the special cases where the encoder does not have knowledge to the optimum coding parameter α.

5.1. For optimum values of α

If the transmitter uses the optimum value of the α parameter for each setup, there is no capacity gain nor loss for the particular cases A,B,C, D and F. The achievable capacity in that case is given by

$$R_{\text{Case A}}=R_{\text{Case B}}(\alpha^{\diamond})=R_{\text{Case C}}(\alpha^{\diamond})=R_{\text{Case D}}(\alpha^{\nabla})=\tfrac{1}{2}\ln\left(1+\tfrac{P}{N}\right)=C^{*}. \quad (18)$$

In Case E, the optimum value of α yields a maximum achievable rate loss:

$$R_{\text{loss Case E}}=R_{\text{Case-E}}(\alpha^{\dagger})-R_{\text{Costa}}(\alpha^{\diamond})=-\tfrac{1}{2}\ln\left(1+\tfrac{PQL}{N((Q+L)(P+N)+QL)}\right). \quad (19)$$

Similarly, for the optimum value of α, the maximum achievable rate loss for the general case is

$$R_{\text{loss general}}=R(\alpha^{*})-R_{\text{Costa}}(\alpha^{\diamond})=-\tfrac{1}{2}\ln\left(1+\tfrac{PQLK}{N((P+N)(QK+QL+LK)+QLK)}\right). \quad (20)$$

5.2. For non optimum values of α

However, in actual systems, the transmitter does not have perfect knowledge of the additive variances N, Q, L, and K, so can not always code with the optimum α parameter. Assuming that the coding is done with a non-optimum α, we analyze the rate gain or loss with respect to Costa's coding setup [perfect,$\emptyset$] . For instance, for using the same non-optimal α, there exists a rate gain in Case B [perfect, partial] with respect to Costa's setup which is given by:

$$C_{\text{gain CaseB-C}}(\alpha)=\max\{0,R_{\text{Case-B}}(\alpha)\}-\max\{0,R_{\text{costa}}(\alpha)\} \quad (21)$$

Let us define the Signal to State Ratio (SSR) and Signal to Noise Ratio (SNR) as SSR $= 10\ \log_{10}\left(\frac{P}{Q}\right)$ and SNR $= 10\ \log_{10}\left(\frac{P}{N}\right)$.

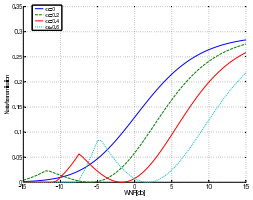
The graphs showing the capacity gains between Case B and Costa's setup for an SNR value ranging between -15 dB and $+15$ dB for different values of the α parameter and of $10\log(Q/K)$ (2.1 dB and 6 dB) in Figure 3. We fix $P = 1$, $L = 0$, SSR $= -6$ dB[1]. We observe that, given the values of P, Q, K and fixing α, the capacity is maximized for a certain SNR value such that $R_{\text{Case-B}}(\alpha) = R_{\text{costa}}(\alpha^{\diamond})$, hence there is 0 capacity gain for that SNR value. However, for other SNR values, there always exists a capacity gain with respect to $R_{\text{costa}}(\alpha)$. It is also evident that, given fixed P, Q, N values and an estimate of α, decreasing the $10\log(Q/K)$ value decreases the capacity gain.

In Case E [partial, $\emptyset$], without optimum α at the transmitter, there is a maximum achievable rate loss with respect to Costa's setup, given by

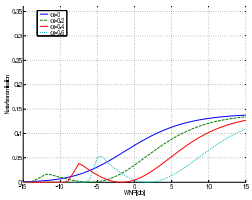
$$C_{\text{loss CaseE-C}}(\alpha)=\max\{0,R_{\text{Case-E}}(\alpha)\}-\max\{0,R_{\text{costa}}(\alpha)\} \quad (22)$$

Finally, without optimum α parameter, there exists a maximum achievable rate gain or loss between the general case [partial,partial] and Costa's setup [perfect,$\emptyset$] expressed as a function of P, N, Q, L, K and α. Figure 4 shows the maximum achievable rate gain/loss versus SNR graph for an SNR value ranging between -15 dB and 15 dB. Please note that $P = L = 1$, SSR $= -6$ dB, $10\log(Q/K) = 2.1$ dB (for Figure 4(a)) and $10\log(Q/K) = 6$ dB (for Figure 4(b)).

[1]Such low SSR values are relevant for practical application such as watermarking.

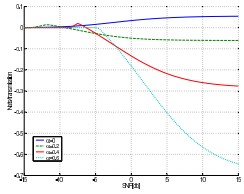


(a) Capacity gain for $10\log(Q/K)= 2.1$ dB.

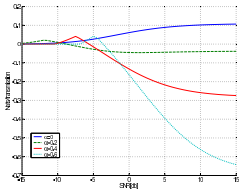


(b) Capacity gain for $10\log(Q/K)= 6$ dB.

Figure 3. Capacity gain (between $R_{\text{Case-B}}(\alpha)$ and $R_{\text{Costa}}(\alpha)$) versus SNR, for different α values where $P = 1$, SSR$= -6$ dB and various $10\log(Q/K)$ values, with perfect knowledge of the channel state information at the encoder ($L = 0$).



(a) Maximum achievable rate gain/loss for $10\log(Q/K)= 2.1$ dB.



(b) Maximum achievable rate gain/loss for $10\log(Q/K)= 6$ dB.

Figure 4. Maximum achievable rate gain or loss (between $R(\alpha)$ and $R_{\text{Costa}}(\alpha)$) versus SNR, for different α values where P$= 1$, SSR$= -6$ dB and various $10\log(Q/K)$ values, with partial knowledge of the channel state information at the encoder (L=1).

6. CONCLUSION

In this paper, we analyzed the maximum achievable rate losses and gains for the general setup where the partial state information is available at the encoder and at the decoder under Gaussian interference. In particular, we derived the capacity for the case [partial or $\emptyset$,perfect], showing that Costa's construction enables to reach it; this is not the case of [partial,partial or $\emptyset$], for which only a maximum achievable rate has been stated. We then analyzed the gain/loss in terms of achievable rates if the optimal coding parameter α is not accessible to the encoder. This general setup is relevant for practical applications such as watermarking under desynchronization attacks. Our perspective is to evaluate our theoretical findings by a practical code design on watermarking of still images and video.

7. REFERENCES

[1] C. E. Shannon, "Channels with side information at the transmitter," in *IBM J. of Research and Development*, 1958, vol. 2, pp. 289–293.

[2] S. I. Gel'fand and M. S. Pinsker, "Coding for channel with random parameters," *Prob. Contr. Inform. Theory*, vol. 9, no. 1, pp. 19–31, 1980.

[3] M. Costa, "Writing on dirty paper (corresp.)," *IEEE Trans. Inf. Theory*, vol. 29, no. 3, pp. 439–441, 1983.

[4] A. Zaidi and P. Duhamel, "On coding with a partial knowledge of the state information," in *Proceedings of the IEEE 39th Asilomar conference on Signals, Systems and Computers*, 2005, pp. 657–661.

[5] P. Piantanida and P. Duhamel, "Dirty-paper coding without channel information at the transmitter and imperfect estimation at the receiver," in *Proceedings of the IEEE International Conference on Communications (ICC '07)*, 2007, pp. 5406–5411.

[6] C. Heegard and A. E. Gamal, "On the capacity of computer memory with defects," *IEEE Trans. Inf. Theory*, vol. 29, no. 5, pp. 731–739, 1983.

[7] P. Moulin and Y. Wang, "Capacity and random-coding exponents for channel coding with side information," *IEEE Trans. Inf. Theory*, vol. 53, no. 4, pp. 1326–1347, 2007.

[8] M. Salehi, "Capacity and coding for memories with real-time noisy defect information at encoder and decoder," in *Proceedings of the IEE Communication, Speech and Vision*, vol. 139, no. 2, 1992, pp. 113–117.

[9] T. M. Cover and M. Chiang, "Duality between channel capacity and rate distortion with two-sided state information," *IEEE Trans. Inf. Theory*, vol. 48, no. 6, pp. 1629–1638, 2002.