Optimal Filter Banks for Multiple Description Coding: Analysis and Synthesis

Pier Luigi Dragotti, *Member, IEEE*, Sergio D. Servetto, *Member, IEEE*, and Martin Vetterli, *Fellow, IEEE*

Abstract—Multiple description (MD) coding is a source coding technique for information transmission over unreliable networks. In MD coding, the coder generates several different descriptions of the same signal and the decoder can produce a useful reconstruction of the source with any received subset of these descriptions. In this paper, we study the problem of MD coding of stationary Gaussian sources with memory. First, we compute an approximate MD rate distortion region for these sources, which we prove to be asymptotically tight at high rates. This region generalizes the MD rate distortion region of El Gamal, Cover, and Ozarow for memoryless Gaussian sources. Then, we develop an algorithm for the design of optimal two-channel biorthogonal filter banks for MD coding of Gaussian sources. We show that optimal filters are obtained by allocating the redundancy over frequency with a reverse "water-filling" strategy. Finally, we present experimental results which show the effectiveness of our filter banks in the low complexity, low rate regime.

Index Terms—Filter bank design, integer-to-integer transforms, multiple description (MD) coding, rate distortion functions, robust source coding.

I. INTRODUCTION

A. The Problem of Multiple Description Source Coding

RECENTLY, the problem of transmitting data over heterogeneous packet switched networks has received considerable attention. Packet losses can be due to transmission errors or congestion. If the network is able to provide preferential treatment to some packets, then the use of multiresolution or layered source coding systems is the obvious solution. But if the network cannot differentiate among packets, and if retransmissions are not allowed (e.g., due to real-time delay constraints or in multicast scenarios), then the source coding strategy should be different. Multiple description (MD) coding offers a potentially attractive framework in which to develop coding algorithms for such scenarios. An MD coder represents an information source using multiple bit streams (descriptions). Each individual description provides an approximation to the original message, and multiple descriptions can refine each other, to produce a better approximation than that attainable by any single one alone.
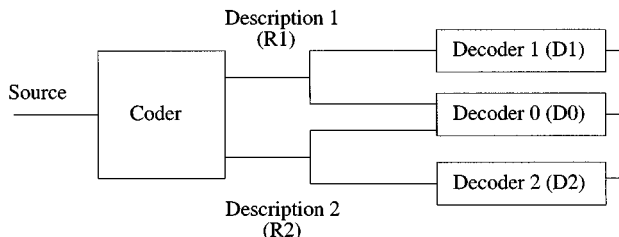
Fig. 1. The multiple description problem.

The simplest formulation of the MD problem is illustrated in Fig. 1 and involves only two descriptions. This is the so-called case of two channels and three receivers. If both descriptions are received then the decoder can reconstruct the source at some small distortion value D_0 (the *central* distortion), but if either one is lost, the decoder can still reconstruct the source at some higher distortion D_1 or D_2 (the *side* distortions).

B. Information-Theoretic Performance Bounds

In a way analogous to classical rate-distortion problems in source coding theory, in MD source coding one is also interested in finding a suitable MD rate-distortion region, defined as the set of quintuples $(R_1, R_2, D_0, D_1, D_2)$ for which there exist codes of rates R_1 and R_2 achieving average distortions D_0, D_1, and D_2. Clearly, we can state that the rate R_1 necessary to achieve the distortion D_1 cannot be smaller than $R(D_1)$ ($R(\cdot)$ is the rate-distortion function for the source); similar arguments apply for the other two cases, so we can state that a first bound for the MD rate region is

$$R_1 \geq R(D_1) \tag{1}$$
$$R_2 \geq R(D_2) \tag{2}$$
$$R_1 + R_2 \geq R(D_0). \tag{3}$$

In general, it is not possible to achieve equality simultaneously in the three equations since two individually good descriptions tend to be similar to each other. Thus, the second description will contribute very little to improve the quality of the first one. On the other hand, two descriptions which are complementary cannot be individually good. Since, in general, $R_1 + R_2 > R(D_0)$ the resulting excess rate is usually called *redundancy* ρ. One possible formulation of the problem of MD coding then consists in minimizing the side distortion, given some allowed redundancy ρ.

Early papers on MD coding are information-theoretic in nature and try to find the set of achievable values for the quintuple $(R_1, R_2, D_0, D_1, D_2)$. El Gamal and Cover determined an achievable rate region for general memoryless sources [10],

while Ozarow showed that this region is tight for the case of memoryless Gaussian sources and squared error distortions [23]. Ahlswede [1] studied the case of *no excess rate* (when there is equality in (3)), and Zhang and Berger [39] considered the *no excess marginal rate* case (when there is equality in (1), (2)). Zhang and Berger also showed by counterexample that in the excess rate case the achievable region of El Gamal and Cover is not tight [38]. More recently, Linder *et al.* [20] found a rate region for memoryless sources and locally quadratic distortion measure which is tight in the limit of small distortions (high bit rate). Finally, Zamir [35], [36] extended the Shannon bounds [3] to the MD case and showed that for a Gaussian source the outer bounds are asymptotically tight.

C. Code Constructions

Several efforts have also been made to design practical MD coding systems. In [27], a design procedure for the construction of fixed-rate scalar quantizers was presented. In [29], that design procedure was extended to the entropy-constrained case. It is shown in [28] that at high rates, for the case of balanced descriptions ($R_1 = R_2 = R$) and Gaussian sources, the distortion product $D_0 D_1$ of the entropy-constrained MD scalar quantizer (MDSQ) takes the form $\frac{1}{4}(\frac{\pi e}{6})^2 2^{-4R}$. At the same time, the MD rate distortion bound (when put in distortion product form) becomes $\frac{1}{4} 2^{-4R}$. This is an important result because it shows that for the MDSQ both the side and the central distortion attain the optimal exponential rate of decay ($D_0 \sim 2^{-2R}$, $D_1 \sim 2^{-2R}$). The only suboptimality of MDSQ at high rates is due to the use of a scalar quantizer which partitions the space into cubic regions instead of an ideal vector quantizer that would optimally partition the space into spheres. Various constructions of MD vector quantizers have been proposed [9], [11], [19], [30] and the MD lattice quantizers of [30] do effectively close the gap between the performance of the entropy constrained MDSQ and the MD rate-distortion bound.

A rather different approach pioneered by Wang *et al.* [22], [33] and then extended by Goyal and Kovačević [13] consists of applying a suitable block-wise transform to the input vector before coding to obtain the MD property. This approach is usually called MD transform coding. The basic idea is to decorrelate the vector components and then to introduce again correlation between coefficients, but in a known and controlled manner, so that erased coefficients can be statistically estimated from those received. Techniques based on overcomplete frame expansions have been proposed in [6], [15], [21].

Most of the previous work on MD coding focuses on the case of memoryless sources or sources with finite memory. In [18], Ingle *et al.* consider the problem of designing differential pulse-code modulation (DPCM) systems for MD coding of sources with memory. Batllo *et al.* considered a similar problem and proposed a solution that combines the use of an orthogonal block transform and of the MDSQ [2]. As in the memoryless case, this system has some good asymptotic properties. At low rates, however, except for some practical results obtained in the context of still image coding [24], much less is known. Note that Batllo and Vaishampayan use the term Multiple Description Transform Coder (MDTC) to refer to this system. From now on, we will also use that name to refer to their system.

D. Contributions and Paper Organization

In this work, we consider sources with infinite memory, specifically, wide-sense stationary Gaussian sources and consider the classical MD scenario of two channels and three receivers. We present two new results: an MD rate-distortion region for stationary Gaussian sources which is asymptotically tight at high rates (Theorem 1 of Section II) and an algorithm for the design of optimal two-channel filter banks for MD coding of Gaussian sources (Theorem 2 of Section III). The filter banks are designed using an approach similar to the one proposed in the case of block transforms: we construct a first filter bank to decorrelate the two input sequences and then we use a second filter bank to efficiently recorrelate them. The frequency responses of this second filter bank depend on the total amount of available redundancy and on the allocation strategy of the redundancy over frequency. In Theorem 2, it is shown that the optimal allocation of the redundancy over frequency is obtained using a reverse water-filling strategy.

In a recent paper [34], Yang and Ramchandran have independently worked on the same problem of designing filter banks for MD Coding. The main difference between their work and the one presented here in Section III is that we have moved the quantization step before the transform and approximated the continuous transform with a discrete one. In practical applications it is very important to put the quantization before a nonorthogonal transform so that the square partition cells are maintained (see, for instance, [12], [14], [22], [32]). This different approach leads to a different formulation of the optimization problem and, in our case, to the analytical solution presented in Theorem 2, which is not present in [34].

In Section IV, we assess the performance of our system. We compare it against the ideal bounds, the MDTC [2] and the MD-DPCM system [18]. For this comparison we consider two possible scenarios. The first one is high rates and infinite delay/complexity. This first comparison is interesting because it is under these hypotheses that the optimization problem is stated and solved both in [2] and in this paper. The second scenario is based on practical requirements of low rates and finite delay/complexity. In this regime, performance of the systems considered will not be predicted well by the theory, and there is much to be learned by means of numerical simulations. We conclude in Section V.

II. ASYMPTOTIC MD RATE REGION

For a memoryless Gaussian source with variance σ^2, Ozarow [23] gave an explicit characterization of the set of achievable distortions (D_0, D_1, D_2) for a given pair of rates R_1, R_2:

$$D_1 \geq \sigma^2 \cdot 2^{-2R_1} \tag{4}$$

$$D_2 \geq \sigma^2 \cdot 2^{-2R_2} \tag{5}$$

$$D_0 \geq \frac{\sigma^2 \cdot 2^{-2(R_1+R_2)}}{1 - \left(\sqrt{\Pi} - \sqrt{\Delta}\right)^2} \tag{6}$$

where

$$\Pi = \left(1 - D_1/\sigma^2\right)\left(1 - D_2/\sigma^2\right)$$

and

$$\Delta = \left(D_1 D_2 / \sigma^4\right) - 2^{-2(R_1 + R_2)}.$$

The inverse of these functions are the following [20], [23]:

$$R_1 \geq \frac{1}{2} \log\left(\frac{\sigma^2}{D_1}\right) \qquad (7)$$

$$R_2 \geq \frac{1}{2} \log\left(\frac{\sigma^2}{D_2}\right) \qquad (8)$$

$$R_1 + R_2 \geq \frac{1}{2} \log\left(\frac{\sigma^2}{D_1}\right) + \frac{1}{2} \log\left(\frac{\sigma^2}{D_2}\right) + \delta \qquad (9)$$

where δ is defined by

$$\delta = \begin{cases} \frac{1}{2} \log\left(\frac{1}{1-\rho^2}\right), & D_0 \leq D_0^{\max} \\ 0, & D_0 > D_0^{\max} \end{cases} \qquad (10)$$

where

$$D_0^{\max} = \frac{D_1 D_2}{D_1 + D_2 - (D_1 D_2 / \sigma^2)} \qquad (11)$$

and

$$\rho = -\frac{\sqrt{\Pi \epsilon_0^2 + \gamma} - \sqrt{\Pi \epsilon_0^2}}{(1 - \epsilon_0)\sqrt{\epsilon_1 \epsilon_2}}$$
$$\gamma = (1 - \epsilon_0)[(\epsilon_1 - \epsilon_0)(\epsilon_2 - \epsilon_0) + \epsilon_0 \epsilon_1 \epsilon_2 - \epsilon_0^2]$$
$$\Pi = (1 - \epsilon_1)(1 - \epsilon_2)$$
$$\epsilon_i = D_i / \sigma^2 \qquad (i = 0, 1, 2).$$

Notice that δ depends on the three distortions (D_0, D_1, D_2) and on the variance σ^2 (see (10)) However, by rearranging (9), one can see the relationship between δ and the rates R_1 and R_2 and interpret δ as the excess rate that is used to reduce the central distortion given the two side distortions or

$$\delta = R_1 - \frac{1}{2} \log\left(\frac{\sigma^2}{D_1}\right) + R_2 - \frac{1}{2} \log\left(\frac{\sigma^2}{D_2}\right).$$

Now, if $\delta = 0$ then

$$R_1 = \frac{1}{2} \log\left(\frac{\sigma^2}{D_1}\right) \quad \text{and} \quad R_2 = \frac{1}{2} \log\left(\frac{\sigma^2}{D_2}\right).$$

This means that all the rate is used to minimize the side distortions and in this case D_0 equals its maximum value $(D_0^{\max})$. If $\delta > 0$ it means that part of the rate is used to reduce the central distortion which becomes smaller than $D_0^{\max}$. In particular, D_0 decreases from $D_0^{\max}$ to zero as δ increases from zero to infinity. This is why δ is also called the excess marginal rate [35], [39].

Consider now the high rate situation, namely, the case where the three distortions (D_0, D_1, D_2) are very small compared to the variance σ^2 or, in other words, the case where the three ratios $D_0/\sigma^2, D_1/\sigma^2, D_2/\sigma^2$ go to zero. In this situation, the excess marginal rate δ (see (10)) and the maximum central distortion $D_0^{\max}$ (see (11)) do not depend on the variance of the source anymore. In particular, we have [20]

$$\delta_{HR}(D_0, D_1, D_2) \triangleq \lim_{\sigma \to \infty} \delta(\sigma^2, D_0, D_1, D_2)$$
$$\triangleq \lim_{\lambda \to 0} \delta(\sigma^2, \lambda D_0, \lambda D_1, \lambda D_2)$$
$$= \frac{1}{2} \log\left(\frac{1}{1 - \rho_{HR}^2}\right) \qquad (12)$$

where

$$\rho_{HR} = -\frac{\left(\sqrt{D_1/D_0 - 1}\right)\left(\sqrt{D_2/D_0 - 1}\right) - 1}{\sqrt{D_1 D_2 / D_0^2}}$$

and

$$\lim_{\sigma \to \infty} D_0^{\max} = \frac{D_1 D_2}{D_1 + D_2}.$$

Now, based on these preliminary results we can state the following theorem.

Theorem 1: In the limit of small distortions (i.e., D_0, D_1, $D_2 \to 0$), the asymptotic multiple description rate region for a stationary Gaussian source and mean square error (MSE) distortions is given by the following equations:

$$R_1 \geq \frac{1}{4\pi} \int_{-\pi}^{\pi} \log\left(\frac{S(\omega)}{D_1}\right) d\omega$$
$$R_2 \geq \frac{1}{4\pi} \int_{-\pi}^{\pi} \log\left(\frac{S(\omega)}{D_2}\right) d\omega$$
$$R_1 + R_2 \geq \frac{1}{4\pi} \left(\int_{-\pi}^{\pi} \log\left(\frac{S(\omega)}{D_1}\right) d\omega \right.$$
$$\left. + \int_{-\pi}^{\pi} \log\left(\frac{S(\omega)}{D_2}\right) d\omega + 2\delta_{HR} \right)$$

where $S(\omega)$ is the power spectral density of the Gaussian source.

Proof: Let $\{X_t, t = 0, \pm 1, \ldots\}$ be a discrete-time stationary Gaussian source. We begin by considering N successive elements of this source and by calculating the asymptotic MD rate region of this N-sequence. Call Φ_N the $N \times N$ correlation matrix related to any N successive components of $\{X_t\}$. Since the source is stationary, Φ_N is a symmetric Toeplitz matrix. Apply a Karhunen–Loève transform (KLT) to the N-sequence to get uncorrelated (and so independent) components. Because the KLT is unitary and invertible and we are considering MSE as our distortion measure, the problem of finding the MD rate region in the new coordinates is identical to that in the original ones, except that the new components are statistically independent. Call $Y = (y_1, y_2, \ldots, y_N)$ the N-dimensional vector with independent components obtained after applying the KLT to the original N-sequence and call $\hat{Y}_0, \hat{Y}_1, \hat{Y}_2$ the appropriate reproducing vectors at the three receivers. Moreover, $\hat{y}_{ki}$ will represent the ith component of the reproducing vector $\hat{Y}_k$, $k = 0, 1, 2$ and Y^i is a vector with the first i elements of Y. Extending the El Gamal–Cover results [10] to the vector case, we have that the MD rate region is given by

$$R_1 \geq \frac{1}{N} \min I\left(Y; \hat{Y}_1\right) \qquad (13)$$

$$R_2 \geq \frac{1}{N} \min I\left(Y; \hat{Y}_2\right) \qquad (14)$$

$$R_1 + R_2 \geq \frac{1}{N} \min\left(I\left(Y; \hat{Y}_0, \hat{Y}_1, \hat{Y}_2\right) + I\left(\hat{Y}_1; \hat{Y}_2\right) \right) \qquad (15)$$

where the minima are over all the probability density functions $p(Y, \hat{Y}_0, \hat{Y}_1, \hat{Y}_2)$ satisfying

$$E\left[\frac{1}{N} \sum_{i=1}^{N} (y_i - \hat{y}_{ki})^2 \right] \leq D_k, \qquad k = 0, 1, 2.$$

First notice that the term $I(Y; \hat{Y}_0, \hat{Y}_1, \hat{Y}_2) + I(\hat{Y}_1; \hat{Y}_2)$ in (15) can be equivalently expressed as

$$I\left(Y; \hat{Y}_0, \hat{Y}_1, \hat{Y}_2\right) + I\left(\hat{Y}_1; \hat{Y}_2\right) \tag{16}$$

$$= I\left(Y; \hat{Y}_0, \hat{Y}_1, \hat{Y}_2\right) + H\left(\hat{Y}_1\right) + H\left(\hat{Y}_2\right) - H\left(\hat{Y}_1, \hat{Y}_2\right) \tag{17}$$

$$= I\left(Y; \hat{Y}_0, \hat{Y}_1, \hat{Y}_2\right) + I\left(Y; \hat{Y}_1\right) + I\left(Y; \hat{Y}_2\right) - H\left(\hat{Y}_1, \hat{Y}_2\right) \tag{18}$$

where in the last equality we have used the fact that $\hat{Y}_1$, $\hat{Y}_2$ are deterministic function of Y and thus $H(\hat{Y}_k) = I(Y; \hat{Y}_k)$, $k = 1, 2$. In the rest of the proof we will use (18) rather than (16). Now, the first thing we want to show is that the MD rate region of this N-vector reduces to the sum of the MD rate region of each component of Y and that the problem of minimizing (13)–(15) reduces to the problem of finding the right allocation strategy of the rates R_1, R_2 to the different components. Consider, first, (13), it results [7]

$$I\left(Y; \hat{Y}_1\right) = h(Y) - h\left(Y \middle| \hat{Y}_1\right) \tag{19}$$

$$= \sum_{i=1}^{N} h(y_i) - \sum_{i=1}^{N} h\left(y_i \middle| Y^{i-1}, \hat{Y}_1\right) \tag{20}$$

$$\geq \sum_{i=1}^{N} h(y_i) - \sum_{i=1}^{N} h(y_i | \hat{y}_{1i}) \tag{21}$$

$$= \sum_{i=1}^{N} I(y_i; \hat{y}_i) \tag{22}$$

$$\geq \sum_{i=1}^{N} \frac{1}{2} \log\left(\frac{\lambda_i^2}{D_{1i}}\right) \tag{23}$$

where (20) follows from the independence of the components y_i and from the chain rule for entropy. The inequality in (21) follows from the fact that conditioning reduces entropy and we can achieve equality by choosing $p(Y | \hat{Y}_1) = \prod_{i=1}^{N} p(y_i | \hat{y}_{1i})$. The last inequality follows from the expression of the rate distortion function of a Gaussian source and equality can be achieved by choosing each $\hat{y}_{1i} \sim \mathcal{N}(0, \lambda_i^2 - D_{1i})$, where λ_i^2 is the variance of the ith component and $D_{1i} = E[(y_i - \hat{y}_{1i})^2]$ is the distortion related to that component. Hence, from (19)–(23), we get that the minimization in (13) reduces to [7]

$$R_1 \geq \frac{1}{N} \min \sum_{i=1}^{N} \frac{1}{2} \log\left(\frac{\lambda_i^2}{D_{1i}}\right) \tag{24}$$

where the minimum, now, is over all the possible distortions D_{1i} such that

$$E\left[\frac{1}{N} \sum_{i=1}^{N} (y_i - \hat{y}_{1i})^2\right] = \frac{1}{N} \sum_{i=1}^{N} D_{1i} \leq D_1.$$

Similar arguments apply to (14) and that minimization reduces to

$$R_2 \geq \frac{1}{N} \min \sum_{i=1}^{N} \frac{1}{2} \log\left(\frac{\lambda_i^2}{D_{2i}}\right) \tag{25}$$

where the minimum is over all the distortions D_{2i} such that $\frac{1}{N} \sum_{i=1}^{N} D_{2i} \leq D_2$. Consider now (15) and its alternative representation in (18). Consider, first, the term $I(Y; \hat{Y}_0, \hat{Y}_1, \hat{Y}_2)$; following the same procedure as in (19)–(23) we have

$$I\left(Y; \hat{Y}_0, \hat{Y}_1, \hat{Y}_2\right)$$
$$= h(Y) - h\left(Y \middle| \hat{Y}_0, \hat{Y}_1, \hat{Y}_2\right) \tag{26}$$

$$= \sum_{i=1}^{N} h(y_i) - \sum_{i=1}^{N} h\left(y_i \middle| Y^{i-1}, \hat{Y}_0, \hat{Y}_1, \hat{Y}_2\right) \tag{27}$$

$$\geq \sum_{i=1}^{N} h(y_i) - \sum_{i=1}^{N} h(y_i | \hat{y}_{0i}, \hat{y}_{1i}, \hat{y}_{2i}) \tag{28}$$

$$= \sum_{i=1}^{N} I(y_i; \hat{y}_{0i}, \hat{y}_{1i}, \hat{y}_{2i}) \tag{29}$$

where inequality in (28) follows from the fact that conditioning reduces entropy and equality is achieved if

$$p(Y | \hat{Y}_0, \hat{Y}_1, \hat{Y}_2) = \prod_{i=1}^{N} p(y_i | \hat{y}_{0i}, \hat{y}_{1i}, \hat{y}_{2i}).$$

For the second term of (18), we obtain

$$I\left(Y; \hat{Y}_1\right) + I\left(Y; \hat{Y}_2\right) - H\left(\hat{Y}_1, \hat{Y}_2\right) \tag{30}$$

$$\geq \sum_{i=1}^{N} I(y_i; \hat{y}_{1i}) + \sum_{i=1}^{N} I(y_i; \hat{y}_{2i}) - H\left(\hat{Y}_1, \hat{Y}_2\right) \tag{31}$$

$$\geq \sum_{i=1}^{N} I(y_i; \hat{y}_{1i}) + \sum_{i=1}^{N} I(y_i; \hat{y}_{2i}) - \sum_{i=1}^{N} H(\hat{y}_{1i}, \hat{y}_{2i}) \tag{32}$$

where inequality in (31) follows from (19)–(22) and we can achieve equality by choosing

$$p(Y | \hat{Y}_k) = \prod_{i=1}^{N} p(y_i | \hat{y}_{ki}), \qquad \text{for } k = 1, 2.$$

The last inequality in (32) follows from the fact that

$$H(\hat{Y}_1, \hat{Y}_2) \leq \sum_{i=1}^{N} H(\hat{y}_{1i}, \hat{y}_{2i})$$

and equality is achieved if

$$p(\hat{Y}_1, \hat{Y}_2) = \prod_{i=1}^{N} p(\hat{y}_{1i}, \hat{y}_{2i}).$$

Thus, combining the results from (26)–(29) and (30)–(32), we have

$$I\left(Y; \hat{Y}_0, \hat{Y}_1, \hat{Y}_2\right) + I\left(Y; \hat{Y}_1\right) + I\left(Y; \hat{Y}_2\right) - H\left(\hat{Y}_1, \hat{Y}_2\right) \tag{33}$$

$$\geq \sum_{i=1}^{N} (I(y_i; \hat{y}_{0i}, \hat{y}_{1i}, \hat{y}_{2i}) + I(y_i; \hat{y}_{1i})$$
$$+ I(y_i; \hat{y}_{2i}) - H(\hat{y}_{1i}, \hat{y}_{2i})) \tag{34}$$

$$\geq \sum_{i=1}^{N} \frac{1}{2} \log\left(\frac{\lambda_i^2}{D_{1i}}\right) + \frac{1}{2} \log\left(\frac{\lambda_i^2}{D_{2i}}\right) + \delta_i \tag{35}$$

where the last inequality comes from the Ozarow equations and equality can be achieved by a correct choice of each triple ($\hat{y}_{0i}$, $\hat{y}_{1i}$, $\hat{y}_{2i}$). This choice depends on the three distortions D_{0i}, D_{1i}, D_{2i} and for an explicit characterization refer to [23]. Equations (33)–(35) show that minimization in (15) reduces to

$$R_1 + R_2 \geq \min\left(\frac{1}{2N}\sum_{i=1}^{N}\log\left(\frac{\lambda_i^2}{D_{1i}}\right)\right.$$
$$\left. + \frac{1}{2N}\sum_{i=1}^{N}\log\left(\frac{\lambda_i^2}{D_{2i}}\right) + \frac{1}{N}\delta_i\right) \quad (36)$$

where the minimum, now, is over all the distortions D_{0i}, D_{1i}, D_{2i} such that $\frac{1}{N}\sum_{i=1}^{N}D_{ki} \leq D_k$, $k = 0, 1, 2$. Now, combining (24), (25), and (36), we can see that the MD rate region of Y reduces, indeed, to the sum of the MD rate region of each component and that the original minimization problem reduces to

$$R_1 \geq \min_{\frac{1}{N}\sum_i D_{1i}=D_1} \frac{1}{2N}\sum_{i=1}^{N}\log\left(\frac{\lambda_i^2}{D_{1i}}\right) \quad (37)$$

$$R_2 \geq \min_{\frac{1}{N}\sum_i D_{2i}=D_2} \frac{1}{2N}\sum_{i=1}^{N}\log\left(\frac{\lambda_i^2}{D_{2i}}\right) \quad (38)$$

$$R_1 + R_2 \geq \min_{\substack{\frac{1}{N}\sum_i D_{0i}=D_0 \\ \frac{1}{N}\sum_i D_{1i}=D_1 \\ \frac{1}{N}\sum_i D_{2i}=D_2}}\left(\frac{1}{2N}\sum_{i=1}^{N}\log\left(\frac{\lambda_i^2}{D_{1i}}\right)\right.$$
$$\left. + \frac{1}{2N}\sum_{i=1}^{N}\log\left(\frac{\lambda_i^2}{D_{2i}}\right) + \frac{1}{N}\delta_i\right). \quad (39)$$

Thus, the problem now is to understand how each component should contribute to the total distortion to minimize the quantities in (37)–(39) or, stated in a different way, the problem is to understand how the rates R_1, R_2 should be allocated to the various components to minimize (37)–(39). Using Lagrange multipliers we can construct the following three functionals:

$$J_1 = \frac{1}{2N}\sum_{i=1}^{N}\log\left(\frac{\lambda_i^2}{D_{1i}}\right) + \nu_1\sum_i D_{1i} \quad (40)$$

$$J_2 = \frac{1}{2N}\sum_{i=1}^{N}\log\left(\frac{\lambda_i^2}{D_{2i}}\right) + \nu_2\sum_i D_{2i} \quad (41)$$

$$J_3 = \frac{1}{2N}\sum_{i=1}^{N}\log\left(\frac{\lambda_i^2}{D_{1i}}\right) + \frac{1}{2N}\sum_{i=1}^{N}\log\left(\frac{\lambda_i^2}{D_{2i}}\right)$$
$$+ \frac{1}{N}\delta_i + \nu_0\sum_i D_{0i} + \nu_1\sum_i D_{1i} + \nu_2\sum_i D_{2i}. \quad (42)$$

The problem of minimizing the first two functionals is equivalent to the problem of finding the optimal allocation strategy for the single description case. Differentiating with respect to D_{1i} and D_{2i} and setting equal to zero, we have

$$\frac{\partial J_1}{\partial D_{1i}} = -\frac{1}{2N}\frac{1}{D_{1i}} + \nu_1 = 0 \quad (43)$$

$$\frac{\partial J_2}{\partial D_{2i}} = -\frac{1}{2N}\frac{1}{D_{2i}} + \nu_2 = 0 \quad (44)$$

and

$$D_{1i} = C_1 = D_1 \quad (45)$$
$$D_{2i} = C_2 = D_2 \quad (46)$$

where C_1 and C_2 are constants. Hence, the optimum allocation of the rates to the various components results in equal distortion for each component [3], [7]. This is due to the fact that the slopes of the curves (40), (41) are independent of the variances. This argument is not valid for the third functional (42) since the slope of δ depends on the variance. However, in the limit of small distortions, δ becomes independent of the variance ($\delta \to \delta_{HR}$) and the minimization strategy for the third functional becomes the same as for the first two functionals (i.e. $D_{1i} = D_1$, $D_{2i} = D_2$, $D_{0i} = D_0$, $i = 1, 2, \ldots, N$). Then the MD rate region becomes

$$R_1 \geq \frac{1}{2N}\sum_{i=1}^{N}\log\left(\frac{\lambda_i^2}{D_1}\right)$$

$$R_2 \geq \frac{1}{2N}\sum_{i=1}^{N}\log\left(\frac{\lambda_i^2}{D_2}\right)$$

$$R_1 + R_2 \geq \frac{1}{2N}\sum_{i=1}^{N}\log\left(\frac{\lambda_i^2}{D_1}\right) + \frac{1}{2N}\sum_{i=1}^{N}\log\left(\frac{\lambda_i^2}{D_2}\right) + \delta_{HR}.$$

Notice that, since the Ozarow's MD rate region is achievable and tight [23], then also our MD rate region is (asymptotically) achievable and tight. Indeed, we have seen that the MD rate-distortion functions (13)–(15) of the vector Y are lower-bounded by the sum of the MD rate-distortion functions of each component y_i and this lower bound is achieved by coding each component independently. Now, since the direct and converse part of the Ozarow theorem apply to each component, the minima in (37)–(39) not only represent an achievable region, but they also represent a tight region.

Now, using the result of the Toeplitz distribution theorem [3], [16] (see the Appendix), we can go to the limit of infinite N and find the MD rate region of the complete source $\{X_t\}$

$$R_1 \geq \frac{1}{4\pi}\int_{-\pi}^{\pi}\log\left(\frac{S(\omega)}{D_1}\right)d\omega$$

$$R_2 \geq \frac{1}{4\pi}\int_{-\pi}^{\pi}\log\left(\frac{S(\omega)}{D_2}\right)d\omega$$

$$R_1 + R_2 \geq \frac{1}{4\pi}\left(\int_{-\pi}^{\pi}\log\left(\frac{S(\omega)}{D_1}\right)d\omega\right.$$
$$\left. + \int_{-\pi}^{\pi}\log\left(\frac{S(\omega)}{D_2}\right)d\omega + 2\delta_{hr}\right). \quad \square$$

A similar result in terms of the entropy rate power of the Gaussian source can be found in [35], [36]. In these papers, Zamir extended the Shannon bounds [3] to the MD case and then showed that the outer bound is asymptotically tight. His results are valid either for a memoryless source or for a source with memory.

Theorem 1 shows that at high rates the single description allocation strategy is also optimal in the MD case. That is because
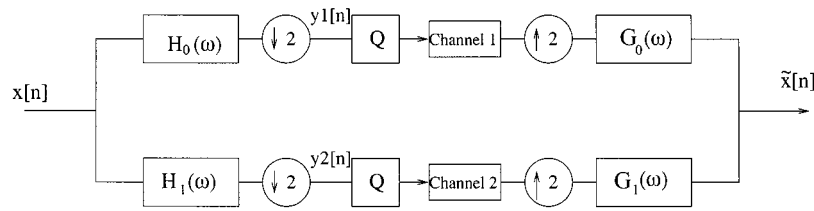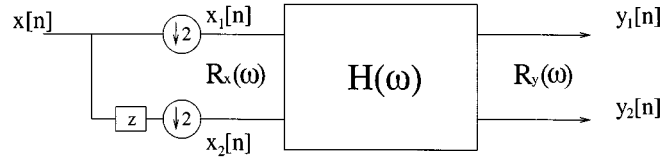
Fig. 2.   Two-channel filter bank.

Fig. 3.   The polyphase representation of the analysis stage.

the slopes of the three functionals (40)–(42) are independent of the source. At low rates, this last assumption is not valid. The functional (42) has a slope dependent on the input source variance and, in general, it is not minimized with a single description allocation strategy. So we can state the following corollary.

Corollary 1: Under a high rate assumption, and for stationary Gaussian sources, the single description rate allocation strategy is also optimal in the MD case. At low rates it is, in general, suboptimal.

III. OPTIMAL TWO-CHANNEL FILTER BANKS FOR MD CODING

A. Problem Formulation and Notation

Consider the classical two-channel filter bank scheme shown in Fig. 2. Here, the input $x[n]$ is assumed to be a stationary Gaussian random process with known statistics and is fed through an analysis filter bank. The two output sequences are then separately quantized and sent over two different erasure channels. We suppose that the channels are independent, that they have the same erasure probability and that $R_1 = R_2$.[1] For convenience, we will formulate our problem in the polyphase domain [26], [31]. In this case, the analysis stage can be equivalently represented by the block scheme shown in Fig. 3.

First we move the quantization step before the transform and approximate our continuous polyphase transform with a discrete one.[2] The discrete transform can be obtained by factoring the continuous one into a product of lifting steps and then sequentially rounding all these intermediate factors [4], [8]. It can be shown that the error due to this approximation can be bounded and that it goes to zero at high rates [12], [14]. The reason why we use this kind of structure is that if the quantization is performed before the transform, then the square partition cells are maintained. This enables the use of nonorthogonal transforms without increasing the quantization error. The importance of performing quantization before the transform in the MD case was pointed out for the first time in [22] (see also [32]). Since at high rates the difference between the discrete and the continuous transforms is small, our analysis will be based on properties of the continuous transform.

Now consider again Fig. 3. The input–output relation can be expressed in matrix notation introducing the analysis polyphase matrix $H(\omega)$

$$\begin{pmatrix} Y_1(\omega) \\ Y_2(\omega) \end{pmatrix} = \begin{pmatrix} H_{11}(\omega) & H_{12}(\omega) \\ H_{21}(\omega) & H_{22}(\omega) \end{pmatrix} \begin{pmatrix} X_1(\omega) \\ X_2(\omega) \end{pmatrix}. \quad (47)$$

Call $R_x(\omega)$ the 2×2 polyphase power spectral density (p.s.d.) matrix of the input process. Likewise, $R_y(\omega)$ is the p.s.d. matrix of the outputs. The system response has the following form:

$$R_y(\omega) = H(\omega) R_x(\omega) H^+(\omega) \quad (48)$$

where $H^+(\omega)$ denotes the Hermitian transpose of $H(\omega)$.

The synthesis part of the system can be analyzed in a similar fashion. Recall that, given the analysis matrix, the synthesis polyphase matrix $G(\omega)$ is uniquely defined (up to a phase factor). In fact, $G(\omega)$ must be such that the condition $G(\omega)H(\omega) = I$ is satisfied [31].

Now, assume that the target central distortion is D_0 and that both channels are coded independently. Since $y_1[n]$, $y_2[n]$ are stationary Gaussian sources and quantization is fine, the minimum bit rates necessary to scalar code the two sequences is [3]

$$R_1(D_0) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{1}{2} \log \frac{R_{y_{11}}(\omega)}{D_0} \, d\omega + \frac{1}{2} \log_2\left(\frac{\pi e}{6}\right)$$

$$R_2(D_0) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{1}{2} \log \frac{R_{y_{22}}(\omega)}{D_0} \, d\omega + \frac{1}{2} \log_2\left(\frac{\pi e}{6}\right). \quad (49)$$

In case we do not use any filter decomposition, the bit rate necessary to get the same central distortion D_0 is [3]

$$R^*(D_0) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{1}{2} \log \frac{S(\omega)}{D_0} \, d\omega + \frac{1}{2} \log_2\left(\frac{\pi e}{6}\right) \quad (50)$$

where $S(\omega)$ is the p.s.d. of the input process. We call redundancy the difference rate between these two cases

$$\rho = \frac{R_1(D_0) + R_2(D_0)}{2} - R^*(D_0)$$

$$= \frac{1}{4\pi} \int_{-\pi}^{\pi} \frac{1}{2} \log \frac{R_{y_{11}}(\omega) R_{y_{22}}(\omega)}{S^2(\omega)}. \quad (51)$$

---

[1]This last hypothesis, although reasonable, is not strictly necessary; but it simplifies the solution.

[2]By continuous transform we mean a generic linear operator in $l_2(\mathcal{Z})$. The discrete transform is a perfectly invertible operator that converts quantized sequences into quantized sequences [4], [17], [37].
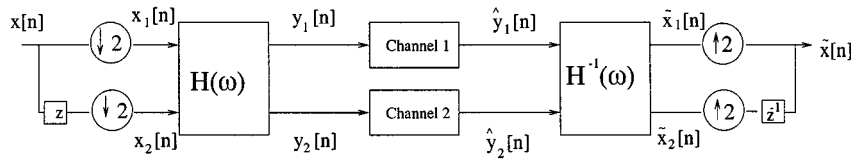
Fig. 4.   The complete MD system in the polyphase domain.

Note that (49) holds because the transform is performed after the quantization. If the transform were performed before the quantization, the shape of the quantization cells would be affected and one should also consider this effect to compute the correct rates. This is one of the limitations of the approach taken in [34].

Now consider the case when one channel (i.e., channel 1) is cut off and $y_1[n]$ must be estimated from the received sequence $y_2[n]$. The optimal estimation is obtained by Wiener filtering

$$\hat{Y}_1(\omega) = \frac{R_{y_{12}}(\omega)}{R_{y_{22}}(\omega)} Y_2(\omega). \tag{52}$$

Call $\eta(\omega)$ the error in predicting $Y_1(\omega)$ from $Y_2(\omega)$

$$\eta(\omega) = Y_1(\omega) - \hat{Y}_1(\omega). \tag{53}$$

Since we have used a nonorthogonal transform, we must return to the original space in order to compute the distortion (MSE distortion in our case); therefore (see also Fig. 4)

$$\begin{pmatrix} \tilde{X}_1(\omega) \\ \tilde{X}_2(\omega) \end{pmatrix} = \begin{pmatrix} G_{11}(\omega) & G_{12}(\omega) \\ G_{21}(\omega) & G_{22}(\omega) \end{pmatrix} \begin{pmatrix} Y_1(\omega) + \eta(\omega) \\ Y_2(\omega) \end{pmatrix}$$
$$= \begin{pmatrix} X_1(\omega) \\ X_2(\omega) \end{pmatrix} + \begin{pmatrix} G_{11}(\omega)\eta(\omega) \\ G_{21}(\omega)\eta(\omega) \end{pmatrix} \tag{54}$$

and

$$\left\| X(\omega) - \tilde{X}(\omega) \right\|^2 = \left\| \begin{pmatrix} G_{11}(\omega)\eta(\omega) \\ G_{21}(\omega)\eta(\omega) \end{pmatrix} \right\|^2. \tag{55}$$

Considering the fact that the error η is still a Gaussian process with p.s.d.

$$R_{y_{11}}(\omega) - \frac{|R_{y_{12}}(\omega)|^2}{R_{y_{22}}(\omega)}$$

and using Parseval's relation we obtain

$$D_2 = E\left[\|x[n] - \tilde{x}[n]\|^2\right]$$
$$= \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{1}{2} (G_{21}^*(\omega)G_{21}(\omega) + G_{11}^*(\omega)G_{11}(\omega))$$
$$\cdot \left(R_{y_{11}}(\omega) - \frac{|R_{y_{12}}(\omega)|^2}{R_{y_{22}}(\omega)} \right) d\omega \tag{56}$$

and, finally, using the biorthogonal relations, we can express the distortion as a function of the analysis filters

$$D_2 = \frac{1}{4\pi} \int_{-\pi}^{\pi} (H_{21}^*(\omega)H_{21}(\omega) + H_{22}^*(\omega)H_{22}(\omega))$$
$$\cdot \left(\frac{\det[R_x(\omega)]}{R_{y_{22}}(\omega)} \right) d\omega. \tag{57}$$

Likewise, we can obtain an expression for the distortion D_1 associated with the loss of $y_2[n]$. Since the two channels have the

same erasure probability, the expected distortion due to erasure (side distortion) is

$$D = \frac{1}{2}(D_1 + D_2). \tag{58}$$

Note that in our formulation we have only considered the distortion due to erasure and have neglected the one due to quantization, since at high rates, it is much smaller.

Our target is to find a perfect reconstruction filter bank which minimizes the side distortion D. The perfect reconstruction condition is realized by the constraint $\det[H(\omega)] = 1$.[3] The design of the filter bank is also constrained by the redundancy through (48) and (51). Thus, our optimization problem is to find a perfect reconstruction filter bank which minimizes the side distortion D for a given, fixed redundancy ρ.

B. Optimal Solution

As a first step, we decompose the matrix $H(\omega)$ into the product of two matrices $M(\omega)$ and $T(\omega)$

$$H(\omega) = T(\omega)M(\omega). \tag{59}$$

$M(\omega)$ is a unitary decorrelating matrix that diagonalizes the input covariance matrix $R_x(\omega)$. Thus,

$$R_x(\omega) = M(\omega)\Lambda(\omega)M^*(\omega)$$

where $\Lambda(\omega)$ is a diagonal matrix which contains the spectral eigenvalues of $R_x(\omega)$

$$\Lambda(\omega) = \begin{bmatrix} \lambda_1^2(\omega) & 0 \\ 0 & \lambda_2^2(\omega) \end{bmatrix}. \tag{60}$$

For a stationary input process, the decorrelating matrix can be found analytically and has the following form [25]:

$$M(\omega) = \frac{\sqrt{2}}{2} \begin{bmatrix} e^{j\omega/2} & 1 \\ -1 & e^{-j\omega/2} \end{bmatrix} \tag{61}$$

the filter bank related to $M(\omega)$ is usually called principal component filter bank [25]. Now, this factorization does not reduce the generality of the solution, since $M(\omega)$ is a unitary invertible matrix independent of ρ and we are considering square error distortions. So it is enough to solve the simpler problem of optimally designing the matrix $T(\omega)$ for the two input sequences with p.s.d. matrix $\Lambda(\omega)$. Then, the final solution will be represented by the product between this matrix and the decorrelating matrix $M(\omega)$. From now on, we will always assume that the two sequences $(x_1[n], x_2[n])$ have already been decorrelated and are represented by the diagonal p.s.d. matrix $\Lambda(\omega)$. Notice

---

[3]Strictly speaking, the perfect reconstruction condition is satisfied if and only if $\det[H(\omega)] \neq 0$ on the unit circle. However, a factorization into lifting steps is possible only if $\det[H(\omega)]$ is a monomial [8]. Since the side distortion (57) does not depend on the value of the determinant, we can assume, without loss of generality, $\det[H(\omega)] = 1$.

that these two sequences are still a realization of a stationary Gaussian process.

To develop our formulation we need to briefly review the results presented in [13], [14]. Here, Goyal *et al.* focus on the problem of designing an optimal block transform to transmit two Gaussian decorrelated variables over two independent erasure channels.[4] In the case the two channels have the same erasure probability and the two components are coded at the same rate $(R_1 = R_2)$, they show that the optimal MD transform, also called correlating transform, is

$$T = \begin{bmatrix} a & \frac{1}{2a} \\ -a & \frac{1}{2a} \end{bmatrix} \tag{62}$$

where the value of a depends on the redundancy ρ

$$a = \sqrt{\frac{\sigma_2}{2\sigma_1 \left(2^{2\rho} - \sqrt{2^{4\rho} - 1}\right)}} \tag{63}$$

σ_1^2 and σ_2^2 are the variances of the two Gaussian components, with the usual assumption that $\sigma_1^2 > \sigma_2^2$. Finally, the side distortion is given by[5]

$$D = \frac{\sigma_1^2}{2} - \frac{1}{4 \cdot 2^{2\rho} \left(2^{2\rho} - \sqrt{2^{4\rho} - 1}\right)} \left(\sigma_1^2 - \sigma_2^2\right). \tag{64}$$

We can now state the following theorem:

Theorem 2: Assume that $\rho \gg 0$ and that the two p.s.d. $\lambda_1^2(\omega)$, $\lambda_2^2(\omega)$ of the two decorrelated input sequences $x_1[n]$, $x_2[n]$ are such that $\delta_1 > \Delta_2$, where δ_1 is the essential infimum of $\lambda_1^2(\omega)$ and Δ_2 is the essential supremum of $\lambda_2^2(\omega)$. Then the optimal analysis filters for MD coding of $x_1[n]$ and $x_2[n]$ are represented by the following polyphase matrix:

$$T(\omega) = \begin{bmatrix} a(\omega) & \frac{1}{2a(\omega)} \\ -a(\omega) & \frac{1}{2a(\omega)} \end{bmatrix}$$

where

$$a(\omega) = \sqrt{\frac{\lambda_2(\omega)}{2\lambda_1(\omega) \left(2^{2\rho(w)} - \sqrt{2^{4\rho(\omega)} - 1}\right)}}$$

and

$$\rho(\omega) = \rho + \frac{1}{4} \log(\lambda_1^2(\omega) - \lambda_2^2(\omega))$$
$$- \frac{1}{8\pi} \int_{-\pi}^{\pi} \log(\lambda_1^2(\omega) - \lambda_2^2(\omega)) \, d\omega.$$

Proof: Consider only N consecutive elements of the first channel sequence $x_1[n]$, and N consecutive elements of the second channel sequence $x_2[n]$ which are located at the same temporal interval. Call ϕ_{1N} and ϕ_{2N} the two $N \times N$ corresponding correlation matrices. Apply a KLT to each of the two N-sequences to get independent components and name Y_{1N} and Y_{2N} the two N-sequences after the transformation. Call λ_{1i}^2, $i = 1, \ldots, N$, the variances related to the N-sequence Y_{1N} and

λ_{2i}^2, $i = 1, \ldots, N$, the variances related to the second N-sequence (Y_{2N}). Since ϕ_{1N} and ϕ_{2N} are Hermitian–Toeplitz matrices the result is (see the Appendix) that

$$\delta_1 \leq \lambda_{1i}^2 \leq \Delta_1, \qquad \forall i \tag{65}$$

and that

$$\delta_2 \leq \lambda_{2i}^2 \leq \Delta_2, \qquad \forall i \tag{66}$$

where Δ_j and δ_j, $j = 1, 2$, are the essential suprema and the essential infima of the power spectral densities $\lambda_1^2(\omega)$ and $\lambda_2^2(\omega)$. Equations (65) and (66) imply that $\lambda_{1i}^2 > \lambda_{2i}^2$, $i = 1, \ldots, N$, since we assumed $\delta_1 > \Delta_2$.

Now, consider the generic ith couple of element (y_{1i}, y_{2i}). We can apply the results of [13] to this pair and say that if we are allowed to use a redundancy ρ_i then the optimal correlating transform for that pair is

$$T_i = \begin{bmatrix} a_i & \frac{1}{2a_i} \\ -a_i & \frac{1}{2a_i} \end{bmatrix} \tag{67}$$

where a_i is given by

$$a_i = \sqrt{\frac{\lambda_{2i}}{2\lambda_{1i} \left(2^{2\rho_i} - \sqrt{2^{4\rho_i} - 1}\right)}} \tag{68}$$

and that the side distortion is

$$D_i = \frac{\lambda_{1i}^2}{2} - \frac{1}{4 \cdot 2^{2\rho_i} \left(2^{2\rho_i} - \sqrt{2^{4\rho_i} - 1}\right)} (\lambda_{1i}^2 - \lambda_{2i}^2). \tag{69}$$

However, we want to minimize the global side distortion

$$D = \frac{1}{N} \sum_i D_i \tag{70}$$

given a global redundancy budget

$$\rho = \frac{1}{N} \sum_i \rho_i. \tag{71}$$

This is a typical problem of constrained minimization, so we define a new cost function L which combines the distortion and the redundancy through a positive Lagrange multiplier ν

$$L = D + \nu\rho$$
$$L_i = D_i + \nu\rho_i, \qquad i = 1, 2, \ldots, N. \tag{72}$$

Finding a minimum of L amounts to finding minima for each L_i (because the costs are additive). Writing distortion as a function of the redundancy $D_i(\rho_i)$ and taking the derivative we get

$$\frac{\partial L_i}{\partial \rho_i} = \frac{\partial D_i}{\partial \rho_i} + \nu = 0. \tag{73}$$

Thus, for a solution to be optimal, the set of chosen redundancy ρ_i has to correspond to constant-slope points on their respective distortion–redundancy curves. Uniqueness follows from the convexity of these curves and from the use of the Kuhn–Tucker conditions when necessary [5]. A constant-slope solution is obtained for any fixed value of ρ. To enforce the constraint (71) exactly, one has to search over all the values of ν until the budget is met. However, if we suppose that ρ_i is sufficiently large then

---

[4]Actually, they consider also the case of larger vectors. For the two-channel case their work is an extension of the results presented in [22], [33].

[5]It is interesting to notice that if the Gaussian source has a circularly symmetric probability density, i.e., $\sigma_1 = \sigma_2$, then the distortion is independent of ρ. In this case, the side distortion cannot be reduced with the addition of redundancy, so the approach based on correlating transforms is useless.

it is possible to give a closed form for the allocation problem. In fact, it follows that

$$\frac{\partial D_i}{\partial \rho_i} = -\frac{\ln 2(\lambda_{1i}^2 - \lambda_{2i}^2)}{4 \cdot 2^{2\rho_i} \left(\sqrt{2^{4\rho_i} - 1}\right)}$$

$$\approx -\frac{\ln 2}{4}(\lambda_{1i}^2 - \lambda_{2i}^2)2^{-4\rho_i} = -\nu. \tag{74}$$

The constant-slope solution forces the redundancies to be of the following form:

$$\rho_i = \alpha + \frac{1}{4}\log(\lambda_{1i}^2 - \lambda_{2i}^2). \tag{75}$$

Using the redundancy constraint (71)

$$\sum_i \rho_i = N\alpha + \frac{1}{4}\sum_i \log(\lambda_{1i}^2 - \lambda_{2i}^2) = N\rho \tag{76}$$

we find

$$\alpha = \rho - \frac{1}{4N}\sum_i \log(\lambda_{1i}^2 - \lambda_{2i}^2) \tag{77}$$

and finally

$$\rho_i = \rho + \frac{1}{4}\log(\lambda_{1i}^2 - \lambda_{2i}^2) - \frac{1}{4N}\sum_i \log(\lambda_{1i}^2 - \lambda_{2i}^2). \tag{78}$$

The approximation in (74) holds if ρ_i is sufficiently large. Its value depends on the total redundancy budget ρ and on the difference $\lambda_{1i}^2 - \lambda_{2i}^2$. The difference $\lambda_{1i}^2 - \lambda_{2i}^2$ influences the slope of the distortion–redundancy curves (74). Now, the global distortion is minimized when the set of chosen redundancy ρ_i corresponds to constant slope points. If $\lambda_{1i}^2 - \lambda_{2i}^2 = 0$, the slope of the ith curve is zero and the optimal solution is always found imposing the Kuhn–Tucker condition: $\rho_i = 0$. For this reason, the approximation in (74) holds only when both the conditions $\rho \gg 0$ and $\delta_1 > \Delta_2$ are verified ($\delta_1 > \Delta_2$ implies $\lambda_{1i}^2 > \lambda_{2i}^2$, $\forall i$). In general, we can say that the difference $\lambda_{1i}^2 - \lambda_{2i}^2$ influences the allocation strategy of the redundancy. The redundancy is mainly allocated in the region where this difference is higher.

Now we can let N go to infinity and find, in this way, the optimal spectral distribution of the redundancy

$$\rho(\omega) = \rho + \frac{1}{4}\log(\lambda_1^2(\omega) - \lambda_2^2(\omega))$$
$$- \frac{1}{2\pi}\int_{-\pi}^{\pi}\frac{1}{4}\log(\lambda_1^2(\omega) - \lambda_2^2(\omega))\,d\omega. \tag{79}$$

Once $\rho(\omega)$ is known, we can obtain the expression of the side distortion D

$$D = \frac{1}{2\pi}\int_{-\pi}^{\pi} D(\omega)\,d\omega \tag{80}$$

where

$$D(\omega) = \frac{\lambda_1^2(\omega)}{2} - \frac{1}{4 \cdot 2^{2\rho(\omega)}\left(2^{2\rho(\omega)} - \sqrt{2^{4\rho(\omega)} - 1}\right)}$$
$$\cdot (\lambda_1^2(\omega) - \lambda_2^2(\omega)) \tag{81}$$

and the expression of the polyphase matrix $T(\omega)$:

$$\begin{bmatrix} a(\omega) & \frac{1}{2a(\omega)} \\ -a(\omega) & \frac{1}{2a(\omega)} \end{bmatrix} \tag{82}$$

where

$$a(\omega) = \sqrt{\frac{\lambda_2(\omega)}{2\lambda_1(\omega)\left(2^{2\rho(\omega)} - \sqrt{2^{4\rho(\omega)} - 1}\right)}}. \tag{83}$$

$\square$

When the approximation (74) is not verified, namely, when at least one of the two hypotheses $\rho \gg 0$ and $\delta_1 > \Delta_2$ is not satisfied, the optimal allocation of the redundancy over frequency can only be found numerically. This means that, for any fixed ν, one has to numerically solve (74) and then has to search over all the values of ν until the constraint (71) is met.

Consider, now, (80) and (81). They express the side distortion in function of the spectral distribution of the redundancy ρ. The side distortion is maximum when we are not allowed to allocate any redundancy over the frequency and its maximum value is

$$D = \frac{1}{8\pi}\int_{-\pi}^{\pi}(\lambda_1^2(\omega) + \lambda_2^2(\omega))\,d\omega. \tag{84}$$

Its minimum value occurs when we can allocate an infinite amount of redundancy over the frequency and it is equal to

$$D = \frac{1}{4\pi}\int_{-\pi}^{\pi}\lambda_2^2(\omega)\,d\omega. \tag{85}$$

This value represents the *systematic error* due to the estimation of one subsequence with the other one and cannot be eliminated even at infinite redundancy. The systematic error typically occurs in MD systems based on correlating transforms [14], [32]. This is in contrast with the performance of other systems (i.e., MDTC), where at high rates both side and central distortions decrease with the rate. Thus, this result gives us a first insight about the performance of the filter bank system.

Corollary 2: The filter bank system is not useful at high rates since, independently of the amount of redundancy allocated, the side distortion has a constant factor (the systematic error) that cannot be eliminated.

C. Approximate Finite Impulse Response (FIR) Solutions

Usually, the filters obtained with the optimization algorithm of the previous section are of infinite length. However, in some applications it is important to approximate them with FIR filters. Let us call $H_f(\omega)$ the polyphase matrix related to the FIR filter bank and D_f the corresponding side distortion obtained with this set of filters. Clearly, $D_f \geq D$, where D is the ideal side distortion given by (58), since the best performance is usually achieved with infinite length filters. Now, the problem is to design a perfect reconstruction FIR filter bank that minimizes the performance gap $D_f - D$ for each fixed redundancy.

We solve this problem numerically by running a constrained minimization algorithm using a gradient descendent approach. The convex function to minimize is $\|D - D_f\|^2$, while the constraints are the perfect reconstruction condition $\det[H_f(\omega)] = 1$ and the maximum allowed redundancy ρ.

Recall that given an FIR analysis filter bank, perfect reconstruction with FIR filters is possible if and only if $\det[H(\omega)]$ is a monomial [31]. So, once we have designed FIR analysis filters with the constraint $\det[H_f(\omega)] = 1$, we know that it is possible

to reconstruct the signal with FIR synthesis filters. These synthesis filters are obtained in the usual way

$$G_0(\omega) = e^{j\omega} H_1(\omega + \pi)$$
$$G_1(\omega) = -e^{j\omega} H_0(\omega + \pi).$$

Finally, recall that once the FIR filter bank is obtained, it can always be factored into a finite number of lifting steps. These steps can be sequentially rounded and, in this way, one can obtain the discrete version of the continuous transform.

D. Application to a Gauss–Markov Process

To conclude this section, we apply our filter design techniques to a Gaussian source and analyze the filter responses.

We consider a Gauss–Markov source $x[n] = \alpha x[n-1] + w[n]$, where the regression coefficient α has magnitude less than 1 and where $w[n]$ is a zero mean, unit variance, independent and identically distributed (i.i.d.) Gaussian source. The p.s.d. of this process is

$$S_x(\omega) = \frac{1}{|1 - \alpha e^{-j\omega}|^2}. \tag{86}$$

Now, the polyphase matrix $H(\omega)$ of the optimal filter bank is given by the product of the matrix $T(\omega)$ with the matrix $M(\omega)$. This second one is known and is given by (61). To design $T(\omega)$ we need to compute the spectral eigenvalues of the input p.s.d. matrix. First notice that the two subsequences obtained by downsampling $x[n]$ are still Gauss–Markov processes, but with the regression coefficient α replaced by α^2 and the i.i.d. original Gaussian source $w[n]$ replaced by a new i.i.d. Gaussian source with zero mean and variance $1 + \alpha^2$. Hence, the power spectral densities for these two processes are given by

$$R_{x11}(\omega) = R_{x22}(\omega) = \frac{1 + \alpha^2}{|1 - \alpha^2 e^{-j\omega}|^2}. \tag{87}$$

The cross p.s.d. $R_{x12}(\omega)$ is given by

$$R_{x12}(\omega) = \frac{\alpha\left(1 + e^{(-j\omega)}\right)}{1 + \alpha^2} R_{x11}(\omega) \tag{88}$$

with $R_{x21}(\omega) = R_{x12}^*(\omega)$. Finally, the p.s.d. matrix after decorrelation is shown in (89) at the bottom of the page. Observe that the two spectral eigenvalues are equal only at π (and, of course, at $-\pi$). As previously stated, at the points closest to the frequency values where $\lambda_1^2(\omega) = \lambda_2^2(\omega)$ it is not possible to use the closed-form (79) even in the high redundancy hypothesis. So, for the Gauss–Markov source, $a(\omega)$ (and, consequently, $T(\omega)$) can only be found numerically.

In Fig. 5, we show the frequency responses of the two analysis filters as a function of the redundancy for the case $\alpha = 0.9$. It is interesting to notice that the amplitude of the two frequency

responses is exactly the same, the two filters differ only for the phase response. This is due to the presence of the principal component filter bank given by $M(\omega)$ and to the constraint $R_1 = R_2$ which forces the matrix $T(\omega)$ to have the shape given by (82). Moreover, notice that at high redundancies the two filters tend to be low-pass. In the case of $\alpha = 0.9$, the Gauss–Markov process is a low-pass process; thus, the frequency responses of the two filters tend to preserve the frequency region where the p.s.d. of the input process is mostly concentrated. This is valid in general, that is, at high redundancies the analysis filters better preserve the region where most of the p.s.d. of the input process is concentrated. It is also of interest to note that, at low redundancies, the two filter responses do not tend to be that of a principal component filter bank, that is an ideal low-pass and an ideal high-pass filter. This is because, if quantization is performed before the transform, the principal component filter bank does not represent the only solution that gives minimum coding rates. The same phenomenon happens in the block transform case, where the KLT does not represent the only transform that gives minimum rates if quantization is performed before the transform [12], [13]. Thus, this additional degree of freedom makes it possible to have a filter bank (or a block transform [13]) that achieves, at the same time, minimum coding rates and balanced rates. This is the solution that we have at low redundancies.

Finally, in Fig. 6 we show the frequency response of the two FIR analysis filters obtained with the minimization algorithm presented in Section III-C. The filters are all of length 6. It is interesting to compare these frequency responses with the ones in Fig. 5. In the FIR case, the amplitude responses of the two filters are not equal, but they tend to be close to each others at high redundancies. Moreover, in the high-redundancy region, the two frequency responses tend to be low-pass as in the ideal case.

IV. PERFORMANCE ANALYSIS

In order to assess the performance of the filter bank proposed in the previous section, we compare it with the asymptotic ideal bounds found in Section II and with other two systems: the MD transform coder [2] and the MD-DPCM system [18]. In the next subsection, we briefly review these two systems. In the simulations, we consider two different Gaussian input sources: a classical Gauss–Markov source and a low-pass Gaussian source obtained as illustrated in Fig. 7, where $x_1[n]$ and $x_2[n]$ are two i.i.d. Gaussian sources with variances σ_1^2, σ_2^2, and $G(\omega)$ is an ideal low-pass filter. Moreover, we consider two different scenarios: high rate, infinite delay/complexity and low rate, finite delay/complexity. In the first scenario, the analysis and the results presented in the first part of this paper are valid. The second more realistic simulation is important, because we do not have

$$\Lambda(\omega) = \begin{bmatrix} R_{x11}(\omega)\left(1 + \dfrac{2\alpha\cos(\omega/2)}{(1 + \alpha^2)}\right) & 0 \\ 0 & R_{x11}(\omega)\left(1 - \dfrac{2\alpha\cos(\omega/2)}{(1 + \alpha^2)}\right) \end{bmatrix}. \tag{89}$$
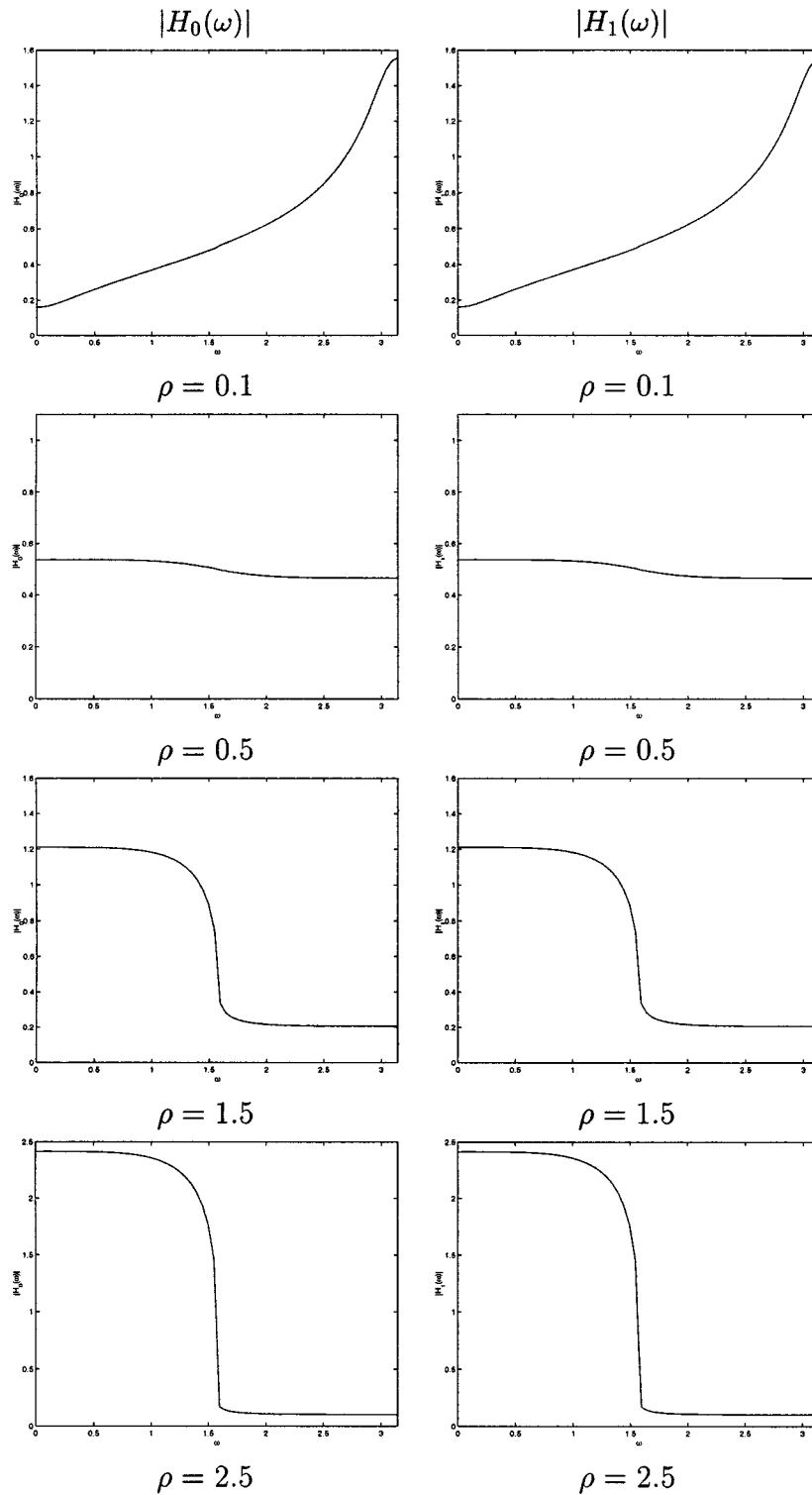
Fig. 5.   Frequency response of the analysis filters in function of the redundancy ρ.

clear theoretical answers on the behavior of the considered systems in this particular context.

A. Other MD Coding Systems

The MDTC [2] is illustrated in Fig. 8. It is represented by two main elements: a linear transform which turns out to be a KLT and a set of entropy-constrained MDSQ. Recall that an MDSQ produces a pair of indexes for each input scalar sample. The behavior of an MDSQ is characterized by two elements: the rate at which it operates and the strategy in the assignment of the two output indexes. This second element defines the tradeoff between side and central distortion. That is, it defines if the indexes are assigned in a way to mainly minimize the central or the

$$|H_0(\omega)| \qquad\qquad |H_1(\omega)|$$

$$\rho = 0.1 \qquad\qquad \rho = 0.1$$

$$\rho = 1.5 \qquad\qquad \rho = 1.5$$

$$\rho = 1.8 \qquad\qquad \rho = 1.8$$

$$\rho = 2.5 \qquad\qquad \rho = 2.5$$
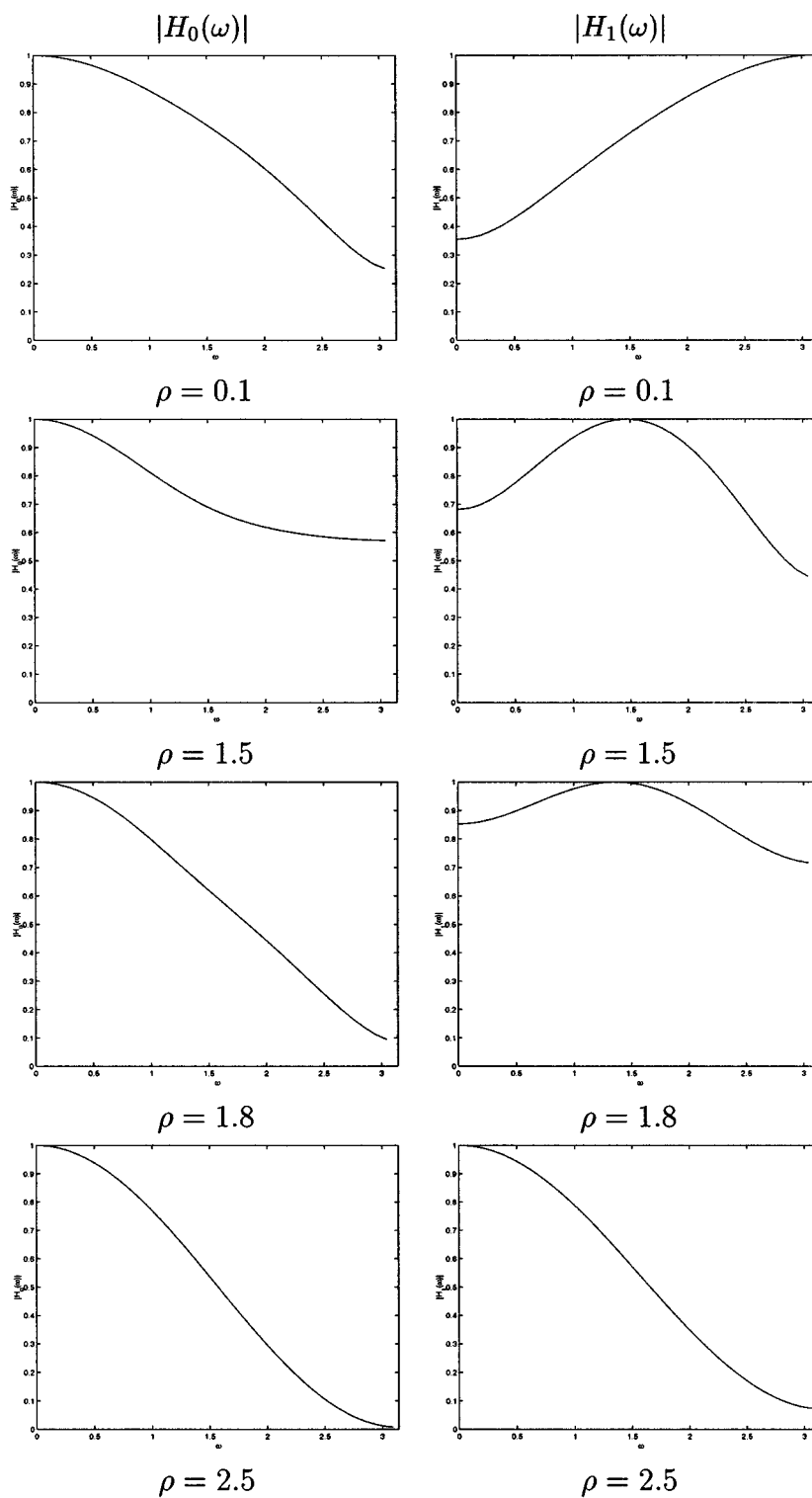
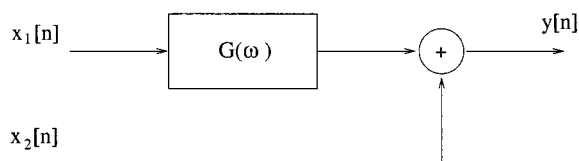Fig. 6. Frequency response of the analysis FIR filters in function of the redundancy ρ.

Fig. 7. A low-pass Gaussian source. $G(\omega)$ is an ideal low-pass filter and $x_1[n]$, $x_2[n]$ are two i.i.d. Gaussian sources.

side distortion. The system works in the following way: it takes a block of N consecutive elements of the input sequence $x[n]$ and applies a KLT to them. Then each of the decorrelated component is encoded with a different entropy-constrained MDSQ and the pair of indexes produced by the MDSQ are transmitted over two separate channels. The result is that in case of Gaussian input sources and at high rates optimal performance is achieved if the index assignment strategy is the same for each MDSQ
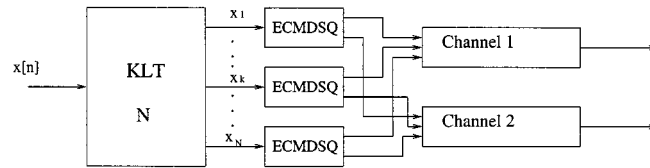
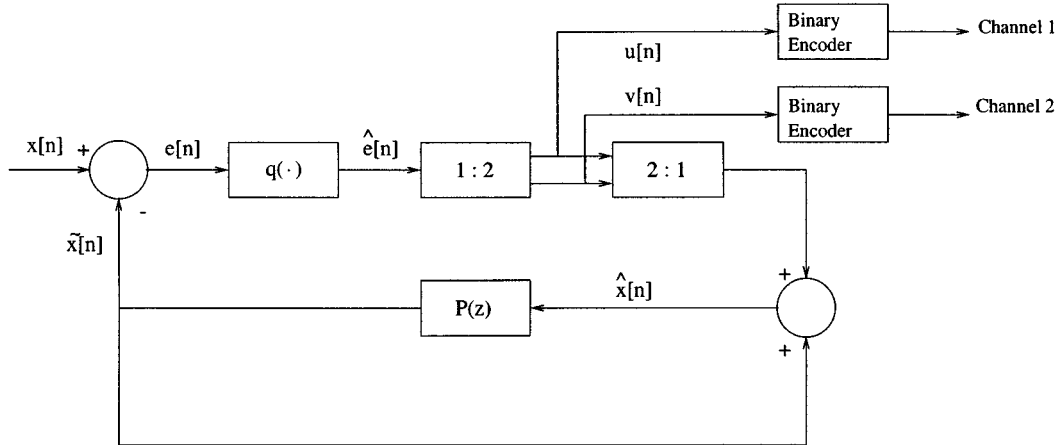Fig. 8. The multiple description transform coder.

Fig. 9. The MD-DPCM system.

and bits are allocated to each component according to a single description allocation strategy [2]. Finally, letting N to go to infinity and in the case of high rates, the performance of this system is given by [2]

$$D_0 D_1 = \gamma 2^{-4R} \exp\left(\frac{1}{\pi} \int_{-\pi}^{\pi} \ln S(\omega) \, d\omega\right) \quad (90)$$

where D_0 is the central distortion, D_1 is the side distortion, R is the average rate per sample per channel, $S(\omega)$ is the input power spectral density and $\gamma = \left(\frac{\pi e}{12}\right)^2$.

The analysis part of the MD-DPCM system [18] is illustrated in Fig. 9.

$$P(z) = b_1 z^{-1} + b_2 z^{-2}$$

is a second-order predictor filter. The quantized predicted sequence $\hat{e}[n]$ is separated into two subsequences containing the even and the odd samples and these subsequences are sent over two different channels. If one subsequence is lost, it is linearly estimated using the received one. Now, if the input source is Gauss–Markov with regression coefficient α, it turns out that the estimating filters present in the synthesis part of the system are realizable filters. Moreover, in this case, the side distortion is given by [18]

$$D_1 = \frac{1}{(1 - \alpha^2)} \left(1 - \frac{(b_2^2 + 2b_2\alpha^2 + \alpha^2)}{2(\alpha^2 + b_2^2)}\right) \quad (91)$$

while the central distortion is [18]:

$$D_0 = (1 + b_2^2/\alpha^2)\epsilon^2 2^{-2R} \quad (92)$$

where ϵ^2 depends on the kind of quantizer used and b_1, b_2 are related by the following equation:

$$b_1/\alpha + b_2/\alpha^2 = 1, \qquad 0 \le b_2 \le \alpha^2.$$

The interesting element to note is that, as for our system, the side distortion of the MD-DPCM system does not go to zero even at infinite rate.

B. High-Rate, Infinite-Complexity Performance

We consider a first-order Gauss–Markov source. In the high-rate and infinite-complexity hypothesis, the performance of the MDTC is given by (90) where $S(\omega)$ is given by (86). The side and central distortions of the MD-DPCM are given by (91) and (92). For the filter bank case, the filter responses are obtained numerically as shown in Section III-D. Given the filter responses, the side distortion at high rates is given by (57) and (58). The central distortion is obtained by numerically inverting the equations in (49).

In Fig. 10, we compare the four performances: MDTC, MD-DPCM, MD filter bank, and ideal bounds, for the case of $\alpha = 0.9$ and $R = 6$ bit/sample/channel. As we can see, the MDTC outperforms the other two systems. This is not astonishing since in the MDTC both the central and the side distortions decrease exponentially with the rate R. The side distortions of the MD-DPCM system and of our system suffer of the systematic estimation error that becomes dominant at high rates and that does not reduce with the rate. It is also interesting to note that the gap between the ideal bounds and the MDTC is constant and equal to 3.06 dB. This confirms that this system attains asymptotically optimal performance.
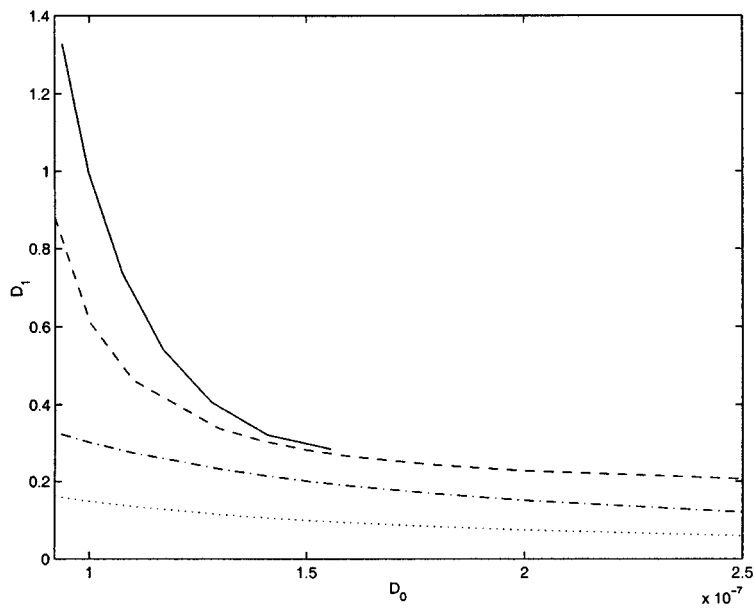
Fig. 10. Asymptotic performance for a Gauss–Markov input source. Abscissa: central distortion, Ordinate: side distortion. Dotted: ideal bounds, dashed–dotted: MDTC, dashed: filter banks for MD coding, line: MD-DPCM.
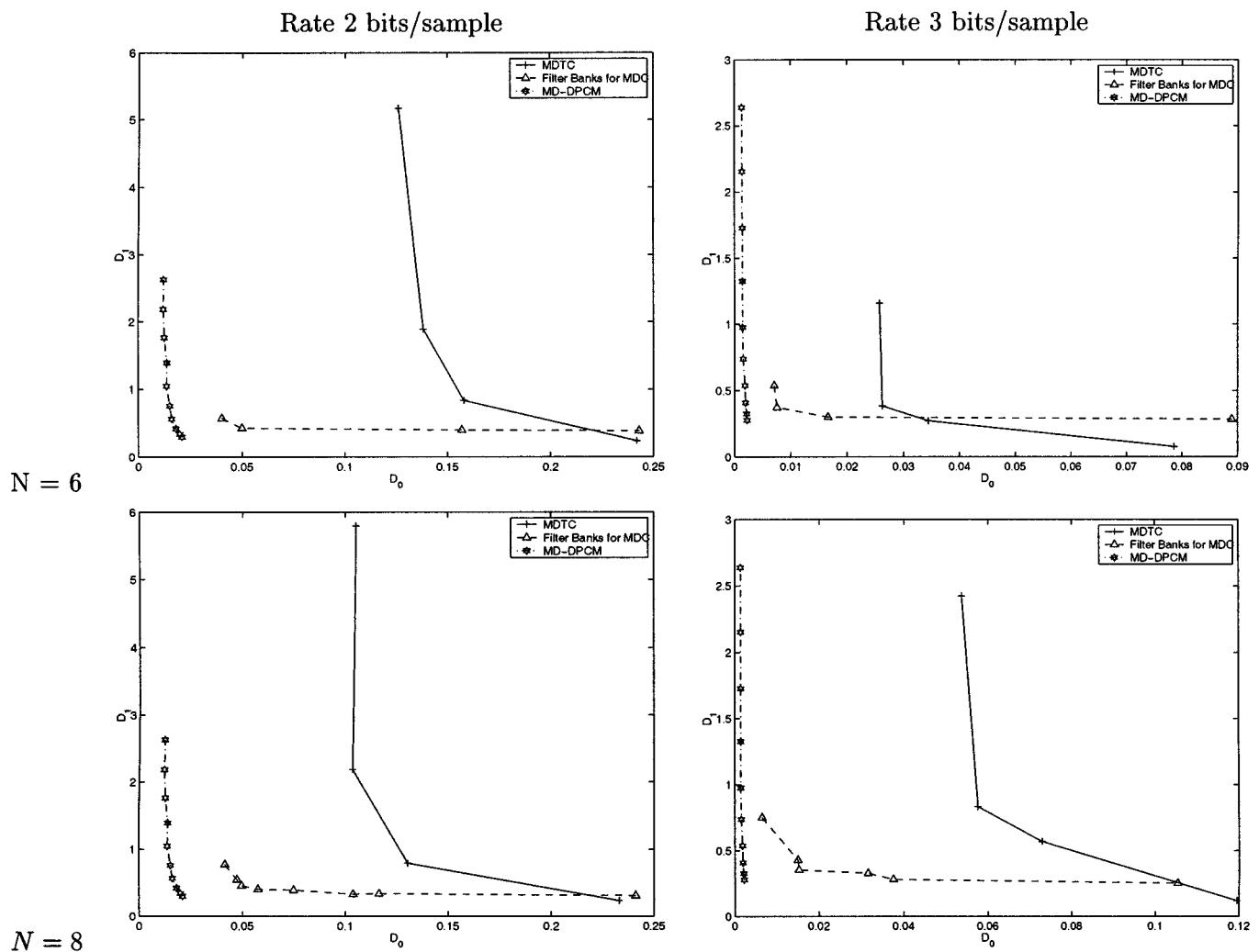

Fig. 11. Comparison between MDTC, filter banks for MD coding, and MD-DPCM system. Input source: Gauss–Markov. Line: MDTC, dashed: filter banks for MD coding, dashed–dotted: MD-DPCM.
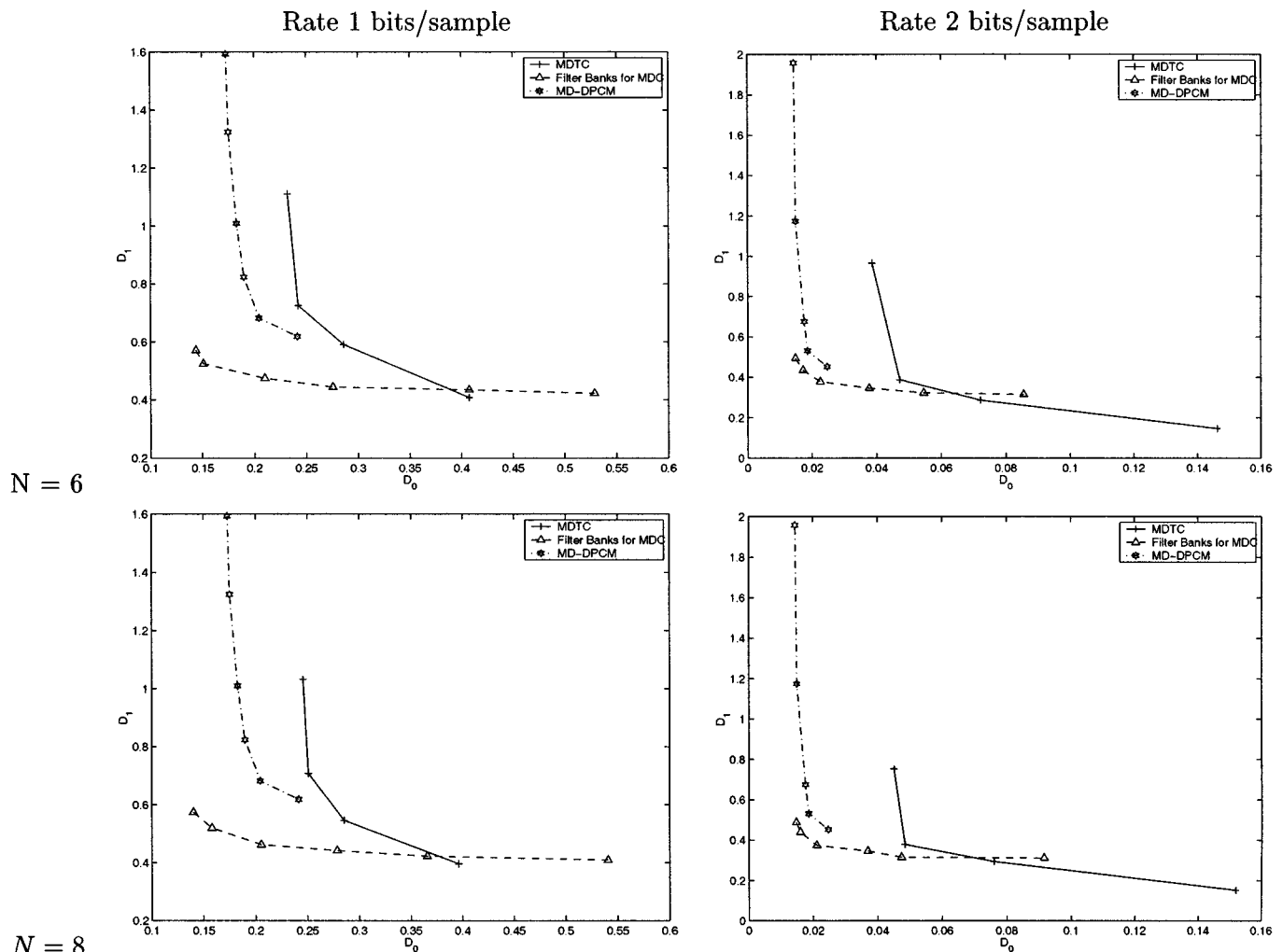
Fig. 12.   Comparison between MDTC, filter banks for MD coding, and MD-DPCM system. Input source: low-pass Gaussian source. Line: MDTC, dashed: filter banks for MD coding, dashed–dotted: MD-DPCM.

C. Low-Rate, Finite Delay/Complexity Performance

In practical settings, we are more interested in low-rate be-haviors and we have to deal with finite delay/complexity con-straints. That means that either the KLT or the filters in the filter bank have finite length N. The FIR filters are designed using the numerical optimization presented in Section III-C.[6] The MDTC is the same shown previously except that the KLT operates on blocks of finite length N. Bits are still allocated according to a single description allocation strategy and the MDSQs are de-signed such that the index assignment strategy is the same for each of the N components. The MD-DPCM system is made of realizable filters and does not need to be approximated.

In the first simulation, we consider again a first-order Gauss–Markov process with memory $\alpha = 0.9$. Numerical results are shown in Fig. 11. Here, we consider two bit rates $R = 2$ and $R = 3$ bit/sample/channel and two different length constraints: $N = 6$ and $N = 8$. The graphs show the tradeoff between side and central distortion for the three systems. The first interesting thing to note is that, in the low rate regime, the MDTC, which is optimal at high rates, is generally out-

[6]For simplicity, we did not decompose the filters into lifting steps. Doing the decomposition would slightly improve the filter bank scheme.

performed by the other two systems. The MD-DPCM system is the best system in this context. Moreover, comparing the results of Figs. 10 and 11, one can conclude that, in this case, our system can attain the same performance of the MD-DPCM system only at the price of infinite delay/complexity (i.e., with infinite length filters).

It seems that one of the reasons why the MD-DPCM system is superior to the other two is because it has been designed as-suming that the input source is Gauss–Markov and, thus, it well exploits the particular structure of this source. The other two systems do not take particular advantage of the characteristics of the input source. For this reason, it is of interest to run sim-ilar simulations with a different Gaussian source. Numerical re-sults are shown in Fig. 12. In this case, the input source is a low-pass Gaussian source obtained as illustrated in Fig. 7, where $x_1[n]$ and $x_2[n]$ are two i.i.d. Gaussian sources with variances $\sigma_2^2 = 1.5$ and $\sigma_2^2 = 0.5$, respectively. We consider two bit-rates $R = 1$ and $R = 2$ bit/sample/channel and two different length constraints: $N = 6$ and $N = 8$. We can see that, in this con-text, our system is the best system in the medium–low redun-dancy region. It is also of interest to note that the performance gap between our system and the MDTC reduces with the rate; in particular, we have noticed that for rates greater than $R = 3$
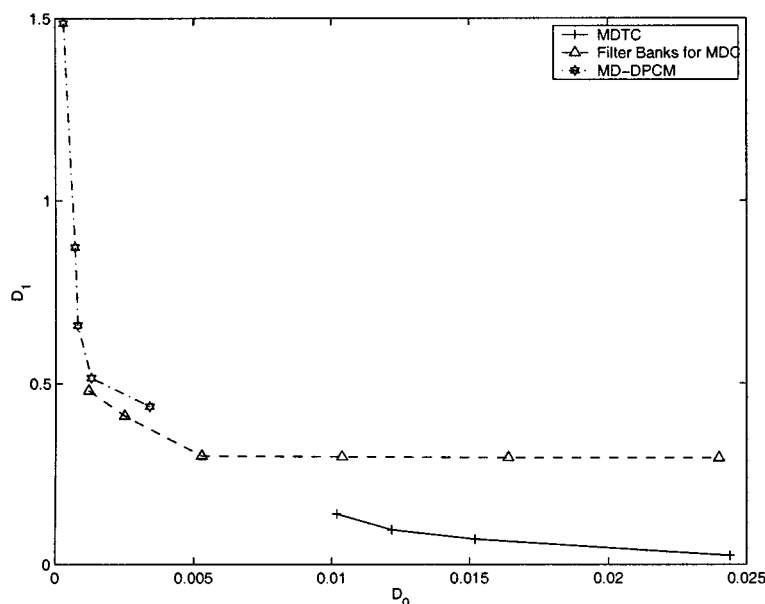
Fig. 13. Comparison between MDTC, filter banks for MD coding, and MD-DPCM system in the case of $R = 4$ bits/sample/channel. Input source: low-pass Gaussian source. Line: MDTC, dashed: filter banks for MD coding, dashed–dotted: MD-DPCM.

bit/sample/channel the MDTC performs better also at low redundancies (see Fig. 13). Finally, the performance of these two systems slightly increases with the length N but, the length, does not changes the performance gap between them.

In conclusion, this set of experiments indicates that, in the low bit-rate regime and at low and medium redundancies, MD-DPCM and filter bank system perform better than the MDTC. Moreover, in this regime and for some classes of Gaussian sources, our system outperforms the other two.

V. CONCLUSION

In this work, we have addressed the problem of MD coding in a subband framework. We have shown how to design perfect reconstruction filter banks that can minimize the side distortion given a certain amount of redundancy. Two other important contributions of this paper are: a) the characterization of a region which is asymptotically the MD rate-distortion region for general stationary Gaussian sources, and b) experimental results showing conditions under which our method outperforms other MD methods.

APPENDIX
TOEPLITZ DISTRIBUTION THEOREM

For a proof of this theorem refer to [16].

Theorem 3 (Toeplitz Distribution Theorem): Let Φ_∞ be an infinite Toeplitz matrix with entry ϕ_k on the kth diagonal. The eigenvalues of Φ_∞ are contained in the interval $\delta \leq \lambda \leq \Delta$ where δ and Δ are the essential infimum and supremum,[7] respectively, of the function

$$\Phi(\omega) = \sum_{k=-\infty}^{\infty} \phi_k e^{-jk\omega}.$$

[7]The essential supremum of a function $f(x)$ is the $\inf_E \sup_{x \notin E} f(x)$ where E ranges over all sets of Lebesgue measure zero. Likewise, the essential infimum is $\sup_E \inf_{x \notin E} f(x)$.

Moreover, if both δ and Δ are finite and $G(\lambda)$ is any continuous function of $\lambda \in [\delta, \Delta]$, then

$$\lim_{n \to \infty} \frac{1}{n} \sum_{k=1}^{n} G\left(\lambda_k^{(n)}\right) = \frac{1}{2\pi} \int_{-\pi}^{\pi} G[\Phi(\omega)] \, d\omega$$

where the $\lambda_k^{(n)}$ are the eigenvalues of the nth-order matrix Φ_n centered about the main diagonal of Φ_∞.

REFERENCES

[1] R. Ahlswede, "The rate distortion region for multiple descriptions without excess rate," *IEEE Trans. Inform. Theory*, vol. IT-31, pp. 721–726, Nov. 1985.
[2] J.-C. Batllo and V. A. Vaishampayan, "Asymptotic performance of multiple description transform codes," *IEEE Trans. Inform. Theory*, vol. 43, pp. 703–707, Mar. 1997.
[3] T. Berger, *Rate Distortion Theory*. Englewood Cliffs, NJ: Prentice-Hall, 1971.
[4] R. C. Calderbank, I. Daubechies, W. Sweldens, and B.-L. Yeo, "Wavelet transforms that map integers to integers," *Appl. Comput. Harmonic Anal.*, vol. 5, no. 3, pp. 332–369, 1998.
[5] E. K. P. Chong and S. H. Zak, *Introduction to Optimization*. New York: Wiley, 1996.
[6] P. A. Chou, S. Mehrotra, and A. Wang, "Multiple description decoding of overcomplete expansions using projection onto convex sets," in *Proc, Data Compression Conf.*, Snowbird, UT, Mar. 1999.
[7] T. Cover and J. A. Thomas, *Elements of Information Theory*. New York: Wiley, 1991.
[8] I. Daubechies and W. Sweldens, "Factoring wavelet transforms into lifting steps," *J. Fourier Anal. Appl.*, vol. 4, no. 3, pp. 247–269, 1998.
[9] S. N. Diggavi, N. J. A. Sloane, and V. A. Vaishampayan, "Design of asymmetric multiple description lattice vector quantizers," in *Proc. Data Compression Conf.*, Snowbird, UT, Mar. 2000.

[10] A. A. El Gamal and T. M. Cover, "Achievable rates for multiple descriptions," *IEEE Trans. Inform. Theory*, vol. IT-28, pp. 851–857, Nov. 1982.

[11] M. Fleming and M. Effros, "Generalized multiple description vector quantization," in *Proc. Data Compression Conf.*, Snowbird, UT, Mar. 1999.

[12] V. K Goyal, "Transform coding with integer-to-integer transforms," *IEEE Trans. Inform. Theory*, vol. 46, pp. 465–473, Mar. 2000.

[13] V. Goyal and J. Kovačević, "Optimal multiple description transform coding of Gaussian vectors," in *Proc. Data Compression Conf.*, Snowbird, UT, Mar. 1998.

[14] ——, "Generalized multiple description coding with correlating transforms," *IEEE Trans. Inform. Theory*, vol. 47, pp. 2199–2224, Sept. 2001.

[15] V. K Goyal, J. Kovačević, and J. A. Kelner, "Quantized frame expansions with erasures," *Appl. Comput. Harmonic Anal.*, vol. 10, no. 3, pp. 203–233, May 2001.

[16] U. Grenander and G. Szego, *Toeplitz Forms and Their Applications*. Berkeley and Los Angeles, CA: Univ. California Press, 1958.

[17] J. Hong, "Discrete Fourier, Hartley, and cosine transforms in signal processing," Ph.D. dissertation, Columbia Univ., New York, 1993.

[18] A. Ingle and V. A. Vaishampayan, "DPCM system design for diversity systems with applications to packetized speech," *IEEE Trans. Speech Audio Processing*, vol. 1, pp. 48–58, Jan. 1995.

[19] J. A. Kelner, V. K Goyal, and J. Kovačević, "Multiple description lattice vector quantization: Variations and extensions," in *Proc. Data Compression Conf.*, Snowbird, UT, Mar. 2000.

[20] T. Linder, R. Zamir, and K. Zeger, "The multiple description rate region for high resolution source coding," in *Proc. Data Compression Conf.*, Snowbird, UT, Mar. 1998.

[21] S. Mehrotra and P. A. Chou, "On optimal frame expansions for multiple description quantization," in *Proc. Int. Symp. Information Theory (ISIT)*, Sorrento, Italy, June 2000.

[22] M. Orchard, Y. Wang, V. A. Vaishampayan, and A. R. Reibman, "Redundancy rate-distortion analysis of multiple description coding using pairwise correlation transforms," in *Proc. IEEE Int. Conf. Image Processing*, Santa Barbara, CA, Oct. 1997.

[23] L. Ozarow, "On a source-coding problem with two channels and three receivers," *Bell Syst. Tech. J.*, vol. 59, no. 10, pp. 1909–1921, 1980.

[24] S. D. Servetto, K. Ramchandran, V. A. Vaishampayan, and K. Nahrstedt, "Multiple description wavelet based image coding," *IEEE Trans. Image Processing*, vol. 9, pp. 813–826, May 2000.

[25] M. K. Tsatsanis and G. B. Giannakis, "Principal component filter banks for optimal multiresolution analysis," *IEEE Trans. Signal Processing*, vol. 43, pp. 1766–1777, Aug. 1995.

[26] P. P. Vaidyanathan, *Multirate Systems and Filter Banks*. Englewood Cliffs, NJ: Prentice-Hall, 1993.

[27] V. A. Vaishampayan, "Design of multiple description scalar quantizers," *IEEE Trans. Inform. Theory*, vol. 39, pp. 821–834, May 1993.

[28] V. A. Vaishampayan and J.-C. Batllo, "Asymptotic analysis of multiple description quantizers," *IEEE Trans. Inform. Theory*, vol. 44, pp. 278–284, Jan. 1998.

[29] V. A. Vaishampayan and J. Domaszewicz, "Design of entropy-constrained multiple-description scalar quantizers," *IEEE Trans. Inform. Theory*, vol. 40, pp. 245–250, Jan. 1994.

[30] V. A. Vaishampayan, N. J. A. Sloane, and S. D. Servetto, "Multiple description vector quantization with lattice codebooks: Design and analysis," *IEEE Trans. Inform. Theory*, vol. 47, pp. 1718–1734, July 2001.

[31] M. Vetterli and J. Kovačević, *Wavelets and Subband Coding*. Englewood Cliffs, NJ: Prentice-Hall, 1995.

[32] Y. Wang, M. Orchard, V. A. Vaishampayan, and A. R. Reibman, "Multiple description coding using pairwise correlation transforms," *IEEE Trans. Image Processing*, vol. 10, pp. 351–366, Mar. 2001.

[33] Y. Wang, T. Orchard, and A. R. Reibman, "Multiple description image coding for noisy channels by pairing transform coefficients," in *Proc. IEEE 1997 1st Workshop Multimedia Signal Processing*, Princeton, NJ, June 1997.

[34] X. Yang and K. Ramchandran, "Optimal subband filter banks for multiple description coding," *IEEE Trans. Inform. Theory*, vol. 46, pp. 2477–2490, Nov. 2000.

[35] R. Zamir, "Gaussian codes and Shannon bounds for multiple descriptions," *IEEE Trans. Inform. Theory*, vol. 45, pp. 2629–2635, Nov. 1999.

[36] ——, "Shannon-type bounds for multiple descriptions of a stationary source," in *J. Comb., Inform. Syst. Sci.*, Dec. 2000.

[37] A. Zandi, J. D. Allen, E. L. Schwartz, and M. Boliek, "CREW: Compression with reversible embedded wavelets," in *Proc. Data Compression Conf.*, Snowbird, UT, Mar. 1995.

[38] Z. Zhang and T. Berger, "New results in binary multiple descriptions," *IEEE Trans. Inform. Theory*, vol. IT-33, pp. 502–521, July 1987.

[39] ——, "Multiple description source coding with no excess marginal rate," *IEEE Trans. Inform. Theory*, vol. 41, pp. 349–357, Mar. 1995.