Multiple Description Coding Using Pairwise Correlating Transforms

Yao Wang, *Senior Member, IEEE*, Michael T. Orchard, *Fellow, IEEE*, Vinay Vaishampayan, *Member, IEEE*, and Amy R. Reibman, *Member, IEEE*

***Abstract*—The objective of multiple description coding (MDC) is to encode a source into multiple bitstreams supporting multiple quality levels of decoding. In this paper, we only consider the two-description case, where the requirement is that a high-quality reconstruction should be decodable from the two bitstreams together, while lower, but still acceptable, quality reconstructions should be decodable from either of the two individual bitstreams. This paper describes techniques for meeting MDC objectives in the framework of standard transform-based image coding through the design of pairwise correlating transforms. The correlation introduced by the transform helps to reduce the distortion when only a single description is received, but it also increases the bit rate beyond that prescribed by the rate-distortion function of the source. We analyze the relation between the redundancy (i.e., the extra bit rate) and the single description distortion using this transform-based framework. We also describe an image coder that incorporates the pairwise transform and show its redundancy-rate-distortion performance for real images.**

***Index Terms*—Image coding, image communication, multiple description coding, redundancy, source coding, transform coding.**

I. Introduction

DIVERSITY is commonly used to enhance the reliability of communication systems. By sending information over multiple channels with independent failure events, the probability of receiving some information from at least one of the channels is greatly increased. *Multiple description coding* (MDC) addresses the problem of coding a source for transmission over a communication system with diversity. This paper focuses on the special case of MDC for a nonwhite Gaussian source to be transmitted over a system with two-channel diversity. In such a system, the MDC objective is to encode the source into two bitstreams such that, if both channels are received, a high-quality reconstruction can be decoded while, if either channel is lost, a lower-quality, but acceptable, reconstruction can be decoded from the remaining channel. It is assumed that the decoder knows when a channel is lost, thus allowing it to use different algorithms for computing the two types of reconstruction.

Manuscript received December 18, 1998; revised October 16, 2000. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Lina J. Karam.

Y. Wang is with Polytechnic University, Brooklyn, NY 11201 USA (e-mail: yao@vision.poly.edu).

M. T. Orchard is with Princeton University, Princeton, NJ 08540 USA (e-mail: orchard@ee.princeton.edu).

V. Vaishampayan and A. R. Reibman are with AT&T Labs—Research, Florham Park, NJ 07932 (e-mail: vinay@research.att.com; amy@research.att.com).

Publisher Item Identifier S 1057-7149(01)01659-1.

Consider a standard, single description coder (SDC) that has been designed to minimize distortion for a given rate R^*, producing distortion $D^* = D(R^*)$, where $D(R)$ is the operational rate-distortion (RD) characteristic of the coder. One approach to multiple description coding would be to partition the SDC's output bitstream into two equal-sized bitstreams for transmission on each of the two channels. When both channels are received, a high-quality reconstruction is decoded with distortion, $D_0 = D^*$ achieved at a total rate of R^*. However, if only one channel is received, even the best partitioning of bits generally results in a single-channel reconstruction with distortion, D_1, that is unacceptably high. In this paper, we assume that the two channels have equal failure probabilities and use D_1 to denote the average distortion when only one channel is working. Theoretical work in the early 1980s developed rate-distortion bounds governing the tradeoff between two-channel (D_0) and single-channel (D_1) distortions for various sources [1]–[3]. For the Gaussian source [2], the lower bound establishes that any partitioning of the SDC bitstream forces D_1 to be very high. To reduce D_1 below that achievable with splitting the SDC stream, the total bit rate over both channels must exceed R^*. The extra bit rate (beyond R^*) does not affect $D_0 = D^*$, but allows dependencies to be incorporated into the two transmitted bitstreams, so that a better reconstruction can be formed from either single channel. We call this extra bit rate the redundancy, denoted by ρ, and a natural performance measure of an MDC system is the amount by which D_1 is reduced for a fixed investment of redundancy rate ρ, when R^* or equivalently D^* is fixed. This can be formulated by a redundancy-rate distortion (RRD) curve, which relates ρ and D_1, for a fixed D^*. The RRD curve quantifies the effectiveness of the redundancy of a representation in reducing single-channel distortion, and provides a useful tool for measuring and optimizing the efficiency of MDC coders.

Previous research in the area have considered the design of MDC coders patterned after different source coding approaches. For example, Vaishampayan proposed a method for structuring the index information of a standard scalar quantizer for meeting MDC objectives [4]. Originally, this idea was applied to code the signal samples directly. Later, it was extended for coding the transform coefficients in a block transform coder [5]. More recently, Servetto *et al*. have considered its application in wavelet-based image coders [6]. In another approach [7], Ingle and Vaishampayan modified standard DPCM coding to meet MDC objectives. Whereas the optimal DPCM coder should select a prediction filter that whitens the prediction error, the prediction filter in [7] is designed to provide a tradeoff between independent prediction errors and interchannel dependency. In

[8], Jafarkhani adapted standard trellis-coded quantization to fulfill MDC objectives. The MDC coders in each of these three cases are built upon specific forms of source coding, and, thus, each is designed to be applied to sources that are most naturally coded with the corresponding single-description coder.

Transform-based source coding algorithms have been widely applied for compressing many types of sources (e.g., audio, image, and video sources), and it is reasonable to expect that a transform-based approach to MDC would have advantages coding such sources, particularly at low redundancy rates. This paper proposes a transform-based approach to MDC that uses linear transforms to provide a smooth tradeoff between source coding efficiency and single channel distortion. We analyze the MDC performance of these transforms for Gaussian test sources.

Standard transform coding methods use a linear transform to decorrelate variables to be coded and transmitted. This coding method is known to be optimal in the RD sense for Gaussian sources. It is natural to consider how to adapt these transforms to force controlled correlation among the coefficients, rather than to decorrelate them completely. Motivated by this rationale, we have been exploring the use of correlating transforms for achieving the MDC objective [9]–[11]. In our proposed scheme, we consider a typical block coding framework in which a source is partitioned into N-sample blocks, or vectors, of random variables. A linear transform is applied to each vector, to generate two sets of coefficients that are correlated between the sets but are such that the coefficients within each set are uncorrelated with each other. Coefficients in each set are quantized and entropy coded independently with scalar quantizers and entropy coders matched to the statistics of each coefficient.

To simplify the design of the correlating transform, we propose to group the N samples into $N/2$ pairs and apply a pairwise correlating transforms (PCT) to each pair. Given this structure, there are two essential design problems. The first is how to optimize the RRD performance of a PCT operating on a pair. For two independent Gaussian variables, we analyze the RRD performance of an arbitrary 2×2 transform and derive the unique optimal transform that minimizes the single description distortion at a given redundancy. The second design problem is how to incorporate pairwise transforms in a system based on N samples. Assuming the N-dimensional vector process is independent and identically distributed (i.i.d.) with diagonal Gaussian distributions, we consider how to optimally allocate redundancy among a prescribed set of pairs, and derive the optimal pairing strategy that achieves the best RRD performance.

Using transforms to introduce correlation between multiple descriptions has also been considered by Goyal *et al*. [12]. The RRD performance of a correlating transform with an arbitrary dimension N was analyzed for a generally asymmetrical channel environment. For the case of $N = 2$ and symmetric channels, a closed-form solution was obtained for the optimal transform. They also proposed to generate $N \geq 2$ descriptions by cascading 2×2 transforms. Goyal *et al*.also explored the use of over-complete frame expansions for accomplishing MDC [13], where a $N \times K$ $(N \geq K)$ linear transform is used to generate N descriptions.

In the remaining sections of this paper, Section II formally defines the concept of redundancy and introduces the RRD curve as a performance measure for MDC. Section III describes the general framework of using block transforms for realizing MDC objectives, and the simplified structure using PCT's. Implementation details, including the quantization process, will be discussed. In Section IV, we analyze the performance of the building block of the proposed system—applying a PCT to a single pair. We derive the RRD curve achievable by an arbitrary 2×2 transform and determine the optimal transform family, which can achieve minimal D_1 for a given ρ. Then, in Section V, we consider the two key questions in applying the PCT to $N > 2$ samples. We derive the optimal redundancy allocation among a given set of pairs and the optimal pairing strategy. Finally, in Section VI, we present an approach for applying the proposed transform-based MDC system to image coding. Our image coder is a modified version of the JPEG coder and the PCT is applied to the DCT coefficients. The redundancy is defined as the additional bit rate compared to that of the JPEG coder.

II. Redundancy-Rate Distortion

All approaches to MDC involve creating redundancy in the bitstreams transmitted over different channels. However, redundancy can take on many forms and may be difficult to quantify and compare. For example, both the simple replication of bits on two channels and the creation of obscure relationships between the bits on two channels create redundancy between the bitstreams. In the first case, it is easy to count the number of replicated bits, while in the second case, it may be unclear how to quantify redundancy. This section formalizes a framework for quantifying arbitrary forms of redundancy, and proposes the RRD function as a measure of the efficiency of an MDC coder.

Given the average channel rate across both channels, R, an MDC coder attempts to jointly minimize two distortion measures $D_0(R)$: average distortion of the two-channel reconstruction and $D_1(R)$: average distortion of the one-channel reconstruction given equi-probable loss of either channel.

The coder minimizing only D_0 for a given rate R is the standard source coder, which we will call SDC. The performance of such a coder is characterized by its RD function for a given source. Intuitively, redundancy is the bit rate sacrificed compared to the SDC for the purpose of lowering D_1. More precisely, the redundancy in coding a source at an average two-channel distortion D_0, is the difference, $\rho = R - R^*$, between the per-variable transmitted bit rate R and $R^* = R(D_0)$, the source RD function evaluated at D_0. Note that R^* is the lowest rate needed by any coder, within some class of coders, to achieve the same two-channel distortion. The purpose of introducing redundancy in a representation is to lower the average one-channel distortion D_1, and we use the function $\rho(D_1; D_0)$ to denote the relationship between ρ and D_1. This function, the *redundancy rate-distortion function,* describes how many bits of redundancy per variable are required by a coder to achieve a desired D_1 at a given average two-channel distortion D_0. Likewise, the *distortion redundancy-rate function* $D_1(\rho; D_0)$ describes the achievable average one-channel distortion D_1 for a given redundancy

ρ and average two-channel distortion D_0. For the coders analyzed in this paper, these functions are approximately independent of D_0, and for these we write $\rho(D_1)$ or $D_1(\rho)$, suppressing the dependency on D_0. For any fixed D_0, redundancy can range from 0 (SDC) to $R(D_0)$ (replicating the SDC bitstream on both channels). The corresponding single channel distortion varies from $(\sigma^2 + D_0)/2$ to D_0, where σ^2 represents the variance of the source variable.

In addition to providing a framework to visualize the ability of additional redundancy to decrease the single-channel distortion, the RRD also provides a useful framework for understanding how to allocate a total available redundancy to different source variables. Suppose there are M different sources, each having a RRD function $D_{1,m}(\rho_m)$. Furthermore, assume that the distortions from different sources are additive. If the desired average redundancy is ρ/M per source variable, we define the Lagrange function

$$L(\rho_m, \forall m, \lambda) = \sum_{m=1}^{M} D_{1,m}(\rho_m) + \lambda \left(\sum_{m=1}^{M} \rho_m - M\rho \right).$$

Setting $\partial L/\partial \rho_m = 0$ yields

$$\lambda = \frac{\partial D_{1,m}(\rho_m)}{\partial \rho_m}, \qquad \forall m. \tag{1}$$

Therefore, similar to the case of optimal bit rate allocation, where each encoder operates at the same slope on its RD curve, *for optimal redundancy allocation, each multiple description coder should operate at the same slope on its RRD curve*. We use these results in Section V to allocate redundancy across multiple pairs of coefficients.

Finally, like other uses of RD curves, we can consider either ideal RRD curves characterizing fundamental properties of a source, or operational RRD curves characterizing the performance of specific classes of coders on the source. This paper is primarily concerned with characterizing operational RRD curves of transform-based coders for i.i.d vector Gaussian processes.

III. Multiple Description Transform Coding

A. Basic Framework

As described in the introduction, we propose to use the standard transform coding framework to realize the objective of MDC. In conventional transform coding, the transform is used to decorrelate the input variables. Here, we use a transform to introduce a controlled amount of correlation among the transformed coefficients. We limit our discussion to nonoverlapping linear transforms, mapping N input variables to N coefficients. For the two description case, the transform coefficients are partitioned into two groups, each consisting of $N/2$ coefficients, which are coded and transmitted through two separate channels. The transform is designed to introduce controlled correlation so that one group, if lost during transmission, can be estimated from the other group with a certain accuracy. To minimize the bit rate, the coefficients in the same group should be independent, so that they can be coded as efficiently as possible. In general, though the N input variables are themselves correlated

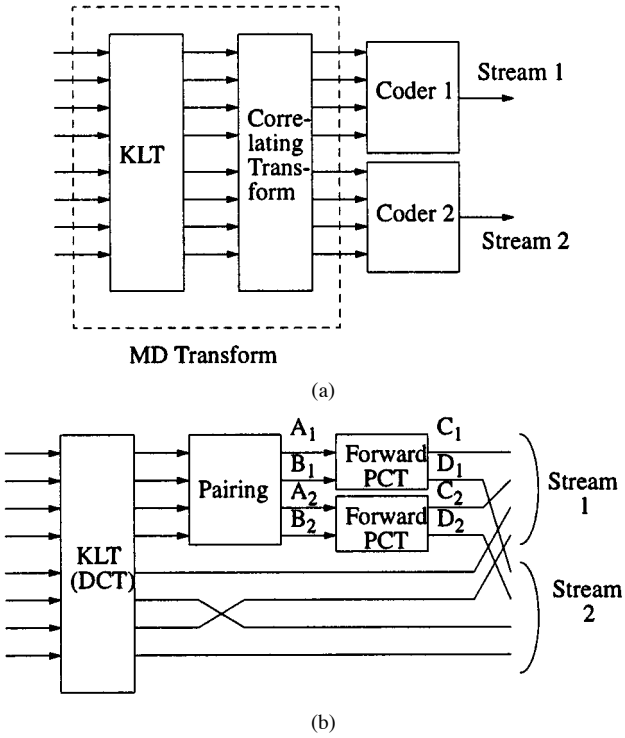

Fig. 1. Multiple description transform coding using a cascade of KL transform and a correlating transform: (a) the general framework using $N \times N$ transform and (b) the implementation using pairwise correlating transform.

with an arbitrary correlation matrix, the design of the transform can be simplified considerably by first decorrelating the input variables using a Karhunen–Loeve transform (KLT), as shown in Fig. 1(a).

To further reduce the complexity of designing the transform, we use the system illustrated in Fig. 1(b). Here, the KLT coefficients are grouped into pairs, a 2×2 transform is applied to each pair, and one coefficient from each output pair is assigned to each channel. This structure maintains decorrelation within each channel (by virtue of the independence among pairs), while providing a means to continuously control the correlation between the two descriptions. The coder design thus involves two issues: 1) how to optimize the MDC performance of a transform operating on a pair and 2) how to incorporate pairwise transforms in a system based on N samples. We consider the first in Section IV and the second in Section V.

B. Pairwise Correlating Transform and its Implementation

Conceptually, our pairwise MDC transform $\mathbf{T}$ takes two independent input variables A and B, and outputs two transformed variables C and D

$$\begin{bmatrix} C \\ D \end{bmatrix} = \mathbf{T} \begin{bmatrix} A \\ B \end{bmatrix}. \tag{2}$$

The transform $\mathbf{T}$ controls the correlation between C and D, which in turn controls the redundancy of the MDC coder.

If the transform $\mathbf{T}$ is nonorthogonal, direct quantization of the transform coefficients defines quantization cells in the signal space that are skewed versions of cubic cells. The mean-squared error of such skewed cells is always greater than the cubic cell with matching volume. To circumvent this problem, we first

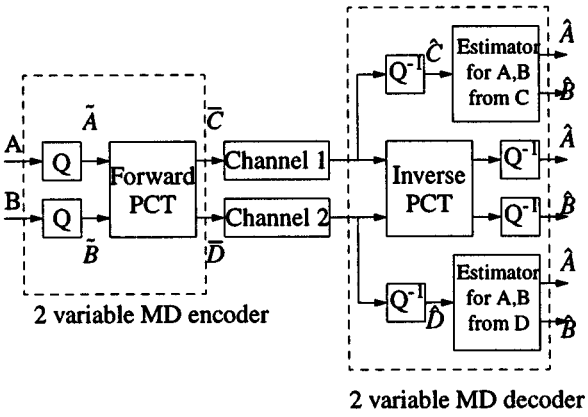

Fig. 2. Coding and decoding process for a single pair: the basic scheme.

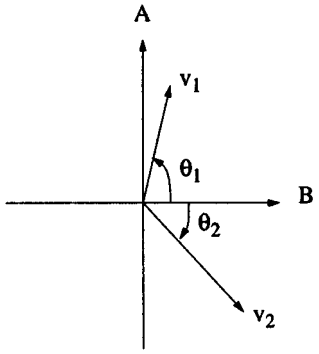

Fig. 3. Parameterization of the pairing transform.

apply a scalar quantizer to the two input variables A and B to yield integer indices $\overline{A}$ and $\overline{B}$, and then apply a discrete version of the transform to $\overline{A}$ and $\overline{B}$ to yield integer indices in the transform domain $\overline{C}$ and $\overline{D}$. The discretized transform should be designed to allow reversible integer-to-integer mapping. It can be thought of as renaming quantization cells in the A–B domain using coordinates in the C–D domain in a way that guarantees one-to-one correspondence. The lifting scheme [14] has been proposed for implementing such lossless discrete transforms by expressing the original transform as a sequence of predictors through an LU decomposition. Recently published work [15] has studied how to minimize quantization noise in lossless discrete transforms, and has proposed an algorithm with lower quantization error than the lifting scheme. In applying the method of [15] in this research, we also find that it creates more balanced quantization errors between C and D.

Therefore, in the MDC system with two input variables, shown in Fig. 2, we implement the pairwise correlating transform as follows. For an arbitrary transform with determinant one,[1]

$$\mathbf{T} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

the forward transform[2] with quantization stepsize Q is implemented as ($[\cdot]$ denotes rounding)

$$\begin{aligned} \overline{A} &= \left[\frac{A}{Q}\right], \quad \overline{B} = \left[\frac{B}{Q}\right] \\ W &= \overline{B} + \left[\frac{1+c}{d}\overline{A}\right] \\ \overline{D} &= [dW] - \overline{A} \\ \overline{C} &= W - \left[\frac{1-b}{d}\overline{D}\right]. \end{aligned}$$

Then, $\overline{C}$ and $\overline{D}$ are individually entropy coded, and their resulting bitstreams are sent on two separate channels.

[1]$\mathbf{T}$ must have determinant one in order to define a one-to-one correspondence between the index pairs $(\overline{A}, \overline{B})$ and $(\overline{C}, \overline{D})$. The one-to-one correspondence is necessary to maintain the coding efficiency of the MDC system.

[2]Before applying this algorithm to a given transform, a remapping to minimize the quantization error may first be necessary [15] so that the basis vectors of the mapped matrix correspond to the Voronoi region of the lattice generated by the original transform.

Assuming both channels are received, the corresponding inverse transform is implemented as

$$\begin{aligned} W &= \overline{C} + \left[\frac{1-b}{d}\overline{D}\right] \\ \overline{A} &= [dW] - \overline{D} \\ \overline{B} &= W - \left[\frac{1+c}{d}\overline{A}\right] \\ \hat{A} &= \overline{A}Q, \; \hat{B} = \overline{B}Q. \end{aligned}$$

It can be easily shown that the transform pair is reversible between $(\overline{A},\overline{B})$ and $(\overline{C}, \overline{D})$.

If only one channel is working, say, the channel carrying $\overline{C}$, we consider $\overline{C}$ as the index of C quantized by a step-size Q and obtain $\tilde{C} = \overline{C}Q$. It is possible either to estimate A and B directly from $\tilde{C}$, or to estimate D from $\tilde{C}$ first and then apply the inverse transform. Assuming a linear estimator, both methods are equivalent. In the latter method, D is estimated from $\tilde{C}$ using

$$\hat{D}(\tilde{C}) = \gamma_{D\tilde{C}}\tilde{C} \tag{3}$$

and $\hat{A}$ and $\hat{B}$ are obtained from the inverse transform

$$\begin{bmatrix} \hat{A} \\ \hat{B} \end{bmatrix} = \mathbf{T}^{-1} \begin{bmatrix} \tilde{C} \\ \hat{D} \end{bmatrix}. \tag{4}$$

Similarly, when only $\overline{D}$ is available, we set $\tilde{D} = \overline{D}Q$, estimate $\hat{C}$ and pre-multiply by the inverse transform to get $\hat{A}$ and $\hat{B}$. If both channels are lost, then the best estimates for A and B are their respective mean values.

This implementation of the MDC pairwise transform minimizes the contribution of quantization error into the distortion of the single-channel reconstruction. In particular, assuming the rounding error at different steps in the forward transform are independent and each has a variance of $1/12$, we can easily show that the quantization errors $q_c = C - \overline{C}Q$ and $q_d = D - \overline{D}Q$ are characterized by

$$\begin{aligned} \sigma_{qd}^2 &= (1 + 2d^2 + c^2)\sigma_{qab}^2, \\ \sigma_{qc}^2 &= \left(1 + a^2 + 2b^2 + \left(\frac{1-b}{d}\right)^2\right)\sigma_{qab}^2 \end{aligned} \tag{5}$$

where $\sigma_{qab}^2 = Q^2/12$ is the mean-square quantization error in A and B. We can see that in general, $\sigma_{qc}^2 \geq \sigma_{qab}^2$, $\sigma_{qd}^2 \geq \sigma_{qab}^2$.

Note also that these errors are generally smaller than those obtained when using a lifting implementation of the transform, which also produces unequal quantization errors in C and D.

While in general, $\sigma_{qc}^2 > \sigma_{qd}^2$, we have found in our numerical simulations that, for most transform parameters and quantization stepsizes of interest, they are nearly equal. Throughout the following analysis, we assume $\sigma_{qc}^2 = \sigma_{qd}^2$ which we denote σ_{qcd}^2.

IV. Redundancy Rate Distortion Analysis for Pairwise Transform

In this section, we analyze the redundancy and distortion associated with a single pair using a 2×2 transform. We will assume the input variables, A and B, are two independent Gaussian random variables, with zero mean and variances σ_A^2 and σ_B^2, respectively. Without loss of generality, we assume $\sigma_A \geq \sigma_B$, and define their ratio to be $r = \sigma_A/\sigma_B \geq 1$.

We first derive the RRD function for an arbitrary transform. We then determine the optimal transform which achieves the minimal D_1 for a given ρ. Finally, we consider the RRD performance of the orthogonal transform family, and compare it to the optimal transform. For each case, we present the one-channel distortion as a function of the redundancy, and then examine the allowable range of redundancies for the given transform. Throughout, we verify the equations with numerical simulation results.

A. General Case: An Arbitrary Transform

In this section, we derive the RRD function for an arbitrary 2×2 linear transform T. For convenience of the analysis, we parameterize $\mathbf{T}$ by

$$\mathbf{T}^{-1} = \begin{bmatrix} r_1 \sin\theta_1 & r_2 \sin\theta_2 \\ r_1 \cos\theta_1 & r_2 \cos\theta_2 \end{bmatrix} = [\mathbf{v}_1 \quad \mathbf{v}_2]$$

or

$$\mathbf{T} = \begin{bmatrix} r_2 \cos\theta_2 & -r_2 \sin\theta_2 \\ -r_1 \cos\theta_1 & r_1 \sin\theta_1 \end{bmatrix}. \tag{6}$$

This transform essentially represents the original variables using two new basis vectors, $\mathbf{v}_1$ and $\mathbf{v}_2$, which are in general nonorthogonal. The parameters r_1 and r_2 control the lengths of the two basis vectors, while θ_1 and θ_2 specify their directions, as illustrated in Fig. 3. To support a lossless integer implementation of the matrix multiplication by $\mathbf{T}$, we require that $\mathbf{T}$ have a determinant one, giving

$$r_1 r_2 = \frac{1}{\sin\Delta\theta} \tag{7}$$

where $\Delta\theta = \theta_1 - \theta_2$. Because $r_1 r_2 > 0$, we have $0 < \Delta\theta < \pi$. Furthermore, without loss of generality, we limit the range of θ_1 to $0 \leq \theta_1 \leq \pi/2$. The variances of C and D follow easily from (6)

$$\sigma_C^2 = r_2^2(\cos^2\theta_2\sigma_A^2 + \sin^2\theta_2\sigma_B^2)$$

and

$$\sigma_D^2 = r_1^2(\cos^2\theta_1\sigma_A^2 + \sin^2\theta_1\sigma_B^2). \tag{8}$$

To derive the RRD function, we first analyze the redundancy introduced by a given a transform, deriving an expression for the transform parameters needed to obtain a given redundancy. Finally, we determine the one-channel distortion as a function of these transform parameters.

Transform redundancy is controlled by the amount of correlation introduced by $\mathbf{T}$, which we parameterize with the correlation angle ϕ between C and D: $E\{CD\} = \sigma_C\sigma_D \cos\phi$. A simple relationship between the variances of the four variables comes from relating the covariance matrices $\mathbf{R}_{AB}$ and $\mathbf{R}_{CD}$, given by

$$\mathbf{R}_{AB} = \begin{bmatrix} \sigma_A^2 & 0 \\ 0 & \sigma_B^2 \end{bmatrix}$$

and

$$\mathbf{R}_{CD} = \begin{bmatrix} \sigma_C^2 & \sigma_C\sigma_D\cos\phi \\ \sigma_C\sigma_D\cos\phi & \sigma_D^2 \end{bmatrix}.$$

Taking the determinant of both sides of $\mathbf{R}_{CD} = \mathbf{T}\mathbf{R}_{AB}\mathbf{T}^T$, we have

$$\sigma_C^2\sigma_D^2 \sin^2\phi = \sigma_A^2\sigma_B^2. \tag{9}$$

Because C and D are correlated, more bits are required to send them than to send A and B. The redundancy, defined as this excess bit rate per variable,[3] can be determined from the rate-distortion functions for Gaussian variables. Allocating bit rate optimally within each pair, the rates required for coding the pairs (C, D) and (A, B) are (in bits per variable)

$$R = \frac{1}{2}\log_2\frac{\sigma_C\sigma_D}{D_0} + K, \quad R^* = \frac{1}{2}\log_2\frac{\sigma_A\sigma_B}{D_0} + K \tag{10}$$

for some constant K. For Gaussian variables and with entropy coding, $K = (1/2)\log_2(\pi e/6)$. Note that the above relation is valid only when each variable is coded at a relatively high rate so that 1) high-resolution approximation can be used in relating rate and distortion [16] and 2) the approach of [17] for optimal rate allocation between two variables can be applied. We will, however, consider the effect of coarse quantization later in Section V-C.

Thus, using (9) and (10), the redundancy per variable is

$$\rho = R - R^* = \frac{1}{2}\log_2\frac{\sigma_C\sigma_D}{\sigma_A\sigma_B} = -\frac{1}{2}\log_2\sin\phi. \tag{11}$$

Note that the effect of the constant K, which depends on the source statistics, cancels in measuring the redundancy, so that the above relation between ρ and signal variances is valid for sources other than Gaussian when high-rate approximation is applicable. This relation is, however, not accurate when the rate is very low, when the quantization error is on the same order as the signal variance. This is further discussed in Section V-C.

To express ϕ (and hence the redundancy) explicitly in terms of the transform parameters and σ_A^2 and σ_B^2, we substitute (8)

[3]Note that in [10], [11], redundancy was defined *per pair*. In this paper, it is defined *per variable*.

into (9) and incorporate (7) to yield

$$\cot\,\Delta\theta = \frac{2r\cot\phi - (r^2-1)\sin 2\theta_1}{2x}, \quad \text{with } x = r^2\cos^2\theta_1 + \sin^2\theta_1 \tag{12}$$

describing the angle $\Delta\theta$ that provides a given redundancy ρ (equivalently ϕ), for a fixed θ_1.

We now evaluate the one-channel distortion as a function of the correlation angle ϕ. When only $\tilde{C}$ is received, we linearly estimate D as in (3), using the optimal linear estimation coefficient $\gamma_{D\tilde{C}} = (\sigma_D/\sigma_{\tilde{C}})\cos\tilde{\phi}$, and producing the minimum mean-squared estimation error $E\{(D-\hat{D}(\tilde{C}))^2\} = \sigma_D^2\sin^2\tilde{\phi}$. (Expressions for estimating C from $\tilde{D}$ follow from symmetry.) Here, $\tilde{\phi}$ is the correlation angle between the random variables D and $\tilde{C}$. Let $q_c = C - \tilde{C}$ and $q_d = D - \tilde{D}$ denote the quantization errors. Under the high-rate assumption, C and D are both quantized with sufficiently fine quantizers, so that q_c and q_d are independent, and each independent of C and D. Therefore, $E\{D\tilde{C}\} = E\{DC\}$, and so $\cos\tilde{\phi} = (\sigma_C/\sigma_{\tilde{C}})\cos\phi$ and $\sin^2\tilde{\phi} = \sin^2\phi + (\sigma_{q_c}^2/\sigma_{\tilde{C}}^2)\cos^2\phi$. Thus, using the optimal linear estimate

$$\hat{D}(\tilde{C}) = \frac{\sigma_C\sigma_D}{\sigma_C^2+\sigma_{qc}^2}\cos\phi\tilde{C} \tag{13}$$

the minimum mean-squared estimation error is

$$E\{(D-\hat{D}(\tilde{C}))^2\} = \sigma_D^2\sin^2\phi\left(1-\frac{\sigma_{qc}^2}{\sigma_C^2+\sigma_{qc}^2}\right) + \frac{\sigma_D^2\sigma_{qc}^2}{\sigma_C^2+\sigma_{qc}^2}.$$

Variables A and B can be reconstructed from $\tilde{C}$ and $\hat{D}(\tilde{C})$ as

$$\begin{bmatrix}\hat{A}\\ \hat{B}\end{bmatrix} = \mathbf{T}^{-1}\begin{bmatrix}\tilde{C}\\ \hat{D}(\tilde{C})\end{bmatrix} = \begin{bmatrix}A\\ B\end{bmatrix} + q_c\mathbf{v}_1 + (D-\hat{D}(\tilde{C}))\mathbf{v}_2.$$

Ignoring the correlation between the quantization error q_c and the estimation error $(D-\hat{D}(\tilde{C}))$, the reconstruction error is

$$E\{(A-\hat{A})^2+(B-\hat{B})^2|\tilde{C}\} = \sigma_{qc}^2r_1^2 + E\{(D-\hat{D}(\tilde{C}))^2\}r_2^2.$$

Similar equations can be obtained for estimating C from $\tilde{D}$, using symmetry.

Assuming that the two channels have equal failure probabilities, the average one-channel distortion per variable is

$$D_1 = \tfrac{1}{4}(E\{(A-\hat{A})^2+(B-\hat{B})^2|\tilde{C}\} + E\{(A-\hat{A})^2 + (B-\hat{B})^2|\tilde{D}\}) = D_{1,0}+D_{1,q} \tag{14}$$

where

$$D_{1,0} = \tfrac{1}{4}(r_1^2\sigma_C^2\sin^2\phi + r_2^2\sigma_D^2\sin^2\phi) \tag{15}$$

is the estimation error in the absence of quantization error, and

$$D_{1,q} = \frac{\sigma_{qcd}^2}{4}\left(r_1^2\left(1+\frac{\sigma_C^2}{\sigma_D^2}\cos^2\phi\right) + r_2^2\left(1+\frac{\sigma_D^2}{\sigma_C^2}\cos^2\phi\right)\right) \tag{16}$$

is the contribution from the quantization error. Equation (16) assumes that $\sigma_{qc}^2 = \sigma_{qd}^2 = \sigma_{qcd}^2$ (as discussed in Section III-B), and that $\sigma_{qc}^2 \ll \sigma_C^2$, $\sigma_{qd}^2 \ll \sigma_D^2$, leading us to ignore terms smaller than σ_{qc}^2/σ_C^2, σ_{qd}^2/σ_D^2. Equations (14)–(16) describe the RRD function in terms of the correlation angle ϕ and the variances of C and D.

To express the RRD function explicitly in terms of the transform parameters and σ_A^2 and σ_B^2, we first substitute (7)–(9) into (15) and (16) to obtain

$$D_{1,0} = \frac{r\sin\phi\sigma_B^2}{4\sin\Delta\theta}\left(\frac{x\sin\phi}{r\sin\Delta\theta}+\frac{r\sin\Delta\theta}{x\sin\phi}\right) \tag{17}$$

and

$$D_{1,q} = \frac{\sigma_{qcd}^2}{4\sin\Delta\theta}\left(r_1^2\sin\Delta\theta+\frac{1}{r_1^2\sin\Delta\theta}\right) + \frac{\cos^2\phi\sigma_{qcd}^2}{2\sin\Delta\theta}\cdot\left(\frac{r_1^2x^2\sin^2\phi}{r^2\sin\Delta\theta}+\frac{r^2\sin\Delta\theta}{r_1^2x^2\sin^2\phi}\right). \tag{18}$$

Though this form of $D_{1,0}$ is useful for deriving the optimal transform, x and $\Delta\theta$ still hide dependencies on θ_1 and ϕ through (12). To give an explicit characterization of the relationship between D_1 and ρ, we use (12) along with the identity $1/(\sin^2\Delta\theta) = 1+\cot^2\Delta\theta$, giving (19), shown at the bottom of the page.

For a given ρ or ϕ, (19) describes the RRD function for any choice of r_1 and θ_1, which are the free parameters in the transform. The other two parameters, r_2 and θ_2 (or $\Delta\theta$), are constrained by (7) and (12). Notice that the length parameter r_1 does not affect $D_{1,0}$ and only affects the quantization error, $D_{1,q}$, in (18).

Because (19) is a periodic function of ϕ, for a given θ_1 redundancies are limited to the range $\rho < \rho_{\max}(\theta_1)$, where $\rho_{\max}(\theta_1)$

$$D_{1,0} = \frac{(\sigma_A^2-\sigma_B^2)((\sigma_A^2\cos^2\theta_1-\sigma_B^2\sin^2\theta_1)\sin^2\phi-\sigma_A\sigma_B\sin 2\theta_1\sin\phi\cos\phi)+2\sigma_A^2\sigma_B^2}{4(\sigma_A^2\cos^2\theta_1+\sigma_B^2\sin^2\theta_1)} \tag{19}$$

corresponds to the ϕ satisfying $\partial D_{1,0}/\partial\phi = 0$. For a given θ_1, this minimum-D_1 correlation angle is

$$\cot\ \phi_{\min}(\theta_1) = r\ \cot\ \theta_1 \tag{20}$$

and the corresponding $\rho_{\max}(\theta_1)$ is $\rho_{\max}(\theta_1) = \log_2(1 + r\cot\ \theta_1)$. The $\Delta\theta$ that achieves $\rho_{\max}(\theta_1)$ is found by substituting (20) into (12), giving

$$\cot\ \Delta\theta_{\min} = \cot\ \theta_1 \quad \text{or} \quad \Delta\theta_{\min} = \theta_1 \quad \text{or} \quad \theta_{2,\max} = 0. \tag{21}$$

Substituting into (17) gives the minimum achievable D_1

$$D_{1,0,\min}(\theta_1) = \frac{\sigma_B^2}{4} + \frac{\sigma_A^2\sigma_B^2}{4(\sigma_A^2\ \cos^2\ \theta_1 + \sigma_B^2\ \sin^2\ \theta_1)} \geq \frac{\sigma_B^2}{2}. \tag{22}$$

The maximum D_1 occurs at zero redundancy, which is achieved when $\phi = \pi/2$ and C and D are uncorrelated. Thus, from (12), for a given θ_1, $\Delta\theta$ must satisfy

$$\cot\ \Delta\theta_{\max} = -\frac{(r^2-1)\ \sin\ 2\theta_1}{2x} \tag{23}$$

to yield the associated zero-redundancy estimation error $D_{1,0,\max} = (\sigma_A^2 + \sigma_B^2)/4$. This error is equivalent to the average single-channel distortion if we had sent A and B individually on the two channels.

Fig. 4 illustrates the RRD function in (19) as $\Delta\theta$ is varied for three different values of θ_1. Source statistics are $\sigma_A = 1$, $\sigma_B = 0.4$, and the quantization step-size is $Q = 0.02$. The simulations implement the process in Section III-B. We generate pairs of independent Gaussian random variables A and B with the specified variances, uniformly quantize A and B with step-size Q to yield $\overline{A}$ and $\overline{B}$, and then apply the appropriate transform using the method described above to yield $\overline{C}$ and $\overline{D}$. The bit rates R^* and R are the entropies estimated from the histograms of $\overline{A}$ and $\overline{B}$, and of $\overline{C}$ and $\overline{D}$, respectively. Using statistics generated from the resulting $\overline{C}$ and $\overline{D}$, we also find the optimal estimator according to (13), from which we reconstruct $\hat{A}$ and $\hat{B}$ to estimate the average single-channel distortion. The simulation results (stars) obtained using 8100 samples match the theoretical results for $D_{1,0}$ (lines) well for this fine quantizer.

B. Optimal Transform and its Redundancy Rate-Distortion Function

The RRD relation derived in the previous section is for arbitrary values of r_1 and θ_1. This section derives conditions on the transform parameters to minimize D_1 for a given ρ. First, a class of transforms is identified that minimizes $D_{1,0}$ in (17), all yielding the same RRD curve when quantization error is ignored. Then, a unique optimum transform is derived by considering quantization error and minimizing $D_{1,q}$ in (18).

The parenthetical terms in (17) and (18) have the form $z + 1/z$, which has an extremum when $z = 1/z$. For the parenthetical term in (17), this is when

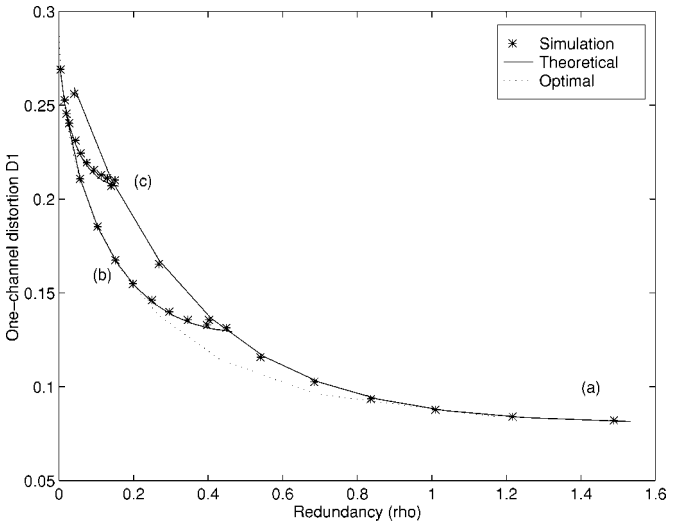

Fig. 4. Theoretical and simulated RRD functions of the general transform for three different values of θ_1 (a) $\pi/20$, (b) $3\pi/10$, and (c) $4\pi/10$.

$$x = \frac{r\ \sin\ \Delta\theta}{\sin\ \phi}. \tag{24}$$

This condition simultaneously optimizes the multiplicative factor in (17). Thus, combining (24) with (12) yields the optimal $\Delta\theta$ for minimizing $D_{1,0}$

$$\Delta\theta = 2\theta_1 \quad \text{or} \quad \theta_2 = -\theta_1. \tag{25}$$

As noted, minimizing $D_{1,0}$ constrains only $\Delta\theta$ or θ_2, but not the length of the basis vectors, r_1 and r_2. The single parameter θ_1 controls both the correlation angle ϕ and the redundancy ρ. The necessary θ_1 to achieve a desired redundancy is given by

$$\tan\ \theta_1 = r\tan\ \frac{\phi}{2} = r2^{2\rho}\left(1 - \sqrt{1 - 2^{-4\rho}}\right) \tag{26}$$

and the minimum $D_{1,0}$ [substituting (26) into (19)] is

$$\begin{aligned} D_{1,0,\text{opt}}(\rho) &= \tfrac{1}{4}\left(\sigma_A^2\left(1 - \sqrt{1-2^{-4\rho}}\right)\right. \\ &\qquad \left. + \sigma_B^2\left(1 + \sqrt{1-2^{-4\rho}}\right)\right) \qquad (27) \\ &= \tfrac{1}{4}\left((\sigma_A^2 + \sigma_B^2) - (\sigma_A^2 - \sigma_B^2)\sqrt{1-2^{-4\rho}}\right). \qquad (28) \end{aligned}$$

Next, considering the contribution of quantization error, we minimize $D_{1,q}$ in (18), giving

$$r_1^2\ \sin\ \Delta\theta = \frac{1}{r_1^2\ \sin\Delta\theta};$$

and

$$\frac{r_1^2\ \sin^2\ \phi x^2}{r^2\ \sin\ \Delta\theta} = \frac{r^2\ \sin\ \Delta\theta}{r_1^2\ \sin^2\ \phi x^2}. \tag{29}$$

The last two equations combine with (24) and (7) to yield

$$r_1 = r_2 = \frac{1}{\sqrt{\sin\ \Delta\theta}}. \tag{30}$$

Thus, necessary and sufficient conditions for the optimality of the transform are

$$r_1 = r_2 = \frac{1}{\sqrt{\sin 2\theta_1}}, \qquad \theta_2 = -\theta_1. \tag{31}$$

Substituting (26) and (30) in (18), the influence of quantization error is incorporated in the optimal RRD function as

$$D_{1,\,opt}(\rho) = (1 + \epsilon_{q,\,opt}(\rho)) D_{1,0,\,opt}(\rho) \tag{32}$$

where

$$\begin{aligned}\epsilon_{q,\,opt}(\rho) &= \frac{\sigma_{qcd^2}}{\sigma_A \sigma_B}\left(\frac{2}{\sin\phi} - \sin\phi\right) \\ &= \frac{\sigma_{qcd}^2}{\sigma_A \sigma_B}(2^{1+2\rho} - 2^{-2\rho}).\end{aligned} \tag{33}$$

When the quantization error $\sigma_{qcd}^2 \ll \sigma_A \sigma_B$, $\epsilon_{q,\,opt} \approx 0$.

The optimal transform, which satisfies the conditions in (31), has the form

$$\mathbf{T}^{-1} = \begin{bmatrix} \sqrt{\frac{\tan\theta_1}{2}} & -\sqrt{\frac{\tan\theta_1}{2}} \\ \sqrt{\frac{\cot\theta_1}{2}} & \sqrt{\frac{\cot\theta_1}{2}} \end{bmatrix}$$

or

$$\mathbf{T} = \begin{bmatrix} \sqrt{\frac{\cot\theta_1}{2}} & \sqrt{\frac{\tan\theta_1}{2}} \\ -\sqrt{\frac{\cot\theta_1}{2}} & \sqrt{\frac{\tan\theta_1}{2}} \end{bmatrix}. \tag{34}$$

The condition in (31) shows that *the optimal transform is formed by two equal length basis vectors that are rotated away from the original basis by the same angle in opposite directions*. The transform parameters r_1 and r_2 control the quantization density along the direction of the two new basis vectors. The equal length condition $r_1 = r_2$ essentially minimizes the impact of quantization error by equalizing the contribution from quantization errors on C and D. A second consequence of the equal length condition is that the transmitted pair C and D have equal energy: $\sigma_C^2 = \sigma_D^2$. Thus, aside from quantization considerations, the optimal transform forms two "balanced" descriptions, meaning that the two variables use the same rate, and generate the same estimation error independent of which channel is lost. Goyal and Kovacevic [12] independently established the optimality of the same transform without considering quantization error by explicitly constraining the transform to generate two balanced descriptions. As will be shown in Section V-C, in practice, for large quantization stepsizes, the descriptions generated by the optimal transform are decidedly unbalanced, and more so for the lifting implementation than the quantization scheme used here.

Fig. 5 shows the theoretical and simulated RRD curves for $\sigma_A = 1$, $\sigma_B = 0.4$, and $Q = 0.02$, corresponding to the low quantization error case. The theoretical curve marked "optimal" shows the RRD performance calculated according to (28), while the simulation results labeled "optimal" are obtained by implementing the process in Section III-B using the optimal transform

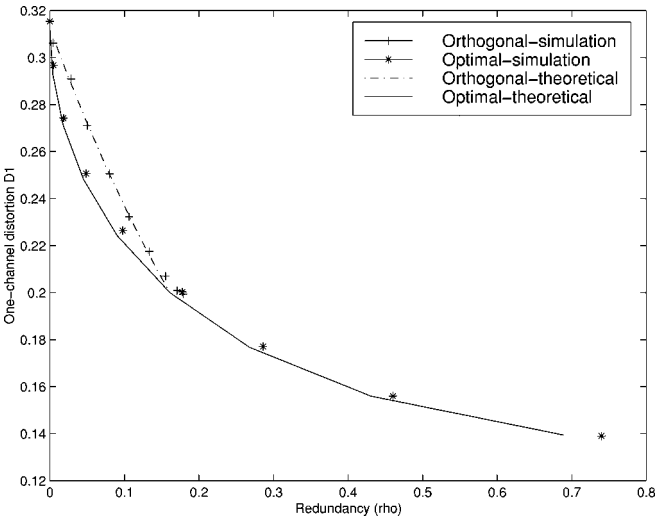

Fig. 5. Theoretical and simulated redundancy rate-distortion: optimal and orthogonal transforms.

in (34) with different values of θ. Curves labeled "orthogonal" correspond to the transform discussed later in Section III-B. We see that theoretical curves fit the simulation results well.

As shown in Fig. 5, $D_{1,\,opt}$ is largest for zero redundancy and decays rapidly for small redundancies, but decreases much more slowly for higher redundancies. At zero redundancy, we have $\tan\theta_{1,\max,\,opt} = r$ from (26), $\sigma_C^2 = \sigma_D^2 = \sigma_A \sigma_B$, and the associated zero-redundancy estimation error $D_{1,0,\max,\,opt} = (\sigma_A^2 + \sigma_B^2)/4$.

It is interesting to note that the optimal RRD function decays faster than exponentially at small values of redundancy. Equation (27) shows that increasing redundancy actually *amplifies* the contribution of σ_B^2 to D_1, while achieving very rapid reduction in the contribution of σ_A^2 to D_1. The overall RRD decay at small redundancies is dominated by the reduction in the contribution of σ_A^2, such that the limiting slope of the RRD curve approaches minus infinity as ρ approaches zero.

In the high redundancy region ($\rho \gg 1/2$), we can approximate $\sqrt{1 - 2^{-4\rho}}$ by $1 - 2^{-4\rho-1}$. Then (28) can be written as

$$D_{1,0,\,optimal} \approx \frac{\sigma_A^2 - \sigma_B^2}{8} 2^{-4\rho} + \frac{\sigma_B^2}{2} \to \frac{\sigma_B^2}{2}. \tag{35}$$

Therefore, D_1 decays exponentially with ρ for high redundancies at a rate that depends on the *difference* of the variances of the two variables, but converges to a nonzero error $\sigma_B^2/2$. This minimum D_1 is achieved in the limit as $\theta_1 \to 0$ so that C and D only carry information about A and $\sigma_C = \sigma_D = \sigma_A$. The corresponding redundancy is

$$\rho_{\max,\,opt} = \frac{1}{2}\log_2 \frac{\sigma_C \sigma_D}{\sigma_A \sigma_B} = \frac{1}{2}\log_2 r.$$

Note that $\rho_{\max,\,opt}$ equals half of the average rate for coding A to distortion σ_B^2.

Fig. 6 examines the impact of the requirement that the basis vectors of the optimal transform be equal length. In particular, it compares via simulation the performance of the optimal transform (34) and the transform in [10], which was motivated by the lifting transform and satisfies the constraint in (25) but has

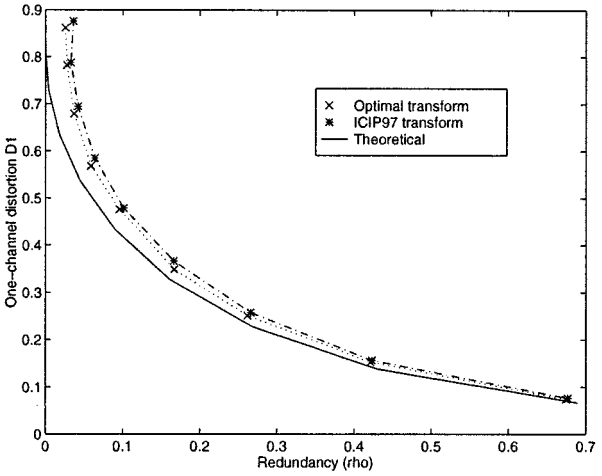

Fig. 6. Effect of equal length basis vectors.

$r_1 = (1/2)\sin\theta_1$. Both transforms have identical theoretical RRD curves in the absence of quantization errors, and the difference between them is only apparent when the ratio between the variances of the two variables is large and when the quantization step-size is large compared to the variance of the smaller variable. In particular, the factor $\epsilon_{q,\,lift}$ due to quantization is

$$\epsilon_{q,\,lift}(\rho) = \frac{\epsilon_{q,\,opt}(\rho)}{2}\left(2r\,\tan(\phi/2) + \frac{1}{2r\,\tan(\phi/2)}\right) \geq \epsilon_{q,\,opt}(\rho).$$

In Fig. 6, $\sigma_A = 1.8$, $\sigma_B = 0.1$ and the stepsize $Q = 0.09$. The solid line is the theoretical RRD performance without considering quantization errors ($D_{1,\,0,\,opt}$), the crosses and stars are simulation results for the optimal transform and the transform with basis vectors of unequal length, respectively. The optimal transform outperforms the unequal-length transform, particularly for low redundancy.

C. Orthogonal Transform

The optimal transform derived in the previous section is in general nonorthogonal. In this section, we present the RRD performance obtainable by orthogonal transforms and compare it to the optimal transform. If $\mathbf{T}$ is an orthogonal transform, quantization can be performed in the C–D domain instead of the A–B domain, thus reducing the quantization error. In addition, standard transform coders use orthogonal transforms, so it is natural to consider using one for a multiple description transform coder as well.

By letting $r_1 = r_2 = 1$, $\theta_2 = -\theta$, and $\theta_1 = -\theta + \pi/2$, we obtain an orthogonal transform with $\Delta\theta = \pi/2$

$$\mathbf{T}^{-1} = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$$

or

$$\mathbf{T} = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix}. \tag{36}$$

The rotation angle θ controls the redundancy. From (12), we obtain

$$\cot\,\phi = \frac{r^2 - 1}{2r}\,\sin 2\theta. \tag{37}$$

The RRD curve for the orthogonal case can be found by substituting this into (19) and (18), giving

$$D_{1,\,0,\,orth}(\rho) = \frac{\sigma_A^2 + \sigma_B^2}{4}\,\sin^2\,\phi = \frac{\sigma_A^2 + \sigma_B^2}{4}\,2^{-4\rho} \tag{38}$$

$$D_{1,\,q,\,orth}(\rho) = \epsilon_{q,\,orth}(\rho) D_{1,\,0,\,orth} \tag{39}$$

with

$$\epsilon_{q,\,orth}(\rho) = \sigma_{qcd}^2\left(\frac{2}{\sigma_A^2 + \sigma_B^2} + \left(\frac{1}{\sigma_A^2} + \frac{1}{\sigma_B^2}\right)(1 - 2^{-4\rho})\right).$$

If one directly quantizes the C and D variables, the quantization error $\sigma_{qcd}^2 = \sigma_{qab}^2$, which is smaller than if one quantizes A and B as described in Section III-B.

Unlike the optimal transform which can achieve redundancies in the range $(0,\,(1/2)\log_2 r)$, the orthogonal transform has a smaller range of redundancies, $(0,\,\rho_{\max,\,orth})$. The maximum redundancy can be found by setting $\theta = \pi/4$ which leads to the minimum correlation angle achievable by the orthogonal transform, $\phi_{\min,\,orth} = \arctan(2r/(r^2 - 1))$. Thus, the maximum redundancy is

$$\rho_{\max,\,orth} = -\frac{1}{2}\,\log_2\,\sin\,\phi_{\min,\,orth} = \frac{1}{2}\,\log_2\,\frac{r^2 + 1}{2r} \tag{40}$$

with corresponding single-channel distortion

$$D_{1,\,0,\,\min,\,orth} = \frac{\sigma_A^2\sigma_B^2}{\sigma_A^2 + \sigma_B^2} = \frac{r^2}{r^2 + 1}\,\sigma_B^2 \geq \frac{\sigma_B^2}{2}.$$

Note that the maximum redundancy is larger for larger r.

As can be seen in Fig. 5, which compares the RRD curves of the orthogonal and optimal transforms, the optimal transform outperforms the orthogonal transform everywhere except at two points: when $\rho = 0$ and when $\rho = \rho_{\max,\,orth}$. At these two points, the two transforms are identical. As we noted earlier, in the low-redundancy region ($\rho < \rho_{\max,\,orth}$), the RRD of the optimal transform has super-exponential decay, since it *amplifies* the contribution of σ_B^2 to D_1 in order to achieve a very large reduction in the contribution of σ_A^2. On the contrary, the orthogonal transform has an RRD function (38) that decays exponentially, with multiplicative constant equal to the *average* of the variances of the two variables. Thus, redundancy is used in a balanced, but suboptimal, way to reduce the contributions of both σ_A^2 and σ_B^2 in D_1.

V. Redundancy Allocation and Optimal Pairing Scheme

Given the desired redundancy to add when coding N (N even) KL coefficients, two design issues remain. The first is the choice of which of the $\prod_{m=1}^{M}(2m-1)$ possible pairing combinations should be used (where $M = N/2$ is the number of pairs), and the second is how much redundancy should be added

to each pair. In this section, we consider how to optimize the redundancy rate-distortion performance with respect to these two issues for the optimal pairwise transform of Section IV-B.

We begin by examining these issues without considering the effect of quantization. For a given targeted redundancy per variable, ρ, we find the optimal redundancy allocation among pairs, for a fixed pairing. For large redundancies, we derive the optimal pairing strategy. We also demonstrate numerically that this pairing strategy is optimal for all redundancies.

Next, we consider the impact of quantization. Specifically, we consider how the pairing strategy is affected when the quantizer becomes coarse with respect to the smaller variables.

A. Optimal Redundancy Allocation

For the case where the M MDC coders each use the optimal pairing transform in Section IV-B, we assume a fixed pairing combination, and define $\alpha_m^2 = (\sigma_{A,m}^2 + \sigma_{B,m}^2)/4$ and $\gamma_m^2 = (\sigma_{A,m}^2 - \sigma_{B,m}^2)/4$, where $\sigma_{A,m}^2$ and $\sigma_{B,m}^2$ are the variances of the two variables in the mth pair (A_m, B_m). Then, (28) can be written as

$$D_{1,m} = \alpha_m^2 - \gamma_m^2 \sqrt{1 - 2^{-4\rho_m}}. \tag{41}$$

As discussed in Section II, optimal redundancy allocation is achieved when we operate at the same slope on all RRD curves. Using (1) and collecting constants into λ, we have

$$\lambda = \gamma_m^2 2^{-4\rho_m} / \sqrt{1 - 2^{-4\rho_m}}.$$

Now, solving for $2^{-4\rho_m}$ gives us

$$2^{-4\rho_m}\left(\sqrt{1 + \frac{4\gamma_m^4}{\lambda^2}} + 1\right) = 2, \qquad m = 1, 2, \cdots, M. \tag{42}$$

Taking the product of the equations in (42) and using the constraint on the total redundancy yields

$$2^{-4M\rho}\prod_m \left(\sqrt{1 + \frac{4\gamma_m^4}{\lambda^2}} + 1\right) = 2^M. \tag{43}$$

Since it is generally difficult to find a closed-form solution of (43) for a given ρ, we solve it numerically to find λ, and then find the individual redundancies allocated to each pair using (42), which can be rewritten as

$$\rho_m = \frac{1}{4}\log_2\left(\sqrt{1 + \frac{4\gamma_m^4}{\lambda^2}} + 1\right) - \frac{1}{4}. \tag{44}$$

B. Optimal Pairing Strategy with Fine Quantization

Because (43) is difficult to solve analytically, it is difficult to determine the optimal pairing strategy for any redundancy. However, in the high redundancy case ($\rho_m \gg 1/2$), we can use the approximation in (35) to write

$$D_{1,m} \approx \frac{\sigma_{B,m}^2}{2} + \frac{\gamma_m^2}{2} 2^{-4\rho_m}, \qquad m = 1, 2, \cdots, M.$$

The first term in each equation is constant, whereas each of the second terms decay exponentially with ρ_m. Using the well-known optimal bit-allocation result for exponentially decaying RD functions [16] yields the following redundancy allocation:

$$\rho_m = \rho + \frac{1}{4}\log_2 \frac{\gamma_m^2}{\left(\prod_i \gamma_i^2\right)^{1/M}}. \tag{45}$$

The average single channel distortion per variable under this redundancy allocation is

$$D_1 = \frac{1}{M}\sum_m D_{1,m}(\rho) \tag{46}$$

$$= \frac{1}{2M}\sum_m \sigma_{B,m}^2 + \frac{1}{2}\left(\prod_m \gamma_m^2\right)^{1/M} 2^{-4\rho}. \tag{47}$$

To minimize D_1, the pairing combination should be such that $\sum_m \sigma_{B,m}^2$ and $(\prod_m \gamma_m^2)^{1/M}$ are simultaneously minimized. To minimize the first term, one should choose the M smallest variables to be the second variable in each pair. To minimize the second term, which is geometric mean of γ_m^2, one should choose the pairing so that γ_m^2 are as different as possible. This can be accomplished by pairing the kth largest variable with the $(N-k)$th largest one, which is consistent with the first requirement. Therefore, this is the optimal pairing scheme for large ρ.

Although we have derived the optimal pairing scheme above only for the case of high redundancy, we have also numerically solved the redundancy allocation problem for both high and low redundancies, for the cases of two, three and four pairs. Even for very small redundancies, the above pairing strategy is still optimal.

Fig. 7 shows theoretical and simulated RRD for the case of six variables using the optimal transform family. The variances of the variables are $\sigma_i^2 = 0.4^i$, and the step-size is $Q = 0.02$. Of the possible 15 ways to pair six variables, we show performance for three possible pairings: pairing neighbors [(1,2), (3,4), (5,6)], the pairing [(1,4), (2,3), (5,6)], and the pairing that pairs the biggest variable with the smallest [(1,6), (2,5), (3,4)]. Simulation results are also shown, with the three pairings indicated by stars, plusses and crosses, respectively.

C. Pairing Strategies Considering Quantization

The above analysis suggests that it is optimal to pair all variables, with the kth largest paired with the $(N-k)$th largest. However, when the smaller variable in a pair is too small relative to the quantization error, it is no longer meaningful to pair this variable. In this case, the two transformed coefficients are very similar, both essentially equal to A_m, so that the actual redundancy is larger than that predicted by theory, particularly for very low redundancies.

Therefore, for a given redundancy, the pairing should be applied only to variables having a large enough variance. Given N variables, only the L variables with largest variances should be paired. The remaining $N-L$ small variables can be simply split among the two descriptions. In the event that they are lost due to channel failure, they can be simply estimated by their mean values. The estimation errors for these small variables are on the same order as the quantization error, or D_0.

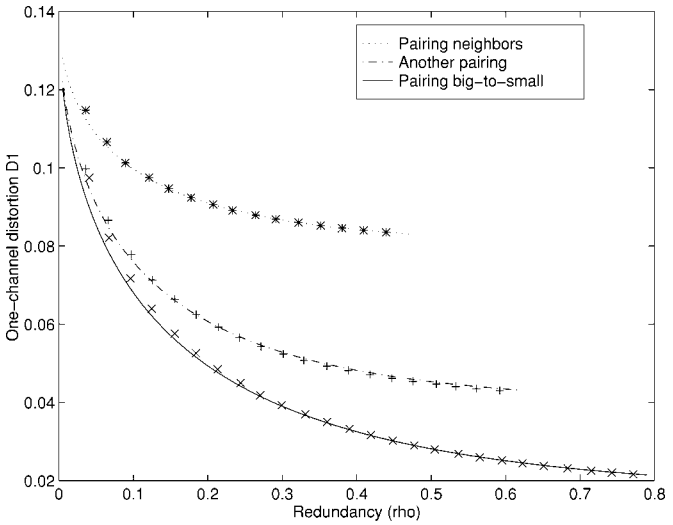

Fig. 7. Redundancy rate-distortion for six-variable pairing combinations (optimal transform).

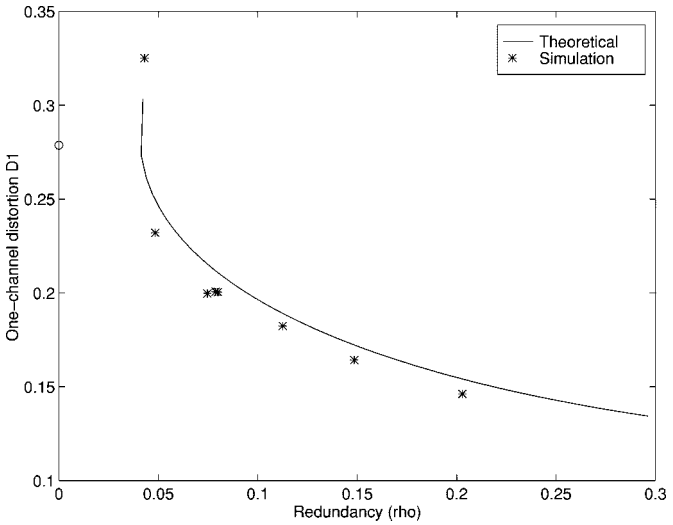

Fig. 8. Redundancy rate-distortion for coarse quantization.

To understand how to choose the number of variables, L, to pair for a given redundancy, we first examine the impact of coarse quantization on the RRD function. For a given pair, the quantization increases the single-channel distortion as described before in (32). However, the actual redundancy is also affected by the quantization. For a given stepsize Q, the redundancy in (11) should be changed to

$$\rho = \frac{1}{4}\log_2\left(\frac{(\sigma_C^2+\sigma_{qc}^2)(\sigma_D^2+\sigma_{qd}^2)}{\left(\sigma_A^2+\frac{Q^2}{12}\right)\left(\sigma_B^2+\frac{Q^2}{12}\right)}\right). \quad (48)$$

Recall that the minimal redundancy is achieved with $\theta_1 = \arctan(r)$. Substituting this transform parameter into the first equation in (5) and assuming $\sigma_{qd}^2 = \sigma_{qc}^2$, we obtain the following minimum achievable redundancy using coarse quantization:

$$\rho_{coarse} = \frac{1}{4}\log_2\left(\frac{\left(\sigma_A\sigma_B+\left(1+r+\frac{1}{2r}\right)\frac{Q^2}{12}\right)^2}{\left(\sigma_A^2+\frac{Q^2}{12}\right)\left(\sigma_B^2+\frac{Q^2}{12}\right)}\right).$$

The effect of coarse quantization is illustrated in Fig. 8, where $\sigma_A^2 = 1$, $\sigma_B^2 = 0.1$, and $Q = 0.3$. Here, the solid line is calculated using (48) for ρ and using (32) for D_1, the stars are simulation points. The circle indicates the distortion $((\sigma_A^2+\sigma_B^2)/4)$ achievable with zero redundancy, which is the average single channel distortion if one codes A and B directly in the two descriptions, respectively.

Thus, we can draw two observations as to whether to apply MDTC to a given pair or not. First, it is clear that any time $D_1 > (\sigma_A^2+\sigma_B^2)/4$, it will be more efficient not to apply MDTC to the variables A and B. Second, when the desired ρ is less than ρ_{coarse}, it is also clearly more efficient to send A and B rather than C and D.

In general, however, it is difficult to determine an optimal algorithm for when to omit a given variable from the pairings. In the next section, we use the following simple algorithm. We choose a multiplier α, and pair all variables whose variance $\sigma_i^2 > \alpha(Q^2/12)$. For very small redundancies, $\alpha = 24$ works well, while for larger redundancies, $\alpha = 4$ is better.

VI. Image Coding Using MDTC

In this section, we examine the integration of our multiple description transform coding (MDTC) algorithm with a typical image coder. We begin by describing our image coder, then present simulation results for two test images.

A. Our MDTC Coder

The coder is based on a modification of a standard JPEG coder [18]. In our baseline encoder, each block of 8×8 samples are transformed using a 8×8-pt discrete cosine transform (DCT). These coefficients are ordered to have decreasing variance. Then for a preset value of maximum estimation error T_{D_1}, we find the first L coefficients with $\sigma_n^2 > T_{D_1}$. The threshold is set proportional to the quantization error corresponding to the chosen step-size, Q, $T_{D_1} = t_{D_1}Q^2/12$. Then using a pairing algorithm, these L coefficients are split into $L/2$ pairs, A_m, B_m, $m = 1, 2, \ldots, L/2$. Each pair of coefficients are quantized using the given step-size to yield $\overline{A}_m$, $\overline{B}_m$. They are then transformed using the integer implementation of Section III-B, to yield $\overline{C}_m$, $\overline{D}_m$. All the $\overline{C}_m$ are then put sequentially into stream one, while all the $\overline{D}_m$ are put into stream two. For the remaining $N-L$ coefficients, after quantization, we append even indexed coefficients to stream one and append odd indexed coefficients to stream two. Each stream is then coded independently and transmitted through separate channels. We apply the run-length coding plus Huffman coding method to each stream.

If both streams are received, then the decoder performs the inverse transform using integer implementation on each pair of received samples, $\overline{C}_m$, $\overline{D}_m$ to recover the quantized indices $\overline{A}_m$, $\overline{B}_m$. An inverse quantizer is then used to recover the quantized values $\tilde{A}_m$, $\tilde{B}_m$. If only one stream is received, say the one containing $\overline{C}_m$, then inverse quantization is applied to yield $\tilde{C}_m$, and then $\hat{D}_m$ is estimated from $\tilde{C}_m$. Finally a direct inverse transform is applied $\tilde{C}_m$, $\hat{D}_m$ to recover $\tilde{A}_m$ and $\tilde{B}_m$. For

TABLE I
INDEX ASSIGNMENTS FOR REUSE INDEX $N = 2$, 3 AND 5. THE QUANTIZER (FOR A UNIFORM THRESHOLD QUANTIZER) BIN INDEX IS l, THE CHANNEL 1 LABEL IS i_N AND THE CHANNEL 2 LABEL IS j_N FOR REUSE INDEX N. IN ALL CASES $i_N(-l) = j_N(l)$ AND $j_N(-l) = i_N(l)$

l	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
i_2	0	0	1	1	2	2	3	3	4	4	5	5	6	6	7	7	8	8	9
j_2	0	1	1	2	2	3	3	4	4	5	5	6	6	7	7	8	8	9	9
i_3	0	0	1	1	2	1	2	2	3	3	4	3	4	4	5	5	6	5	6
j_3	0	1	0	1	1	2	2	3	2	3	3	4	4	5	4	5	5	6	6
i_5	0	0	1	0	2	1	2	1	3	1	2	2	3	2	4	3	4	3	5
j_5	0	1	0	2	0	1	1	2	1	3	2	3	2	4	2	3	3	4	3

the remaining coefficients which are split by the even-odd rule, the missing coefficients are simply replaced with their means, which are assumed to be zero. Finally, all the recovered DCT coefficients are inverse-transformed to produce the samples in the image domain.

For a given set of L coefficients, we always pair the kth largest coefficient with the $(L-k)$th largest one, as suggested by the theoretical analysis and numerical results presented in Section V-B. To determine the transform parameter for each pair, we perform optimal redundancy allocation according to Section V. In our implementation, we specify a slope parameter, λ, and determine the redundancy for each pair ρ_m by requiring the slope of the RRD function associated with that pair to be equal to λ. Ideally, we should find the operational RRD functions for coding DCT coefficient pairs. In our implementation, we assume these coefficients follow the Gaussian distribution, and use the strategy in Section V (44) for deriving ρ_m. With these ρ_m, we can then determine the transform parameter $\theta_{1,m}$ using (26). By varying λ, we achieve different redundancies.

A difficulty in applying the above coder to real images is that images are not statistically stationary. Statistical properties vary significantly among different images, and among different blocks of the same image. Because the linear estimation coefficients are determined based on coefficient variances, a mismatch between the actual variances and the assumed ones could lead to a large estimation error. To overcome this problem, prior to encoding an image we classify its blocks into different classes such that the transform coefficients of blocks in the same class have similar statistical properties. Within each class, the coefficient variances are calculated, ordered, and used as the basis for optimal redundancy allocation. Finally, the transform and estimation parameters are determined for different pairs.

In our present simulation, we consider four classes which essentially correspond to blocks that are smooth, with horizontal edges, with vertical edges, and the rest. The classification is accomplished by examining the coefficient concentration pattern. If the energy of the DC coefficient exceeds a certain percentage (specified as a threshold, T_c) of the total coefficient energy, then the block is classified as smooth. Otherwise, if the energy along the first column (or row) of the DCT block exceeds the preset threshold, the block is classified as with horizontal (or vertical) edges. A threshold of 95% has yielded quite good classification results for several test images.

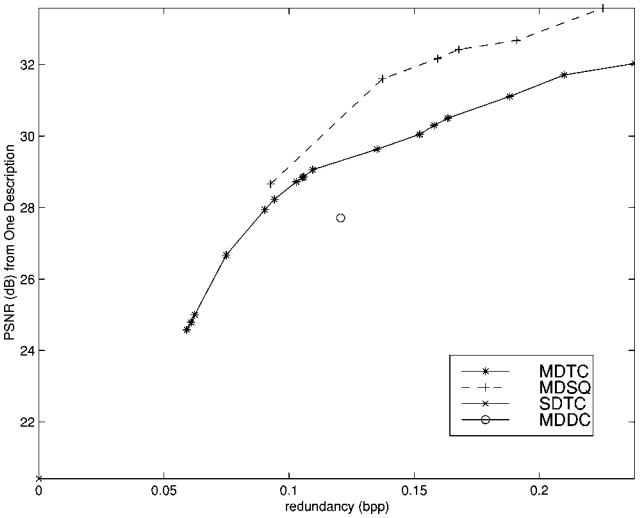

Fig. 9. RRD performance of different coders for image *Lena*. $R^* = 0.63$ bpp, $D_0 = 35.78$ dB.

The DCT coefficients are first scaled by a normalization matrix designed based on visual sensitivities to different frequencies. (We used the one recommended in the JPEG standard [18]). These scaled coefficients are then quantized using a fixed quantization interval, denoted by QP (referred to as the quality factor in the JPEG standard). The coefficients are ordered in decreasing order of the variances prior to applying the PCT.

B. Comparison Coders

We use two different base-lines to evaluate the trade-off between coding efficiency and single-channel reconstruction obtained by the proposed MDTC coder. The first one, called SDTC, is essentially the JPEG coder, in which all the DCT coefficients in a block are coded together using the run-length plus Huffman coding method. In order to split the resulting bit stream into two symmetric descriptions, we put the bits from even indexed blocks into stream one and those from odd indexed blocks into stream two. Two separate Huffman tables are designed, one for each description. In the decoder, if only one stream (e.g., the stream containing even blocks) is received, the blocks in the missing stream (e.g., the odd blocks) are simply reconstructed by assigning each pixel the mean value of the block on the left. As expected, the reconstructed image

Fig. 10. Image reconstruction results. (a) Reconstructed from both descriptions ($R^* = 0.60$ bpp, $D_0 = 35.78$ dB); (b) MDTC, from a single description, $\rho = 0.088$ bpp (15%), $D_1 = 27.94$ dB; (c) MDTC, from a single description, $\rho = 0.133$ bpp (22%), $D_1 = 29.63$ dB; and (d) MDSQ, from a single description, $\rho = 0.090$ bpp (15%), $D_1 = 28.63$ dB.

in the presence of only one stream is very poor, with a fairly strong checker-board artifact.

To obtain a better visual reconstruction, we also consider a coder in which the DC coefficients are duplicated and sent on both channels. The remaining coefficients are split so that even coefficients are put in description one, and odd coefficients in description two. This coder is referred to as MDDC. This coder has a lower coding efficiency than SDTC, not only because the DC coefficient is duplicated, but also because coefficient splitting can break some of the zero run-lengths.

We also implemented the multiple description scalar quantization (MDSQ) method proposed by Vaishampayan [4]. The MDSQ system uses the standard DCT and applies MDSQ to each transform coefficient. MDSQ's are parameterized by a quantity known as the reuse index, N, which is the number of times a particular label is used.[4] In the simulation results presented here, we have used index assignments with a reuse index of 2, 3, and 5 which are illustrated in Table I. Low reuse index corresponds to high redundancy and high reuse index corresponds to low redundancy. In each case, the results obtained for the lowest redundancy point correspond to using $N = 2$ for

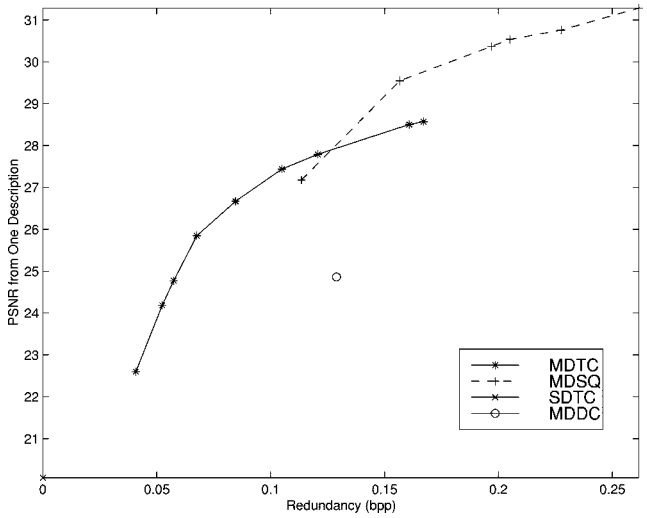

Fig. 11. RRD performance of different coders for image *Horse*. $R^* = 0.070$ bpp, $D_0 - 33.51$ dB.

[4]These assignments have been modified slightly, so the reuse index is eventually equal to the stated quantity.

Fig. 12. Image reconstruction results. (a) Reconstructed from both descriptions ($R^* = 0.70$ bpp, $D_0^* = 33.51$ dB); (b) MDTC, from a single description, $\rho = 0.085$ bpp (12%), $D_1 = 26.74$ dB; (c) MDTC, from a single description, $\rho = 0.121$ bpp (17%), $D_1 = 27.81$ dB; (d) and MDSQ, from a single description, $\rho = 0.110$ bpp (16%), $D_1 = 27.09$ dB.

the DC coefficient and $N = 5$ for all other coefficients. The point with the highest redundancy corresponds to using $N = 2$ for all transform coefficients.

For all four coders, SDTC, MDDC, MDTC, and MDSQ, two sets of Huffman tables are designed based on the actual image being coded, one for the coefficients of each description. The bits required to send the Huffman tables are ignored in all the cases because we are interested in the relative performance of these coders, not the absolute bit count. But for the MDTC coder, we do count the extra bits required for specifying the class index of each block in both descriptions. This bit rate is estimated from the entropy of the class index.

For all the coders compared, the first coefficient in each stream is treated as the DC component. However, we do not apply DC-prediction, to prevent error propagation. The bit rate required by the SDTC coder is used as the reference rate R^*. Redundancy is defined as the additional rate (in bits/pixel or bpp) required by the other coders. Not using DC-prediction has the effect of making R^* and consequently the total rate R higher than a coder using DC-prediction. Because the main purpose of our simulations is to compare the redundancy required by different coders under the same R^* and D_0, not using DC-prediction in all the coders is a reasonable choice.

C. Simulation Results

We have applied the above coders to the common test image *Lena*, with size 512×512. We fix the quantization step-size at $QP = 1$, which yields a two-channel distortion of $D_0 = 35.78$ dB for all coders. The bit rate of SDTC is $R^* = 0.60$ bpp. The RRD curve obtained by varying λ in the MDTC coder is shown in Fig. 9. On the same figure, we also give the points corresponding to SDTC (zero redundancy point), MDDC and MDSQ coders. With MDDC, only a single point of redundancy can be achieved. At this redundancy, our coder outperforms the MDDC coder significantly. Compared to MDSQ, MDTC can reach a lower redundancy range, but it does not perform as well as MDSQ in terms of PSNR in the high redundancy range, for this image.

Fig. 10 shows the reconstructed images from a single description by the MTDC coder at two different redundancy values, together with a reconstructed image using the MDSQ method. For comparison purposes, the image reconstructed from both descriptions is also presented. Our MDTC coder can produce a good quality single-channel image at a redundancy of 15%. At higher redundancies, we can obtain very good quality single-channel images. Comparing Fig. 10(b) with Fig. 10(d), we see

that at similar redundancies, MDTC and MDSQ have different visual artifacts.

We also applied the above coders to the image *Horse*. Using a quantization step-size of QP=1, the two-channel distortion is $D_0 = 33.51$ dB for all coders. The bit rate of SDTC is $R^* = 0.70$ bpp. The RRD curve obtained by varying λ in the MDTC coder is shown in Fig. 11, along with the points corresponding to SDTC (with zero redundancy), MDDC, and MDSQ coders. Again, MDTC outperforms MDDC substantially. At similar redundancies, the MDTC coder [Fig. 12(c)] yields a visually more pleasing image than MDSQ [Fig. 12(d)].

VII. Summary and Concluding Remarks

We have presented a new MDC scheme using the transform coding framework. It introduces correlation between pairs of transform coefficients via a pairwise correlating transform. We derived the optimal transform family and analyzed its RRD performance. We also considered the problem of redundancy allocation among multiple pairs and showed what is the best strategy to pair a given set of coefficients. Finally, we incorporated this MDC scheme in a JPEG-like image coder and demonstrated that with about 15% redundancy, a low but acceptable quality image can be reconstructed from a single description, and with about 20% redundancy, good to very good quality can be obtained.

The characteristics of super-exponential decay of the proposed MDC scheme at the small redundancy regime makes this coder very attractive for practical applications: a very small amount of redundancy can yield a noticeable improvement in quality. On the other hand, slower than exponential decay of the RRD curve in the large redundancy region indicates that the coder performs poorly asymptotically. To overcome this problem, in [11], we presented a generalized MDTC scheme, which maintains the super-exponential decay character of the transform-based scheme at small redundancies, and converges to exponential decay at large redundancies.

The MDTC framework introduced here can also be applied for coding video signals into multiple descriptions. Our initial approach [19] is to apply our transform-based MDC scheme to the prediction errors in video coders using motion-compensated temporal prediction, while considering the mismatch between the prediction frame available at the encoder and that at the decoder when only a single description is available.

References

[1] J. K. Wolf, A. Wyner, and J. Ziv, "Source coding for multiple descriptions," *Bell Syst. Tech. J.*, vol. 59, pp. 1417–1426, Oct. 1980.

[2] L. Ozarow, "On a source coding problem with two channels and three receivers," *Bell Syst. Tech. J.*, vol. 59, p. 1921, Dec. 1980.

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

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

[5] J.-C. Batllo and V. Vaishampayan, "Asymptotic performance of multiple description transform codes," *IEEE Trans. Inform. Theory*, vol. 43, pp. 703–707, Mar. 1997.

[6] S. D. Servetto, K. Ramchandran, V. Vaishampayan, and K. Nahrstedt, "Multiple-description wavelet based image coding," in *Proc. Int. Conf. Image Processing '98*, Chicago, IL, Oct. 1998.

[7] A. Ingle and V. A. Vaishampayan, "DPCM system design for diversity systems with applications to packetized speech," *IEEE Trans. Speech Audio Processing*, vol. 3, pp. 48–57, Jan. 1995.

[8] H. Jafarkhani and V. Tarokh, "Multiple description trellis coded quantization," in *Proc. Int. Conf. Image Processing '98*, Chicago, IL, Oct. 1998.

[9] Y. Wang, M. Orchard, and A. Reibman, "Multiple description image coding for noisy channels by pairing transform coefficients," in *Proc. IEEE 1997 1st Workshop Multimedia Signal Processing (MMSP'97)*, Princeton, NJ, June 1997, pp. 419–424.

[10] M. Orchard, Y. Wang, V. Vaishampayan, and A. Reibman, "Redundancy rate distortion analysis of multiple description image coding using pairwise correlating transforms," in *Proc. IEEE Int. Conf. Image Processing (ICIP'97)*, vol. I, Santa Barbara, CA, Oct. 1997, pp. 608–611.

[11] Y. Wang, M. Orchard, and A. Reibman, "Optimal pairwise correlating transforms for multiple description coding," in *Proc. IEEE Int. Conf. Image Processing (ICIP'98)*, Chicago, IL, Oct. 1998.

[12] V. K. Goyal and J. Kovacevic, "Optimal multiple description transform coding of Gaussian vectors," in *Proc. Data Compression Conf. (DCC'98)*, Snowbird, UT, Mar. 1998, pp. 388–397.

[13] V. K. Goyal, J. Kovacevic, R. Arean, and M. Vetterli, "Multiple description transform coding of images," in *Proc. IEEE Int. Conf. Image Processing (ICIP'98)*, Chicago, IL, Oct. 1998.

[14] A. R. Calderbank, I. Daubechies, W. Sweldens, and B.-L. Yeo, "Wavelet transforms that map integers to integers," *J. Appl. Comput. Harmon. Anal.*, to be published.

[15] X. Li, B. Tao, and M. T. Orchard, "On implementing transforms from integers to integers," in *Proc. Int. Conf. Image Processing '98*, Chicago, IL, Oct. 1998.

[16] A. Gersho and R. Gray, *Vector Quantization and Signal Compression*. Norwell, MA: Kluwer, 1992.

[17] J.-Y. Huang and P. M. Schultheiss, "Block quantization of correlated Gaussian random variables," *IEEEE Trans. Commun.*, vol. CS-11, pp. 289–296, Sept. 1963.

[18] W. B. Pennebaker and J. L. Mitchell, *JPEG: Still Image Data Compression Standard*. New York: Van Nostrand Reinhold, 1993.

[19] A. Reibman, H. Jafarkhani, Y. Wang, M. Orchard, and R. Puri, "Multiple description coding for video using motion compensated prediction," in *Proc. Int. Conf. Image Processing (ICIP'99)*, Kobe, Japan, Oct. 1999.

Yao Wang (M'90–SM'98) received the B.S. and M.S. degrees in electrical engineering from Tsinghua University, Beijing, China, and the Ph.D degree in electrical engineering from the University of California, Santa Barbara.

Since 1990, she has been on the faculty of Polytechnic University, Brooklyn, NY, and is presently Professor of electrical engineering. Since 1992, she has been a Consultant with AT&T Bell Laboratories (now AT&T Labs Research), Florham Park, NJ. She was on sabbatical at Princeton University, Princeton, NJ, in 1998 and was a Visiting Professor at the University of Erlangen, Germany, in the Summer of 1998. Her current research interests include image and video compression for unreliable networks, motion estimation and object-oriented video coding, signal processing using multimodal information, and image reconstruction problems in medical imaging.

Dr. Wang is a member of several IEEE technical committees and has served on the organizing/technical committees of several international conferences and workshops. She won the New York City Mayor's Award for Excellence in Science and Technology in the young investigator category in 2000. She has served as an Associate Editor for the IEEE Transactions on Multimedia and the IEEE Tranactions on Circuits and Systems for Video Technology.

Michael T. Orchard (F'00) was born in Shanghai, China. He received the B.S. and M.S. degrees in electrical engineering from San Diego State University, San Diego, CA, in 1980 and 1986, and the M.A. and Ph.D. degrees in electrical engineering from Princeton University, Princeton, NJ, in 1988 and 1990, respectively.

He was with the Government Products Division of Scientific Atlanta, Atlanta, GA, from 1982 to 1986, developing passive sonar DSP applications, and he has consulted with the Visual Communications Department, AT&T Bell Laboratories, since 1988. From 1990 to 1995, he was an Assistant Professor with the Department of Electrical and Computer Engineering, University of Illinois, Urbana-Champaign, where he served as Associate Director of the Image Laboratory, Beckman Institute. Since 1995, he has been an Associate Professor with the Electrical Engineering Department, Princeton University. During the 2000–2001 academic year, he is serving as Texas Instruments Visiting Professor at Rice University, Houston, TX.

Dr. Orchard received the National Science Foundation Young Investigator Award in 1993, the Army Research Office Young Investigator Award in 1996, and was elected IEEE Fellow in 2000 for "contributions to the theory and development of image and video compression algorithms."

Vinay Vaishampayan (S'87–M'89) received the B.S. degree in 1981 from the Indian Institute of Technology, Delhi, the M.S. degree in 1986 and the Ph.D. degree in 1989, both from the University of Maryland, Baltimore.

He is with AT&T Labs Research, Florham Park, NJ. His research interests are in signal processing and communications. Previously, he was an Associate Professor with the Electrical Engineering Department, Texas A&M University, College Station.

Amy R. Reibman (S'83–M'87) received the B.S., M.S., and Ph.D. degrees in electrical engineering from Duke University, Durham, NC, in 1983, 1984, and 1987, respectively.

From 1988 to 1991, she was an Assistant Professor with the Department of Electrical Engineering, Princeton University, Princeton, NJ. In 1991, she joined AT&T Bell Laboratories and became a Distinguished Member of Technical Staff in 1995. She is currently a Technology Consultant with the Wireless and Signal Processing Research Department, AT&T Laboratories. Her research interests include video compression systems for transport over packet and wireless networks and watermarking and security for video transport.

Dr. Reibman won the IEEE Communications Society Leonard G. Abraham Prize Paper Award in 1998. She was the Technical Program Chair for the 6th International Workshop on Packet Video in 1994 and Technical Co-chair for the First IEEE Workshop on Multimedia Signal Processing in 1997, and is the Technical Co-chair of the IEEE International Conference on Image Processing in 2002. She is also a member of the IEEE Signal Processing Society's Technical Committee on Image and Multidimensional Signal Processing.