

# Transforms for High-Rate Distributed Source Coding

David Rebollo-Monedero, Anne Aaron and Bernd Girod  
Information Systems Laboratory, Dept. of Electrical Eng.  
Stanford University, Stanford, CA 94305  
{drebollo,amaaron,bgirod}@stanford.edu

**Abstract**—We extend high-rate quantization theory to distributed source coding for the case in which the rate is the conditional entropy of the quantization index given the side information. This theory is applied to orthonormal block transforms for distributed source coding. A formula for the optimal rate allocation and an approximation to the optimal transform are derived. We implement a transform-domain Wyner-Ziv video coder that encodes frames independently but decodes them conditionally. Experimental results show a rate-distortion improvement with respect to the pixel-domain coder by using the discrete cosine transform.

## I. INTRODUCTION

Rate-distortion theory for distributed source coding [1]–[3] suggests that the efficiency of coding systems in which both encoder and decoder have access to some side information is close to the case in which the side information is only available to the decoder. One of the many applications of this principle is reducing the complexity of video encoders by eliminating motion compensation, and decoding using past frames as side information, while keeping the efficiency close to that of motion-compensated encoding [4] [5]. To make such systems practical, it seems crucial to extend the building blocks of traditional source coding, such as lossless coding, quantization and transform coding, to distributed source coding.

Distributed lossless coding schemes that perform close to the Slepian-Wolf bound [1] have been proposed (e.g., [6]). As for quantization for distributed source coding, optimal design of quantizers has been analyzed [7] and extended to ideal Slepian-Wolf coding [8]. However, the high-resolution theory of quantization (e.g., [9]) has not yet been extended to the distributed case. In [10], the Karhunen-Loève Transform (KLT) for distributed source coding is investigated, but it is assumed that the covariance matrix of the source vector given the side information does not depend on the values of the side information, and the study is not in the context of a practical coding scheme with quantizers for distributed source coding.

In this paper, we study high-rate quantizers and linear transform coding with side information at the decoder, assuming that a lossless coding system is available with efficiency close to an ideal Slepian-Wolf coder.

Section II contains the theoretical results for high-rate quantization, and Section III, for transforms of the source data and the side information. Experimental results of a video compression scheme using transforms for distributed coding are shown in Section IV.

Throughout the paper, we follow the convention of using uppercase letters for random vectors, and lowercase letters

for particular values they take on. Bold letters are used to emphasize that a variable is a vector or a matrix. Covariance matrices are denoted by  $\Sigma$ . For instance, if  $\mathbf{X}$  and  $\mathbf{Y}$  are random vectors,  $\Sigma_{\mathbf{X}\mathbf{Y}} = \text{Cov}[\mathbf{X}, \mathbf{Y}]$ . The conditional covariance is the matrix function  $\Sigma_{\mathbf{X}|\mathbf{Y}}(\mathbf{y}) = \text{Cov}[\mathbf{X}|\mathbf{Y} = \mathbf{y}]$ .

## II. HIGH-RATE QUANTIZATION FOR DISTRIBUTED SOURCE CODING

We study the properties of high-rate quantizers for the distributed source coding setting in Fig. 1. The source data to

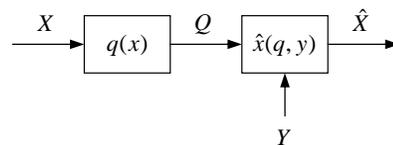


Fig. 1. Quantizer for distributed source coding.

be quantized is modelled by a continuous random vector  $\mathbf{X}$  of finite dimension  $n$ . Let the quantization function  $q(\mathbf{x})$  map the source data into the quantization index  $Q$ . A random vector  $\mathbf{Y}$  plays the role of side information available only at the receiver. The side information and the quantization index are used jointly to estimate the source data. Let  $\hat{\mathbf{X}}$  represent this estimate, obtained with the reconstruction function  $\hat{\mathbf{x}}(q, \mathbf{y})$ .

Mean-squared error is used as a distortion measure, thus the expected distortion per sample is  $\mathcal{D} = \frac{1}{n} \mathbb{E}[\|\mathbf{X} - \hat{\mathbf{X}}\|^2]$ . The formulation in this work assumes that the coding of the index  $Q$  with side information  $\mathbf{Y}$  is carried out by an ideal Slepian-Wolf coder. The expected rate per sample is defined accordingly as  $\mathcal{R} = \frac{1}{n} \mathbb{H}(Q|\mathbf{Y})$  [8]. We emphasize that the quantizer only has access to the source data, not to the side information. However, the joint statistics of  $\mathbf{X}$  and  $\mathbf{Y}$  are assumed to be known, and exploited in the design of  $q(\mathbf{x})$  and  $\hat{\mathbf{x}}(q, \mathbf{y})$ . We consider the problem of characterizing the quantization and reconstruction functions that minimize the expected Lagrangian cost  $\mathcal{J} = \mathcal{D} + \lambda \mathcal{R}$ , with  $\lambda$  a nonnegative real number, for high rate  $\mathcal{R}$ .

The theoretical results are presented in Theorem 1. The hypotheses of the theorem are believed to hold if the Bennett assumptions [11] [12] apply to the conditional probability density function (PDF)  $f_{\mathbf{X}|\mathbf{Y}}(\mathbf{x}|\mathbf{y})$  for each value of the side information  $\mathbf{y}$ , and if Gershó's conjecture [13] is true (known to be the case for  $n = 1$ ), among other technical conditions,

mentioned in [9]. For a rigorous treatment of high-rate theory that does not rely on Gersho's conjecture, see [14] [15].

*Theorem 1:* Let  $M_n$  be the minimum normalized moment of inertia of the convex polytopes tessellating  $\mathbb{R}^n$  (e. g.,  $M_1 = 1/12$ ). Suppose that for each value  $\mathbf{y}$  in the support set of  $\mathbf{Y}$ , the statistics of  $\mathbf{X}$  given  $\mathbf{Y} = \mathbf{y}$  are such that the differential entropy  $h(\mathbf{X}|\mathbf{Y} = \mathbf{y})$  is defined and finite. Suppose further that for each  $\mathbf{y}$ , there exists an asymptotically optimal entropy-constrained lattice quantizer of  $\mathbf{x}$ ,  $q(\mathbf{x}|\mathbf{y})$ , with rate  $\mathcal{R}_{\mathbf{X}|\mathbf{Y}}(\mathbf{y})$  and distortion  $\mathcal{D}_{\mathbf{X}|\mathbf{Y}}(\mathbf{y})$ , with no two cells assigned to the same index and with cell volume  $V(\mathbf{y}) > 0$ , which satisfies, for large  $\mathcal{R}_{\mathbf{X}|\mathbf{Y}}(\mathbf{y})$ ,

$$\mathcal{D}_{\mathbf{X}|\mathbf{Y}}(\mathbf{y}) \simeq M_n V(\mathbf{y})^{\frac{2}{n}}, \quad (1)$$

$$\mathcal{R}_{\mathbf{X}|\mathbf{Y}}(\mathbf{y}) \simeq \frac{1}{n} (h(\mathbf{X}|\mathbf{Y} = \mathbf{y}) - \log_2 V(\mathbf{y})), \quad (2)$$

$$\mathcal{D}_{\mathbf{X}|\mathbf{Y}}(\mathbf{y}) \simeq M_n 2^{\frac{2}{n}h(\mathbf{X}|\mathbf{Y}=\mathbf{y})} 2^{-2\mathcal{R}_{\mathbf{X}|\mathbf{Y}}(\mathbf{y})}. \quad (3)$$

Then, there exists an asymptotically optimal quantizer  $q(\mathbf{x})$  for large  $\mathcal{R}$  (precisely, minimizing  $\mathcal{J}$  as  $\lambda \rightarrow 0^+$ ) for the distributed source coding setting considered such that:

- 1)  $q(\mathbf{x})$  is a lattice quantizer with minimum moment of inertia  $M_n$  and cell volume  $V$ .
- 2) No two cells of the partition defined by  $q(\mathbf{x})$  need to be mapped into the same quantization index.
- 3) The rate and the distortion satisfy

$$\mathcal{D} \simeq M_n V^{\frac{2}{n}}, \quad (4)$$

$$\mathcal{R} \simeq \frac{1}{n} (h(\mathbf{X}|\mathbf{Y}) - \log_2 V), \quad (5)$$

$$\mathcal{D} \simeq M_n 2^{\frac{2}{n}h(\mathbf{X}|\mathbf{Y})} 2^{-2\mathcal{R}}. \quad (6)$$

*Proof:* The proof uses the quantization setting in Fig. 2, which we shall refer to as conditional quantizer, along with an argument of optimal rate allocation for the family of quantizers  $q(\mathbf{x}|\mathbf{y})$ . In this case, the side information  $\mathbf{Y}$  is available to

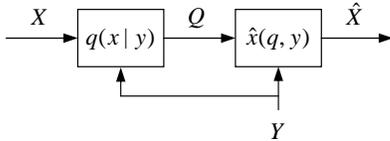


Fig. 2. Conditional quantizer.

the sender, and the design of the quantization function  $q(\mathbf{x}|\mathbf{y})$  on  $\mathbf{x}$  for each value  $\mathbf{y}$  is a non-distributed entropy-constrained quantization problem. More precisely, for all  $\mathbf{y}$  define

$$\mathcal{D}_{\mathbf{X}|\mathbf{Y}}(\mathbf{y}) = \frac{1}{n} \mathbb{E}[\|\mathbf{X} - \hat{\mathbf{X}}\|^2 | \mathbf{Y} = \mathbf{y}],$$

$$\mathcal{R}_{\mathbf{X}|\mathbf{Y}}(\mathbf{y}) = \frac{1}{n} \mathbb{H}(Q | \mathbf{Y} = \mathbf{y}),$$

$$\mathcal{J}_{\mathbf{X}|\mathbf{Y}}(\mathbf{y}) = \mathcal{D}_{\mathbf{X}|\mathbf{Y}}(\mathbf{y}) + \lambda \mathcal{R}_{\mathbf{X}|\mathbf{Y}}(\mathbf{y}).$$

By iterated expectation,  $\mathcal{D} = \mathbb{E}[\mathcal{D}_{\mathbf{X}|\mathbf{Y}}(\mathbf{Y})]$  and  $\mathcal{R} = \mathbb{E}[\mathcal{R}_{\mathbf{X}|\mathbf{Y}}(\mathbf{Y})]$ , thus the overall cost satisfies  $\mathcal{J} = \mathbb{E}[\mathcal{J}_{\mathbf{X}|\mathbf{Y}}(\mathbf{Y})]$ . As a consequence, a family of quantizers  $q(\mathbf{x}|\mathbf{y})$  minimizing  $\mathcal{J}_{\mathbf{X}|\mathbf{Y}}(\mathbf{y})$  for each  $\mathbf{y}$  also minimizes  $\mathcal{J}$ .

Since  $\mathcal{J}_{\mathbf{X}|\mathbf{Y}}(\mathbf{y})$  is a convex function of  $\mathcal{R}_{\mathbf{X}|\mathbf{Y}}(\mathbf{y})$  for all  $\mathbf{y}$ , it has a global minimum where its derivative vanishes, or

equivalently, at  $\mathcal{R}_{\mathbf{X}|\mathbf{Y}}(\mathbf{y})$  such that  $\lambda \simeq 2 \ln 2 \mathcal{D}_{\mathbf{X}|\mathbf{Y}}(\mathbf{y})$ . Suppose that  $\lambda$  is small enough for  $\mathcal{R}_{\mathbf{X}|\mathbf{Y}}(\mathbf{y})$  to be large and the approximations (1)-(3) to hold, for each  $\mathbf{y}$ . Then, all quantizers  $q(\mathbf{x}|\mathbf{y})$  introduce the same distortion (proportional to  $\lambda$ ) and consequently have a common cell volume  $V(\mathbf{y}) \simeq V$ . This, together with the fact that  $\mathbb{E}_{\mathbf{Y}}[h(\mathbf{X}|\mathbf{Y} = \mathbf{Y})] = h(\mathbf{X}|\mathbf{Y})$ , implies (4)-(6). Provided that a translation of the partition defined by  $q(\mathbf{x}|\mathbf{y})$  affects neither the distortion nor the rate, all lattice quantizers  $q(\mathbf{x}|\mathbf{y})$  can be set to be (approximately) the same, which we denote by  $q(\mathbf{x})$ . Since none of the quantizers  $q(\mathbf{x}|\mathbf{y})$  maps two cells into the same indices, neither does  $q(\mathbf{x})$ . Now, since  $q(\mathbf{x})$  is asymptotically optimal for the conditional quantizer and does not depend on  $\mathbf{y}$ , it is also optimal for the distributed quantizer in Fig. 1. ■

Equation (6) means that, asymptotically, there is no loss in performance by not having access to the side information in the quantization.

*Corollary 2:* Under the hypotheses of Theorem 1, asymptotically, there is an quantizer that leads to no loss in performance by ignoring the side information in the reconstruction.

*Proof:* Since index repetition is not required, the distortion (4) would be asymptotically the same if the reconstruction  $\hat{\mathbf{x}}(q, \mathbf{y})$  were of the form  $\hat{\mathbf{x}}(q) = \mathbb{E}[\mathbf{X}|Q = q]$ . ■

*Corollary 3:* Let  $\mathbf{X}$  and  $\mathbf{Y}$  be jointly Gaussian random vectors. Then,  $\Sigma_{\mathbf{X}|\mathbf{Y}}$  is constant with  $\mathbf{y}$ , and for large  $\mathcal{R}$ ,

$$\mathcal{D} \simeq M_n 2\pi e (\det \Sigma_{\mathbf{X}|\mathbf{Y}})^{\frac{1}{n}} 2^{-2\mathcal{R}} \xrightarrow{n \rightarrow \infty} (\det \Sigma_{\mathbf{X}|\mathbf{Y}})^{\frac{1}{n}} 2^{-2\mathcal{R}}.$$

*Proof:* Use  $h(\mathbf{X}|\mathbf{Y}) = \frac{1}{2} \log_2 ((2\pi e)^n \det \Sigma_{\mathbf{X}|\mathbf{Y}})$ , and  $M_n \rightarrow 1/2\pi e$ , together with Theorem 1. ■

### III. TRANSFORMS FOR DISTRIBUTED SOURCE CODING

#### A. Transformation of the Source Data

The following intermediate definitions and results will be useful to analyze transforms for distributed source coding. Define the geometric expectation of a positive random variable  $S$  as  $G[S] = b^{\mathbb{E}[\log_b S]}$  for any  $b > 1$ . Note that if  $S$  were discrete with probability mass function  $p_S(s)$ , then  $G[S] = \prod_s s^{p_S(s)}$ . The constant factor in the rate-distortion approximation (3) can be expressed as

$$M_n 2^{\frac{2}{n}h(\mathbf{X}|\mathbf{Y}=\mathbf{y})} = \epsilon_{\mathbf{X}|\mathbf{Y}}^2(\mathbf{y}) (\det \Sigma_{\mathbf{X}|\mathbf{Y}}(\mathbf{y}))^{1/n},$$

where  $\epsilon_{\mathbf{X}|\mathbf{Y}}^2(\mathbf{y})$  depends only on  $M_n$  and  $f_{\mathbf{X}|\mathbf{Y}}(\mathbf{x}|\mathbf{y})$ , normalized with covariance identity. If  $h(\mathbf{X}|\mathbf{Y} = \mathbf{y})$  is finite then  $\epsilon_{\mathbf{X}|\mathbf{Y}}^2(\mathbf{y}) (\det \Sigma_{\mathbf{X}|\mathbf{Y}}(\mathbf{y}))^{1/n} > 0$ , and (6) is equivalent to

$$\mathcal{D} \simeq G[\epsilon_{\mathbf{X}|\mathbf{Y}}^2(\mathbf{Y})] G[(\det \Sigma_{\mathbf{X}|\mathbf{Y}}(\mathbf{Y}))^{1/n}] 2^{-2\mathcal{R}}.$$

We are now ready to consider the transform coding setting in Fig. 3. Let  $\mathbf{X} = (X_1, \dots, X_n)$  be a continuous random vector of finite dimension  $n$  modelling source data, and  $\mathbf{Y}$  a random vector playing the role of side information available at the decoder (of dimension possibly different from  $n$ ). The source data undergo an orthogonal transformation represented by the matrix  $U$ , precisely,  $\mathbf{X}' = U^T \mathbf{X}$ . Each transformed component  $X'_i$  is coded individually with a scalar quantizer for asymmetric distributed source coding (represented in Fig. 1).

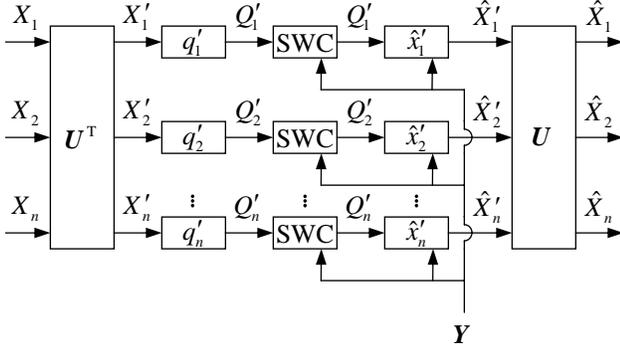


Fig. 3. Transformation of the source vector.

The quantization index is assumed to be coded with an ideal Slepian-Wolf coder, abbreviated as SWC in Fig. 3. The entire vector  $\mathbf{Y}$  is used for Slepian-Wolf decoding and reconstruction to obtain the transformed estimate  $\hat{\mathbf{X}}'$ , which is inversely transformed to recover an estimate of the original source vector according to  $\hat{\mathbf{X}} = \mathbf{U} \hat{\mathbf{X}}'$ .

The expected distortion in band  $i$  is  $\mathcal{D}_i = \mathbb{E}[(X'_i - \hat{X}'_i)^2]$ . The rate required to code the quantization index  $Q'_i$  is  $\mathcal{R}_i = \mathbb{H}(Q'_i|\mathbf{Y})$ . Define the total expected distortion per sample as  $\mathcal{D} = \frac{1}{n} \mathbb{E}[\|\mathbf{X} - \hat{\mathbf{X}}\|^2]$ , and the total expected rate per sample as  $\mathcal{R} = \frac{1}{n} \sum_i \mathcal{R}_i$ . We wish to minimize the Lagrangian cost  $\mathcal{J} = \mathcal{D} + \lambda \mathcal{R}$ .

Define the expected conditional covariance  $\bar{\Sigma}_{\mathbf{X}|\mathbf{Y}} = \mathbb{E}_{\mathbf{Y}}[\Sigma_{\mathbf{X}|\mathbf{Y}}(\mathbf{Y})]$ . Note that  $\bar{\Sigma}_{\mathbf{X}|\mathbf{Y}}$  is the covariance of the error of the best non-linear estimate of  $\mathbf{X}$  given  $\mathbf{Y}$ , i.e.,  $\mathbb{E}[\mathbf{X}|\mathbf{Y}]$ . If  $\mathbb{E}[\mathbf{X}|\mathbf{y}]$  were constant with  $\mathbf{y}$ , then  $\bar{\Sigma}_{\mathbf{X}|\mathbf{Y}} = \Sigma_{\mathbf{X}}$ .

*Theorem 4:* Assume  $R_i$  large so that the results for high-rate approximation of Theorem 1 can be applied to each band in Fig. 3, i.e.,

$$\mathcal{D}_i \simeq \frac{1}{12} 2^{h(X'_i|\mathbf{Y})} 2^{-2R_i}. \quad (7)$$

Suppose further that the change of the shape of the PDF of the transformed components with the choice of  $\mathbf{U}$  is negligible so that  $\prod_i \mathbb{G}[\epsilon_{X'_i|\mathbf{Y}}^2(\mathbf{Y})]$  can be considered constant, and that  $\text{Var}[\sigma_{X'_i|\mathbf{Y}}^2(\mathbf{Y})] \simeq 0$ , which means that the variance of the conditional distribution does not change significantly with the side information.

Then, minimization of the overall Lagrangian cost  $\mathcal{J}$  is achieved when the following conditions hold:

- 1) All bands have a common distortion  $\mathcal{D}$ . All quantizers are uniform, without index repetition, and with a common interval width  $\Delta$  such that  $\mathcal{D} \simeq \frac{1}{12} \Delta^2$ .
- 2)

$$\mathcal{D} \simeq \frac{1}{12} 2^{2\frac{1}{n} \sum_i h(X'_i|\mathbf{Y})} 2^{-2\mathcal{R}}. \quad (8)$$

- 3) An optimal choice of  $\mathbf{U}$  is that diagonalizing  $\bar{\Sigma}_{\mathbf{X}|\mathbf{Y}}$ , i.e., it is the KLT for the expected conditional covariance matrix.
- 4) The transform coding gain  $G_T$ , which we define as the inverse of the relative decrease of distortion due to the

transformation, satisfies

$$G_T \simeq \frac{\prod_i \mathbb{G}[\sigma_{X_i|\mathbf{Y}}^2(\mathbf{Y})]^{1/n}}{\prod_i \mathbb{G}[\sigma_{X'_i|\mathbf{Y}}^2(\mathbf{Y})]^{1/n}} \simeq \frac{\prod_i \mathbb{G}[\sigma_{X_i|\mathbf{Y}}^2(\mathbf{Y})]^{1/n}}{(\det \bar{\Sigma}_{\mathbf{X}|\mathbf{Y}})^{1/n}}.$$

*Proof:* Since  $\mathbf{U}$  is orthogonal,  $\mathcal{D} = \frac{1}{n} \sum \mathcal{D}_i$ . The minimization of the overall Lagrangian cost

$$\mathcal{J} = \frac{1}{n} \sum_i \mathcal{D}_i + \lambda \mathcal{R}_i$$

yields a common distortion condition,  $\mathcal{D}_i = \mathcal{D}$  (proportional to  $\lambda$ ). Equation (7) is equivalent to

$$\mathcal{D}_i \simeq \mathbb{G}[\epsilon_{X'_i|\mathbf{Y}}^2(\mathbf{Y})] \mathbb{G}[\sigma_{X'_i|\mathbf{Y}}^2(\mathbf{Y})] 2^{-2R_i}.$$

Since  $\mathcal{D}_i \simeq \mathcal{D}$  for all  $i$ , then  $\mathcal{D} = \prod_i \mathcal{D}_i^{1/n}$  and

$$\mathcal{D} \simeq \prod_i \mathbb{G}[\epsilon_{X'_i|\mathbf{Y}}^2(\mathbf{Y})]^{1/n} \prod_i \mathbb{G}[\sigma_{X'_i|\mathbf{Y}}^2(\mathbf{Y})]^{1/n} 2^{-2\mathcal{R}}, \quad (9)$$

which is equivalent to (8). The fact that all quantizers are uniform and the interval width satisfies  $\mathcal{D} = \frac{1}{12} \Delta^2$  is a consequence of Theorem 1 for one dimension.

For any positive random variable  $S$  such that  $\text{Var}[S] \simeq 0$ , it can be shown that  $\mathbb{G}[S] \lesssim \mathbb{E}[S]$ . It is assumed in the theorem that  $\text{Var}[\sigma_{X'_i|\mathbf{Y}}^2(\mathbf{Y})] \simeq 0$ , hence

$$\mathbb{G}[\sigma_{X'_i|\mathbf{Y}}^2(\mathbf{Y})] \simeq \mathbb{E}[\sigma_{X'_i|\mathbf{Y}}^2(\mathbf{Y})].$$

This, together with the assumption that  $\prod_i \mathbb{G}[\epsilon_{X'_i|\mathbf{Y}}^2(\mathbf{Y})]$  can be considered constant, implies that the choice of  $\mathbf{U}$  that minimizes the distortion (9) is approximately equal to that minimizing  $\prod_i \mathbb{E}[\sigma_{X'_i|\mathbf{Y}}^2(\mathbf{Y})]$ .

$\bar{\Sigma}_{\mathbf{X}|\mathbf{Y}}$  is nonnegative definite. The spectral decomposition theorem implies that there exists an orthogonal matrix  $\bar{\mathbf{U}}_{\mathbf{X}|\mathbf{Y}}$  and a nonnegative definite diagonal matrix  $\bar{\Lambda}_{\mathbf{X}|\mathbf{Y}}$  such that  $\bar{\Sigma}_{\mathbf{X}|\mathbf{Y}} = \bar{\mathbf{U}}_{\mathbf{X}|\mathbf{Y}} \bar{\Lambda}_{\mathbf{X}|\mathbf{Y}} \bar{\mathbf{U}}_{\mathbf{X}|\mathbf{Y}}^T$ . On the other hand,

$$\forall \mathbf{y} \quad \Sigma_{\mathbf{X}'|\mathbf{Y}}(\mathbf{y}) = \mathbf{U}^T \Sigma_{\mathbf{X}|\mathbf{Y}}(\mathbf{y}) \mathbf{U} \Rightarrow \bar{\Sigma}_{\mathbf{X}'|\mathbf{Y}} = \mathbf{U}^T \bar{\Sigma}_{\mathbf{X}|\mathbf{Y}} \mathbf{U},$$

where a notation analogous to that of  $\mathbf{X}$  is used for  $\mathbf{X}'$ .

Finally, from Hadamard's inequality and the fact that  $\mathbf{U}$  is orthogonal, it follows that

$$\prod_i \mathbb{E}[\sigma_{X'_i|\mathbf{Y}}^2(\mathbf{Y})] \geq \det \bar{\Sigma}_{\mathbf{X}'|\mathbf{Y}} = \det \bar{\Sigma}_{\mathbf{X}|\mathbf{Y}}.$$

Since  $\mathbf{U} = \bar{\mathbf{U}}_{\mathbf{X}|\mathbf{Y}}$  implies that  $\bar{\Sigma}_{\mathbf{X}'|\mathbf{Y}} = \bar{\Lambda}_{\mathbf{X}|\mathbf{Y}}$ , we conclude that the distortion is minimized precisely for that choice of  $\mathbf{U}$ . The expression for the transform coding gain follows immediately. ■

*Corollary 5:* If  $\mathbf{X}$  and  $\mathbf{Y}$  are jointly Gaussian, then it is only necessary to assume the high-rate approximation hypothesis of Theorem 4, in order for it to hold. Furthermore, if  $\mathcal{D}_{VQ}$  and  $\mathcal{R}_{VQ}$  denote the distortion and the rate when an optimal vector quantizer is used, then we have:

- 1)  $\bar{\Sigma}_{\mathbf{X}|\mathbf{Y}} = \Sigma_{\mathbf{X}} - \Sigma_{\mathbf{X}\mathbf{Y}} \Sigma_{\mathbf{Y}}^{-1} \Sigma_{\mathbf{Y}\mathbf{X}}^T$ .
- 2)  $h(\mathbf{X}|\mathbf{Y}) = \sum_i h(X'_i|\mathbf{Y})$ .
- 3)  $\frac{\mathcal{D}}{\mathcal{D}_{VQ}} \simeq \frac{1/12}{M_n} \xrightarrow[n \rightarrow \infty]{} \frac{\pi e}{6} \simeq 1.53$  dB.
- 4)  $\mathcal{R} - \mathcal{R}_{VQ} \simeq \frac{1}{2} \log_2 \frac{1/12}{M_n} \xrightarrow[n \rightarrow \infty]{} \frac{1}{2} \log_2 \frac{\pi e}{6} \simeq 0.25$  b/s.

*Proof:* Conditionals of Gaussian random vectors are Gaussian, and linear transformations preserve Gaussianity, thus  $\prod_i G[\epsilon_{X'_i|Y}^2(\mathbf{Y})]$ , which depends only on the type of PDF, is constant with  $\mathbf{U}$ . Furthermore,

$$\Sigma_{\mathbf{X}|\mathbf{Y}}(\mathbf{y}) = \Sigma_{\mathbf{X}} - \Sigma_{\mathbf{X}\mathbf{Y}} \Sigma_{\mathbf{Y}}^{-1} \Sigma_{\mathbf{X}\mathbf{Y}}^T,$$

constant with  $\mathbf{y}$ , hence  $\text{Var}[\sigma_{X'_i|Y}^2(\mathbf{Y})] = 0$ . The differential entropy identity follows from the fact that for Gaussian random vectors (conditional) independence is equivalent to (conditional) uncorrelation, and that this is verified for each  $\mathbf{y}$ . To complete the proof, apply Corollary 3. ■

As an additional example with  $\text{Var}[\sigma_{X'_i|Y}^2(\mathbf{Y})] = 0$ , consider  $\mathbf{X} = \mathbf{Y} + \mathbf{N}$ , and assume  $\mathbf{N}$  and  $\mathbf{Y}$  are independent.  $\Sigma_{\mathbf{X}'|\mathbf{Y}}(\mathbf{y}) = \mathbf{U}^T \Sigma_{\mathbf{N}} \mathbf{U}$ , constant with  $\mathbf{y}$ .

*Corollary 6:* Suppose that for each  $\mathbf{y}$ ,  $\Sigma_{\mathbf{X}|\mathbf{Y}}(\mathbf{y})$  is Toeplitz with a square summable associated autocorrelation so that it is also asymptotically circulant as  $n \rightarrow \infty$ . In terms of the associated random process, this means that  $(X_i|\{\mathbf{Y} = \mathbf{y}\})_i$  is wide-sense stationary for each  $\mathbf{y}$ . Then, it is not necessary to assume that  $\text{Var}[\sigma_{X'_i|Y}^2(\mathbf{Y})] \simeq 0$  in Theorem 4 in order for it to hold, with the following modifications for  $\mathbf{U}$  and  $G_T$ :

- 1) The Discrete Cosine Transform (DCT) is an asymptotically optimal choice for  $\mathbf{U}$ .
- 2) The transform coding gain is given by

$$G_T \simeq G[G_T(\mathbf{Y})], \quad G_T(\mathbf{Y}) = \frac{\prod_i \sigma_{X'_i|Y}^2(\mathbf{Y})^{1/n}}{(\det \Sigma_{\mathbf{X}|\mathbf{Y}}(\mathbf{Y}))^{1/n}}.$$

*Proof:* The proof goes along the same lines of that of Theorem 4, observing that the DCT matrix asymptotically diagonalizes  $\Sigma_{\mathbf{X}|\mathbf{Y}}(\mathbf{y})$  for each  $\mathbf{y}$ . ■

Observe that the coding performance of the cases considered in Corollaries 5 and 6 would be asymptotically the same if the transform  $\mathbf{U}$  were allowed to be a function of  $\mathbf{y}$ .

### B. Transformation of the Side Information

A very convenient simplification in the setting of Fig. 3 would consist of using scalars instead of vectors in each of the Slepian-Wolf coders and in the reconstruction functions. More precisely, we are interested in linear transformations  $\mathbf{Y}' = \mathbf{V}^T \mathbf{Y}$  that lead to a small loss in terms of rate and distortion.  $\mathbf{Y}$  is assumed to be a random vector of finite dimension  $p$ . It is not required for  $\mathbf{V}$  to define an injective transformation, since no inversion is needed.

*Proposition 7:* Let  $X \in \mathbb{R}$  and  $\mathbf{Y} \in \mathbb{R}^p$  be jointly Gaussian. Let  $\mathbf{c}$  be a  $p$ -dimensional vector, which gives the linear estimate  $\hat{X} = \mathbf{c}^T \mathbf{Y}$ . Then,

$$\min_{\mathbf{c}} h(X|\hat{X}) = h(X|\mathbf{Y}),$$

and the minimum is achieved for  $\mathbf{c}$  such that  $\hat{X}$  is the best linear estimate of  $X$  given  $\mathbf{Y}$  in the mean-squared error sense.

*Proof:* For each  $\mathbf{c}$ ,  $X|\{\hat{X} = \hat{x}\}$  is a Gaussian random variable with variance  $\sigma_{X|\hat{X}}^2$  equal to the variance of the error of the best linear estimate of  $X$  given  $\hat{X}$  (and also of its best affine estimate), constant with  $\hat{x}$ . Minimizing

$$h(X|\hat{X}) = \frac{1}{2} \log_2(2\pi e \sigma_{X|\hat{X}}^2)$$

is equivalent to minimizing  $\sigma_{X|\hat{X}}^2$ . Set  $\mathbf{c}^*$  according to the best estimate of  $X$  given  $\mathbf{Y}$ , and  $\sigma_{X|\hat{X}}^{*2}$  to the variance of the error. Since a linear estimate of  $X$  given  $\mathbf{c}^T \mathbf{Y}$  is also a linear estimate of  $X$  given  $\mathbf{Y}$ ,  $\sigma_{X|\hat{X}}^{*2} \leq \sigma_{X|\hat{X}}^2$ , with equality when  $\mathbf{c} = \mathbf{c}^*$ . ■

*Theorem 8:* Under the hypotheses of Corollary 5, for high rates, the transformation of the side information given by

$$\mathbf{V}^T = \mathbf{U}^T \Sigma_{\mathbf{X}\mathbf{Y}} \Sigma_{\mathbf{Y}}^{-1} \quad (10)$$

minimizes the total rate  $\mathcal{R}$ , with no performance loss in distortion or rate with respect to the transform coding setting of Fig. 3, in which the entire vector  $\mathbf{Y}$  is used for decoding and reconstruction. Precisely, reconstruction functions defined by  $E[X'_i|Q = q, \mathbf{Y} = \mathbf{y}]$  and by  $E[X'_i|Q = q, Y'_i = y'_i]$  give approximately the same distortion  $\mathcal{D}_i$ , and  $\mathcal{R}_i = H(X'_i|Y'_i) \simeq H(X'_i|\mathbf{Y})$ .

*Proof:* Theorems 1 and 4 imply

$$\mathcal{R}_i = H(X'_i|\mathbf{Y}) \simeq h(X'_i|\mathbf{Y}) - \log_2 \Delta,$$

thus the minimization of  $\mathcal{R}_i$  is approximately equivalent to the minimization of  $h(X'_i|\mathbf{Y})$ . Since linear transformations preserve Gaussianity,  $\mathbf{X}'$  and  $\mathbf{Y}$  are jointly Gaussian, and Proposition 7 applies to each  $X'_i$ .  $\mathbf{V}$  is determined by the best linear estimate of  $\mathbf{X}'$  given  $\mathbf{Y}$ . This proves that there is no loss in rate. Corollary 2 implies that a sub-optimal reconstruction is asymptotically as efficient, thus there is no loss in distortion either. ■

Observe that  $\Sigma_{\mathbf{X}\mathbf{Y}} \Sigma_{\mathbf{Y}}^{-1}$  in (10) corresponds to the best linear estimate of  $X_i$  from  $\mathbf{Y}$ . This estimate is transformed according to the same transformation applied to  $\mathbf{X}$ , yielding an estimate of  $X'_i$ .

## IV. EXPERIMENTAL RESULTS

In [5], we apply lossy distributed coding to build a low-complexity, asymmetric video compression scheme where individual frames are encoded independently (*intraframe encoding*) but decoded conditionally (*interframe decoding*). In the proposed scheme we encode the pixel values of a frame independently from other frames. At the decoder, previously reconstructed frames are used as side information and Wyner-Ziv decoding is performed by exploiting the temporal similarities between the current frame and the side information.

In the following experiments, we extend the Wyner-Ziv video codec, outlined in [5], to a transform domain Wyner-Ziv coder. The spatial transform enables the codec to exploit the statistical dependencies within a frame, thus achieving better rate-distortion performance.

For the simulations, the odd frames are designated as key frames which are encoded and decoded using a conventional intraframe codec. The even frames are Wyner-Ziv frames which are intraframe encoded but interframe decoded, adopting the simplified transform coding set-up described in Section III-B.

For encoding a Wyner-Ziv frame  $\mathbf{X}$ , we first apply a blockwise DCT to generate  $\mathbf{X}'$ . Each transform coefficient

band is then independently quantized using uniform scalar quantizers with similar step sizes across bands. Since we use fixed length codes for Slepian-Wolf coding it is not possible to have exactly the same step sizes. Rate-compatible punctured turbo codes are used for Slepian-Wolf coding each band. The parity bits produced by the turbo encoder are stored in a buffer which transmits a subset of these parity bits to the decoder upon request.

At the decoder, we take previously reconstructed frames to generate side information  $Y$  which is used in decoding  $X$ . In the first set-up (*MC-I*), we perform motion-compensated interpolation on the previous and next reconstructed key frames to generate  $Y$ . In the second scheme (*MC-E*), we produce  $Y$  through motion-compensated extrapolation using the two previous reconstructed frames (a key frame and a Wyner-Ziv frame). The DCT is applied on  $Y$  generating the different side information coefficient bands  $Y'_i$ . A bank of turbo decoders reconstruct the quantized coefficient bands independently using the corresponding  $Y'_i$  as side information. Each coefficient band is then reconstructed as the best estimate given the reconstructed symbols and the side information. More details of the proposed scheme and extended results can be found in [16].

The compression results for the first 100 frames of *Mother and Daughter* are shown in Fig. 4. For the plots, we only include the rate and distortion of the luminance of the even frames. The even frame rate is 15 frames per second. We compare our results to (i) DCT-based intraframe coding (the even frames are encoded as I frames) and (ii) H.263+ interframe coding with an I-B-I-B predictive structure, counting only the rate and PSNR of the B frames. We also plot the compression results of the pixel-domain Wyner-Ziv codec.

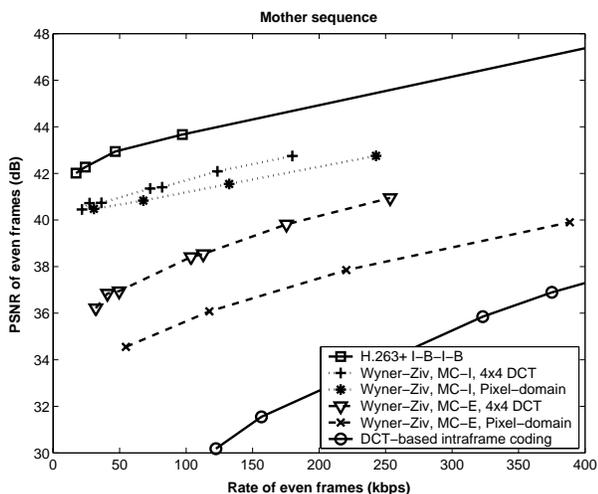


Fig. 4. Rate and PSNR comparison of Wyner-Ziv codec vs. DCT-based intraframe coding and H.263+ I-B-I-B coding. *Mother and Daughter* sequence.

As it can be observed from the plots, when the side information is highly reliable, such as when MC-I is used, the transform-domain codec is only 0.5 dB better than the pixel-domain Wyner-Ziv codec. With less reliable MC-E, using a

transform before encoding results in a 2 to 2.5 dB improvement. Compared to conventional DCT-based intraframe coding the Wyner-Ziv transform codec is about 10 to 12 dB (with MC-I) and 7 to 9 dB (with MC-E) better. The gap from H.263+ interframe coding is 2 dB for MC-I and about 5 dB for MC-E. The proposed system allows low-complexity encoding while coming close to the compression efficiency of interframe video coders.

## V. CONCLUSIONS

If ideal Slepian-Wolf coders are used, lattice quantizers without index repetition are asymptotically optimal. It is known [3] that the rate loss in the Wyner-Ziv problem for smooth continuous sources and quadratic distortion vanishes as  $\mathcal{D} \rightarrow 0$ . Our work shows that this is true also for the operational rate loss and for each finite dimension  $n$ .

The theoretical study of transforms shows that (under certain conditions) the KLT of the source vector is determined by its expected conditional covariance given the side information, which is approximated by the DCT for conditionally stationary processes. Experimental results confirm that the use of the DCT may lead to important performance improvements.

## REFERENCES

- [1] J. D. Slepian and J. K. Wolf, "Noiseless coding of correlated information sources," *IEEE Trans. Inform. Theory*, vol. IT-19, pp. 471–480, July 1973.
- [2] A. D. Wyner and J. Ziv, "The rate-distortion function for source coding with side information at the decoder," *IEEE Trans. Inform. Theory*, vol. IT-22, no. 1, pp. 1–10, Jan. 1976.
- [3] R. Zamir, "The rate loss in the Wyner-Ziv problem," *IEEE Trans. Inform. Theory*, vol. 42, no. 6, pp. 2073–2084, Nov. 1996.
- [4] R. Puri and K. Ramchandran, "PRISM: A new robust video coding architecture based on distributed compression principles," in *Proc. 40th Allerton Conf. Commun., Contr., Comput.*, Allerton, IL, Oct. 2002.
- [5] A. Aaron, R. Zhang, and B. Girod, "Wyner-Ziv coding of motion video," in *Proc. Asilomar Conf. Signals, Syst., Pacific Grove, CA, Nov. 2002*.
- [6] A. Aaron and B. Girod, "Compression with side information using turbo codes," in *Proc. IEEE Data Compression Conf. (DCC)*, Snowbird, UT, Apr. 2002, pp. 252–261.
- [7] M. Fleming, Q. Zhao, and M. Effros, "Network vector quantization," *IEEE Trans. Inform. Theory*, 2001, submitted.
- [8] D. Rebollo-Monedero, R. Zhang, and B. Girod, "Design of optimal quantizers for distributed source coding," in *Proc. IEEE Data Compression Conf. (DCC)*, Snowbird, UT, Mar. 2003, pp. 13–22.
- [9] R. M. Gray and D. L. Neuhoff, "Quantization," *IEEE Trans. Inform. Theory*, vol. 44, pp. 2325–2383, Oct. 1998.
- [10] M. Gastpar, P. Dragotti, and M. Vetterli, "The distributed, partial, and conditional Karhunen-Loève transforms," in *Proc. IEEE Data Compression Conf. (DCC)*, Snowbird, UT, Mar. 2003, pp. 283–292.
- [11] W. R. Bennett, "Spectra of quantized signals," *Bell Syst., Tech. J.* 27, July 1948.
- [12] S. Na and D. L. Neuhoff, "Bennett's integral for vector quantizers," *IEEE Trans. Inform. Theory*, vol. 41, pp. 886–900, July 1995.
- [13] A. Gersho, "Asymptotically optimal block quantization," *IEEE Trans. Inform. Theory*, vol. IT-25, pp. 373–380, July 1979.
- [14] P. L. Zador, "Topics in the asymptotic quantization of continuous random variables," *Bell Lab., Tech. Memo.*, 1966.
- [15] R. Gray, T. Linder, and J. Li, "A Lagrangian formulation of Zador's entropy-constrained quantization theorem," *IEEE Trans. Inform. Theory*, vol. 48, no. 3, pp. 695–707, Mar. 2002.
- [16] A. Aaron, S. Rane, E. Setton, and B. Girod, "Transform-domain Wyner-Ziv codec for video," in *Proc. SPIE Visual Commun., Image Processing (VCIP)*, San Jose, CA, Jan. 2004, to appear.