High-Rate Quantization and Transform Coding with Side Information at the Decoder*

David Rebollo-Monedero†, Shantanu Rane, Anne Aaron and Bernd Girod

Information Systems Lab., Dept. of Electrical Engineering
Stanford University, Stanford, CA 94305, USA

Abstract

We extend high-rate quantization theory to Wyner-Ziv coding, i.e., lossy source coding with side information at the decoder. Ideal Slepian-Wolf coders are assumed, thus rates are conditional entropies of quantization indices given the side information. This theory is applied to the analysis of orthonormal block transforms for Wyner-Ziv coding. A formula for the optimal rate allocation and an approximation to the optimal transform are derived. The case of noisy high-rate quantization and transform coding is included in our study, in which a noisy observation of source data is available at the encoder, but we are interested in estimating the unseen data at the decoder, with the help of side information.

We implement a transform-domain Wyner-Ziv video coder that encodes frames independently but decodes them conditionally. Experimental results show that using the discrete cosine transform results in a rate-distortion improvement with respect to the pixel-domain coder. Transform coders of noisy images for different communication constraints are compared. Experimental results show that the noisy Wyner-Ziv transform coder achieves a performance close to the case in which the side information is also available at the encoder.

Keywords: high-rate quantization, transform coding, side information, Wyner-Ziv coding, distributed source coding, noisy source coding

1. Introduction

Rate-distortion theory for distributed source coding [3–6] shows that under certain conditions, the performance of coders with side information available only at the decoder is close to the case in which both encoder and decoder have access to the side information. Under much more restrictive statistical conditions, this also holds for coding of noisy observations of unseen data [7,8].

One of the many applications of this result 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 [9–11]. In addition, even if the image captured by the video encoder is corrupted by noise, we would still wish to recover the clean, unseen data at the decoder, with the help of side information, consisting of previously decoded frames, and perhaps some additional local noisy image.

In these examples, due to complexity constraints in the design of the encoder, or simply due to the unavailability of the side information at the encoder, conventional, joint denoising and coding techniques are not possible. We need practical systems for noisy source coding with decoder side information, capable of the rate-distortion performance predicted by information-theoretic studies. To this end, it is crucial to extend the

*This work was supported by NSF under Grant No. CCR-0310376. The material in this paper was presented partially at the 37th and 38th editions of the Asilomar Conference on Signals, Systems and Computers, Pacific Grove, CA, Nov. 2003 [1] and 2004 [2].

†Corresponding author. Tel.: +1-650-724-3647. E-mail address: drebollo@stanford.edu (D. Rebollo-Monedero).

building blocks of traditional source coding and denoising, such as lossless coding, quantization, transform coding and estimation, to distributed source coding.

It was shown by Slepian and Wolf [3] that lossless distributed coding can achieve the same performance as joint coding. Soon after, Wyner and Ziv [4,12] established the rate-distortion limits for lossy coding with side information at the decoder, which we shall refer to as Wyner-Ziv (WZ) coding. Later, an upper bound on the rate loss due to the unavailability of the side information at the encoder was found in [5], which also proved that for power-difference distortion measures and smooth source probability distributions, this rate loss vanishes in the limit of small distortion. A similar high-resolution result was obtained in [13] for distributed coding of several sources *without* side information, also from an information-theoretic perspective, that is, for arbitrarily large dimension. In [14] (unpublished), it was shown that tessellating quantizers followed by Slepian-Wolf coders are asymptotically optimal in the limit of small distortion *and* large dimension.

It may be concluded from the proof of the converse to the WZ rate-distortion theorem [4] that there is no asymptotic loss in performance by considering block codes of sufficiently large length, which may be seen as vector quantizers, followed by fixed-length coders. This suggests a convenient implementation of WZ coders as quantizers, possibly preceded by transforms, followed by Slepian-Wolf coders, analogously to the implementation of nondistributed coders.

Practical distributed lossless coding schemes have been proposed, adapting channel coding techniques such as turbo codes and low-density parity-check codes, which are approaching the Slepian-Wolf bound, e.g., [15–23]. See [24] for a much more exhaustive list.

The first studies on quantizers for WZ coding were based on high-dimensional nested lattices [25–27], or heuristically designed scalar quantizers [16,28], often applied to Gaussian sources, with fixed-length coding or entropy coding of the quantization indices. A different approach was followed in [29–32], where the Lloyd algorithm [33] was generalized for a variety of settings. In particular, [32] considered the important case of ideal Slepian-Wolf coding of the quantization indices, at a rate equal to the conditional entropy given the side information. In [34–36], nested lattice quantizers and trellis-coded quantizers followed by Slepian-Wolf coders were used to implement WZ coders.

The Karhunen-Loève Transform (KLT) [37–39] for distributed source coding was investigated in [40,41], but it was 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 was not in the context of a practical coding scheme with quantizers for distributed source coding. Very recently, the distributed KLT was studied in the context of compression of Gaussian source data, assuming that the transformed coefficients are coded at the information-theoretic rate-distortion performance [42,43]. Most of the recent experimental work on WZ coding uses transforms [44,1,24].

There is extensive literature on source coding of a noisy observation of an unseen source. The nondistributed case was studied in [45–47], and [7,48–50,8] analyzed the distributed case from an information-theoretic point of view. Using Gaussian statistics and Mean-Squared Error (MSE) as a distortion measure, [13] proved that distributed coding of two noisy observations *without* side information can be carried out with a performance close to that of joint coding and denoising, in the limit of small distortion *and* large dimension. Most of the operational work on distributed coding of noisy sources, that is, for a fixed dimension, deals with quantization design for a variety of settings [51–54], but does not consider the characterization of such quantizers at high rates or transforms.

A key aspect in the understanding of operational coding is undoubtedly the theoretic characterization of quantizers at high rates [55], which is also fundamental in the theoretic study of transforms for data compression [56]. In the literature reviewed at this point, the studies of high-rate distributed coding are information theoretic, thereby requiring arbitrarily large dimension, among other constraints. On the other hand, the aforementioned studies of transforms applied to

compression are valid only for Gaussian statistics and assume that the transformed coefficients are coded at the information-theoretic limit.

In this paper, we provide a theoretic characterization of high-rate WZ quantizers for a fixed dimension, assuming ideal Slepian-Wolf coding of the quantization indices, and we apply it to develop a theoretic analysis of orthonormal transforms for WZ coding. Both the case of coding of directly observed data, and the case of coding of a noisy observation of unseen data, are considered. We shall refer to these two cases as coding of clean sources and noisy sources, respectively. The material in this paper was presented partially in [1,2].

Section 2 presents a theoretic analysis of high-rate quantization of clean sources, and Section 3, of noisy sources. This analysis is applied to the study of transforms of the source data in Sections 4 and 5, also for clean and noisy sources, respectively. Section 6 analyzes the transformation of the side information itself. In Section 7, experimental results on a video compression scheme using WZ transform coding, and also on image denoising, are shown to illustrate the clean and noisy coding cases.

Throughout the paper, we follow the convention of using uppercase letters for random variables, including random scalars, vectors, or abstract random ensembles, and lowercase letters for the particular values that they take on. The measurable space in which a random variable takes values will be called alphabet. Let X be a random variable, discrete, continuous or in an arbitrary alphabet, possibly vector valued. Its probability function, if it exists, will be denoted by $p_X(x)$, whether it is a probability mass function (PMF) or a probability density function (PDF)[(a)]. For notational convenience, the covariance operator Cov and the letter Σ will be used interchangeably. For example, the conditional covariance of $X|\,y$ is the matrix function $\Sigma_{X|Y}(y) = \mathrm{Cov}[X|\,y]$.

[(a)] As a matter of fact, a PMF is a PDF with respect to the counting measure.

2. High-Rate WZ Quantization of Clean Sources

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

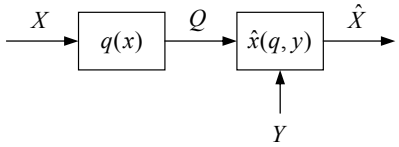


Fig. 1. WZ quantization.

data to be quantized is modeled by a continuous random vector X of finite dimension n. Let the quantization function $q(x)$ map the source data into the quantization index Q. A random variable Y, distributed in an arbitrary alphabet, discrete or continuous, 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 $\hat{X}$ represent this estimate, obtained with the reconstruction function $\hat{x}(q, y)$.

MSE is used as a distortion measure, thus the expected distortion per sample is $\mathcal{D} = \frac{1}{n}\,\mathrm{E}\,\|X - \hat{X}\|^2$. The rate for asymmetric Slepian-Wolf coding of Q given Y has been shown to be $\mathrm{H}(Q|Y)$ in the case when the two alphabets of the random variables involved are finite [3], in the sense that any rate greater than $\mathrm{H}(Q|Y)$ would allow arbitrarily low probability of decoding error, but any rate lesser than $\mathrm{H}(Q|Y)$ would not. In [57], the validity of this result has been generalized to countable alphabets, but it is still assumed that $\mathrm{H}(Y) < \infty$. We show in Appendix A that the asymmetric Slepian-Wolf result remains true under the assumptions in this paper, namely, for any Q in a countable alphabet and any Y in an arbitrary alphabet, possibly continuous, regardless of the finiteness of $\mathrm{H}(Y)$. The formulation in this work assumes that the coding of the index Q with side information Y is carried out by an *ideal* Slepian-Wolf coder, with negligible decoding error probability and rate redundancy. The expected rate per sample is defined accordingly as $\mathcal{R} = \frac{1}{n}\,\mathrm{H}(Q|\,Y)$ [32].

We emphasize that the quantizer only has access to the source data, not to the side informa-

tion. However, the joint statistics of X and Y are assumed to be known, and are exploited in the design of $q(x)$ and $\hat{x}(q, y)$. We consider the problem of characterizing the quantization and reconstruction functions that minimize the expected Lagrangian cost $\mathcal{C} = \mathcal{D} + \lambda \mathcal{R}$, with λ a nonnegative real number, for high rate $\mathcal{R}$.

The theoretic results are presented in Theorem 1. The theorem holds if the Bennett assumptions [58,59] apply to the conditional PDF $p_{X|Y}(x|\,y)$ *for each* value of the side information y, and if Gersho's conjecture [60] is true (known to be the case for $n = 1$), among other technical conditions, mentioned in [55]. For a rigorous treatment of high-rate theory that does not rely on Gersho's conjecture, see [61,62].

We shall use the term uniform tessellating quantizer in reference to quantizers whose quantization regions are possibly rotated versions of a common convex polytope, with equal volume. Lattice quantizers are, strictly speaking, a particular case. In the following results, Gersho's conjecture for nondistributed quantizers, which allows rotations, will be shown to imply that optimal WZ quantizers are also tessellating quantizers, and the uniformity of the cell volume will be proved as well[(b)]. M_n denotes the minimum normalized moment of inertia of the convex polytopes tessellating $\mathbb{R}^n$ (e.g., $M_1 = 1/12$).

Theorem 1 (High-rate WZ quantization)**.** *Suppose that* for each value y *in the alphabet of* Y*, the statistics of* X *given* $Y = y$ *are such that the conditional differential entropy* $\mathrm{h}(X|\,y)$ *exists and is finite. Suppose further that* for each y*, there exists an asymptotically optimal entropy-constrained uniform tessellating quantizer of* x*,* $q(x|\,y)$*, with rate* $\mathcal{R}_{X|Y}(y)$ *and distortion* $\mathcal{D}_{X|Y}(y)$*, with no two cells assigned to the same index and with cell volume* $V(y) > 0$*, which satisfies, for large* $\mathcal{R}_{X|Y}(y)$*,*

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

$$\mathcal{R}_{X|Y}(y) \simeq \tfrac{1}{n}\left(\mathrm{h}(X|\,y) - \log_2 V(y)\right), \tag{2}$$

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

Then, there exists an asymptotically optimal quantizer $q(x)$ *for large* $\mathcal{R}$*, for the WZ coding setting considered such that:*

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

$$\mathcal{D} \simeq M_n\, V^{\frac{2}{n}}, \tag{4}$$

$$\mathcal{R} \simeq \tfrac{1}{n}\left(\mathrm{h}(X|\,Y) - \log_2 V\right), \tag{5}$$

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

Proof: The proof uses the quantization setting in Fig. 2, which we shall refer to as a conditional quantizer, along with an argument of optimal rate allocation for $q(x|\,y)$, where $q(x|\,y)$ can be regarded as a quantizer on the values x and y taken by the source data and the side information, or a family of quantizers on x indexed by y. In this case, the side information Y is available

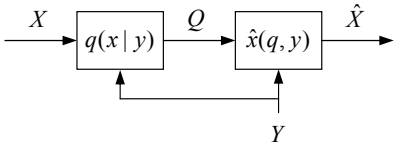


Fig. 2. Conditional quantizer.

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

$$\mathcal{D}_{X|Y}(y) = \tfrac{1}{n}\,\mathrm{E}[\|X - \hat{X}\|^2|\,y],$$
$$\mathcal{R}_{X|Y}(y) = \tfrac{1}{n}\,\mathrm{H}(Q|\,y),$$
$$\mathcal{C}_{X|Y}(y) = \mathcal{D}_{X|Y}(y) + \lambda\,\mathcal{R}_{X|Y}(y).$$

[(b)] A tessellating quantizer need not be uniform. A trivial example is a partition of the real line into intervals of different length. In 1 dimension, uniform tesselating quantizers are uniform lattice quantizers. It is easy to construct simple examples in $\mathbb{R}^2$ of uniform and nonuniform tesselating quantizers that are not lattice quantizers using rectangles. However, the optimal nondistributed, fixed-rate quantizers for dimensions 1 and 2 are known to be lattices: the Z-lattice and the hexagonal lattice respectively.

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

Since $\mathcal{C}_{X|Y}(y)$ is a convex function of $\mathcal{R}_{X|Y}(y)$ for all y, it has a global minimum where its derivative vanishes, or equivalently, at $\mathcal{R}_{X|Y}(y)$ such that $\lambda \simeq 2 \ln 2\, \mathcal{D}_{X|Y}(y)$. Suppose that λ is small enough for $\mathcal{R}_{X|Y}(y)$ to be large and for the approximations (1)-(3) to hold, for each y. Then, all quantizers $q(x|\,y)$ introduce the same distortion (proportional to λ) and consequently have a common cell volume $V(y) \simeq V$. This, together with the fact that $\mathrm{E}_Y[\mathrm{h}(X|\,y)]_{y=Y} = \mathrm{h}(X|\,Y)$, implies (4)-(6). Provided that a translation of the partition defined by $q(x|\,y)$ affects neither the distortion nor the rate, all uniform tessellating quantizers $q(x|\,y)$ may be set to be (approximately) the same, which we denote by $q(x)$. Since none of the quantizers $q(x|\,y)$ maps two cells into the same indices, neither does $q(x)$. Now, since $q(x)$ is asymptotically optimal for the conditional quantizer and does not depend on y, it is also optimal for the WZ 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 (High-rate WZ reconstruction)**.** *Under the hypotheses of Theorem 1, asymptotically, there is a 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 $\hat{x}(q, y)$ were of the form $\hat{x}(q) = \mathrm{E}[X|\,q]$. □

Corollary 3. *Let X and Y be jointly Gaussian random vectors. Then, the conditional covariance $\Sigma_{X|Y}$ does not depend on y, and for large $\mathcal{R}$,*

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

Proof: Use $\mathrm{h}(X|\,Y) = \frac{1}{2} \log_2 \big((2\pi e)^n \det \Sigma_{X|Y}\big)$, and $M_n \to \frac{1}{2\pi e}$ [63], together with Theorem 1. □

3. High-Rate WZ Quantization of Noisy Sources

In this section, we study the properties of high-rate quantizers of a noisy source with side information at the decoder, as illustrated in Fig. 3, which we shall refer to as WZ quantizers of a noisy source. A noisy observation Z of some unseen source data X is quantized at the encoder. The quantizer $q(z)$ maps the observation into a quantization index Q. The quantization index is losslessly coded, and used jointly with some side information Y, available only at the decoder, to obtain an estimate $\hat{X}$ of the unseen source data. $\hat{x}(q, y)$ denotes the reconstruction function at the decoder. X, Y and Z are random variables with known joint distribution, such that X is a continuous random vector of finite dimension n. No restrictions are imposed on the alphabets of Y and Z.

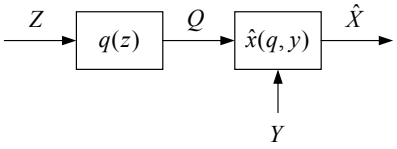


Fig. 3. WZ quantization of a noisy source.

MSE is used as a distortion measure, thus the expected distortion *per sample of the unseen source* is $\mathcal{D} = \frac{1}{n}\,\mathrm{E}\,\|X - \hat{X}\|^2$. As in the previous section, it is assumed that the coding of the index Q is carried out by an ideal Slepian-Wolf coder, at rate per sample $\mathcal{R} = \frac{1}{n}\,\mathrm{H}(Q|\,Y)$. We emphasize that the quantizer only has access to the observation, not to the source data or the side information. However, the joint statistics of X, Y and Z can be exploited in the design of $q(z)$ and $\hat{x}(q, y)$. We consider the problem of characterizing the quantizers and reconstruction functions that minimize the expected Lagrangian cost $\mathcal{C} = \mathcal{D} + \lambda\mathcal{R}$, with λ a nonnegative real number, for high rate $\mathcal{R}$. This includes the problem in the previous section as the particular case $Z = X$.

3.1. Nondistributed Case

We start by considering the simpler case of quantization of a noisy source *without* side information, depicted in Fig. 4. The following

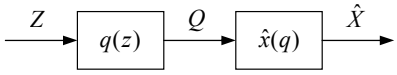


Fig. 4. Quantization of a noisy source without side information.

theorem extends the main result of [46,47] to entropy-constrained quantization, valid for any rate $\mathcal{R} = \mathrm{H}(Q)$, not necessarily high. Define $\bar{x}(z) = \mathrm{E}[X|\, z]$, the best MSE estimator of X given Z, and $\bar{X} = \bar{x}(Z)$.

Theorem 4 (MSE noisy quantization)**.** *For any nonnegative λ and any Lagrangian-cost optimal quantizer of a noisy source* without *side information (Fig. 4), there exists an implementation with the same cost in two steps:*

1. *Obtain the minimum MSE estimate $\bar{X}$.*
2. *Quantize the estimate $\bar{X}$ regarded as a clean source, using a quantizer $q(\bar{x})$ and a reconstruction function $\hat{x}(q)$, minimizing* $\mathrm{E}\,\|\bar{X} - \hat{X}\|^2 + \lambda\,\mathrm{H}(Q)$.

This is illustrated in Fig. 5. Furthermore, the total distortion per sample is

$$\mathcal{D} = \tfrac{1}{n}(\mathrm{E}\,\mathrm{tr}\,\mathrm{Cov}[X|\, Z] + \mathrm{E}\,\|\bar{X} - \hat{X}\|^2), \tag{7}$$

where the first term is the MSE of the estimation step.

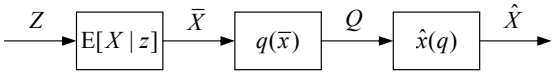


Fig. 5. Optimal implementation of MSE quantization of a noisy source without side information.

Proof: The proof is a modification of that in [47], replacing distortion by Lagrangian cost. Define the modified distortion measure $\tilde{d}(z, \hat{x}) = \mathrm{E}[\|X - \hat{x}\|^2|\, z]$. Since $X \leftrightarrow Z \leftrightarrow \hat{X}$, it is easy to show that $\mathrm{E}\,\|X - \hat{X}\|^2 = \mathrm{E}\,\tilde{d}(Z, \hat{X})$. By the orthogonality principle of linear estimation,

$$\tilde{d}(z, \hat{x}) = \mathrm{E}[\|X - \bar{x}(z)\|^2|\, z] + \|\bar{x}(z) - \hat{x}\|^2.$$

Take expectation to obtain (7). Note that the first term of (7) does not depend on the quantization design, and the second is the MSE between $\bar{X}$ and $\hat{X}$.

Let $r(q)$ be the codeword length function of a uniquely decodable code, that is, satisfying $\sum_q 2^{-r(q)} \leqslant 1$, with $\mathcal{R} = \mathrm{E}\,r(Q)$. The Lagrangian cost of the setting in Fig. 4 can be written as

$$\mathcal{C} = \tfrac{1}{n}(\mathrm{E}\,\mathrm{tr}\,\mathrm{Cov}[X|\, Z] + \inf_{\hat{x}(q), r(q)} \mathrm{E}\inf_q\{\|\bar{x}(Z) - \hat{x}(q)\|^2 + \lambda\, r(q)\}),$$

and the cost of the setting in Fig. 5 as

$$\mathcal{C} = \tfrac{1}{n}(\mathrm{E}\,\mathrm{tr}\,\mathrm{Cov}[X|\, Z] + \inf_{\hat{x}(q), r(q)} \mathrm{E}\inf_q\{\|\bar{X} - \hat{x}(q)\|^2 + \lambda\, r(q)\}),$$

which give the same result. Now, since the expected rate is minimized for the (admissible) rate measure $r(q) = -\log p_Q(q)$ and $\mathrm{E}\,r(Q) = \mathrm{H}(Q)$, both settings give the same Lagrangian cost with a rate equal to the entropy. □

Similarly to the remarks on Theorem 1, the hypotheses of the next theorem are believed to hold if the Bennett assumptions apply to the PDF $p_{\bar{X}}(\bar{x})$ of the MSE estimate, and if Gersho's conjecture is true among other technical conditions.

Theorem 5 (High-rate noisy quantization)**.** *Assume that $\mathrm{h}(\bar{X}) < \infty$ and that there exists a uniform tessellating quantizer $q(\bar{x})$ of $\bar{X}$ with cell volume V that is asymptotically optimal in Lagrangian cost at high rates. Then, there exists an asymptotically optimal quantizer $q(z)$ of a noisy source in the setting of Fig. 4 such that:*

1. *An asymptotically optimal implementation of $q(z)$ is that of Theorem 4, represented in Fig. 5, with a uniform tessellating quantizer $q(\bar{x})$ having cell volume V.*
2. *The rate and distortion per sample satisfy*

$$\mathcal{D} \simeq \tfrac{1}{n}\,\mathrm{E}\,\mathrm{tr}\,\mathrm{Cov}[X|\, Z] + M_n\, V^{\frac{2}{n}},$$
$$\mathcal{R} \simeq \tfrac{1}{n}(\mathrm{h}(\bar{X}) - \log_2 V),$$
$$\mathcal{D} \simeq \tfrac{1}{n}\,\mathrm{E}\,\mathrm{tr}\,\mathrm{Cov}[X|\, Z] + M_n\, 2^{\frac{2}{n}\,\mathrm{h}(\bar{X})}\, 2^{-2\mathcal{R}}.$$

Proof: Immediate from Theorem 4 and conventional theory of high-rate quantization of clean sources. □

3.2. Distributed Case

We are now ready to consider the WZ quantization of a noisy source in Fig. 3. Define $\bar{x}(y,z) = \mathrm{E}[X|\,y,z]$, the best MSE estimator of X given Y and Z, $\bar{X} = \bar{x}(Y,Z)$, and $\mathcal{D}_\infty = \frac{1}{n}\,\mathrm{E}\,\mathrm{tr}\,\mathrm{Cov}[X|\,Y,Z]$. The following theorem extends the results on high-rate WZ quantization in Section 2 to noisy sources. The remark on the hypotheses of Theorem 5 also applies here, where the Bennett assumptions apply instead to the conditional PDF $p_{\bar{X}|Y}(\bar{x}|\,y)$ *for each* y.

Theorem 6 (High-rate noisy WZ quantization)**.** *Suppose that the conditional expectation function* $\bar{x}(y,z)$ *is additively separable, i.e.,* $\bar{x}(y,z) = \bar{x}_Y(y) + \bar{x}_Z(z)$*, and define* $\bar{X}_Z = \bar{x}_Z(Z)$*. Suppose further that* for each value y *in the alphabet of* Y*,* $\mathrm{h}(\bar{X}|\,y) < \infty$*, and that there exists a uniform tessellating quantizer* $q(\bar{x},y)$ *of* $\bar{X}$*, with no two cells assigned to the same index and cell volume* $V(y) > 0$*, with rate* $\mathcal{R}_{\bar{X}|Y}(y)$ *and distortion* $\mathcal{D}_{\bar{X}|Y}(y)$*, such that, at high rates, it is asymptotically optimal in Lagrangian cost and*

$$\mathcal{D}_{\bar{X}|Y}(y) \simeq M_n\,V(y)^{\frac{2}{n}},$$
$$\mathcal{R}_{\bar{X}|Y}(y) \simeq \tfrac{1}{n}\left(\mathrm{h}(\bar{X}|\,y) - \log_2 V(y)\right),$$
$$\mathcal{D}_{\bar{X}|Y}(y) \simeq M_n\,2^{\frac{2}{n}\,\mathrm{h}(\bar{X}|y)}\,2^{-2\mathcal{R}_{\bar{X}|Y}(y)}.$$

Then, there exists an asymptotically optimal quantizer $q(z)$ *for large* $\mathcal{R}$*, for the WZ quantization setting represented in Fig. 3 such that:*

1. $q(z)$ *can be implemented as an estimator* $\bar{x}_Z(z)$ *followed by a uniform tessellating quantizer* $q(\bar{x}_Z)$ *with cell volume* V*.*
2. *No two cells of the partition defined by* $q(\bar{x}_Z)$ *need to be mapped into the same quantization index.*
3. *The rate and distortion per sample satisfy*

$$\mathcal{D} \simeq \mathcal{D}_\infty + M_n\,V^{\frac{2}{n}}, \tag{8}$$
$$\mathcal{R} \simeq \tfrac{1}{n}(\mathrm{h}(\bar{X}|\,Y) - \log_2 V), \tag{9}$$
$$\mathcal{D} \simeq \mathcal{D}_\infty + M_n\,2^{\frac{2}{n}\,\mathrm{h}(\bar{X}|Y)}\,2^{-2\mathcal{R}}. \tag{10}$$

4. $\mathrm{h}(\bar{X}|\,Y) = \mathrm{h}(\bar{X}_Z|\,Y)$.

Proof: The proof is similar to that for clean sources in Theorem 1 and only the differences are emphasized. First, as in the proof of WZ quantization of a clean source, a conditional quantization setting is considered, as represented in Fig. 6. An entirely analogous argument using conditional costs, as defined in the proof for clean sources, implies that the optimal conditional quantizer is an optimal conventional quantizer for each value of y. Therefore, using statistics conditioned on y

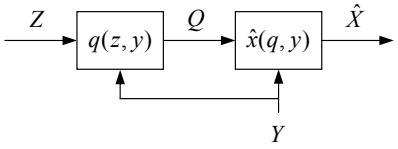


Fig. 6. Conditional quantization of a noisy source.

everywhere, by Theorem 4, the optimal conditional quantizer can be implemented as in Fig. 7, with conditional costs

$$\mathcal{D}_{X|Y}(y) \simeq \tfrac{1}{n}\,\mathrm{E}[\mathrm{tr}\,\mathrm{Cov}[X|\,y,Z]|\,y] + M_n\,V(y)^{\frac{2}{n}},$$
$$\mathcal{R}_{X|Y}(y) \simeq \tfrac{1}{n}(\mathrm{h}(\bar{X}|\,y) - \log_2 V(y)),$$
$$\mathcal{D}_{X|Y}(y) \simeq \tfrac{1}{n}\,\mathrm{E}[\mathrm{tr}\,\mathrm{Cov}[X|\,y,Z]|\,y] + M_n\,2^{\frac{2}{n}\,\mathrm{h}(\bar{X}|y)}\,2^{-2\mathcal{R}_{X|Y}(y)}.$$

The derivative of $\mathcal{C}_{X|Y}(y)$ with respect to

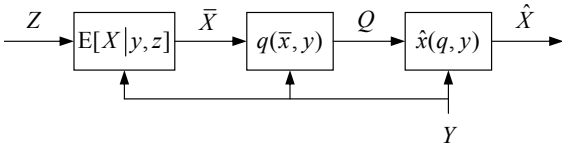


Fig. 7. Optimal implementation of MSE conditional quantization of a noisy source.

$\mathcal{R}_{X|Y}(y)$ vanishes when $\lambda \simeq 2\ln 2\,M_n\,V(y)^{\frac{2}{n}}$, which, as in the proof for clean sources, implies that all conditional quantizers have a common cell volume $V(y) \simeq V$ (however, only the second term of the distortion is constant, not the overall distortion). Taking expectation of the conditional costs proves that (8) and (9) are valid for the conditional quantizer of Fig. 7. The validity of (10) for the conditional quantizer can be shown

by solving for V in (9) and substituting the result into (8).

The assumption that $\bar{x}(y,z) = \bar{x}_Y(y) + \bar{x}_Z(z)$ means that for two values of y, y_1 and y_2, $\bar{x}(y_1,z)$ and $\bar{x}(y_2,z)$, seen as functions of z, differ only by a constant vector. Since the conditional quantizer of $\bar{X}$, $q(\bar{x}\,|\,y)$, is a uniform tessellating quantizer at high rates, a translation will neither affect the distortion nor the rate, and therefore $\bar{x}(y,z)$ can be replaced by $\bar{x}_Z(z)$ with no impact on the Lagrangian cost. In addition, since all conditional quantizers have a common cell volume, the same translation argument implies that a common unconditional quantizer $q(\bar{x}_Z)$ can be used instead, with performance given by (8)-(10), and since conditional quantizers do not reuse indices, neither does the common unconditional quantizer.

The last item of the theorem follows from the fact that $\mathrm{h}(\bar{x}_Y(y) + \bar{X}_Z\,|\,y) = \mathrm{h}(\bar{X}_Z\,|\,y)$. □

Clearly, the theorem is a generalization of Theorem 1, since $Z = X$ implies $\bar{x}(y,z) = \bar{x}_Z(z) = z$, trivially additively separable.

The case in which X can be written as $X = f(Y) + g(Z) + N$, for any (measurable) functions f, g and any random variable N with $\mathrm{E}[N\,|\,y,z]$ constant with (y,z), gives an example of additively separable estimator. This includes the case in which X, Y and Z are jointly Gaussian. Furthermore, in the Gaussian case, since $\bar{x}_Z(z)$ is an affine function and $q(\bar{x}_Z)$ is a uniform tessellating quantizer, the overall quantizer $q(\bar{x}_Z(z))$ is also a uniform tessellating quantizer, and if Y and Z are uncorrelated, then $\bar{x}_Y(y) = \mathrm{E}[X\,|\,y]$ and $\bar{x}_Z(z) = \mathrm{E}[X\,|\,z]$, but *not* in general.

Observe that, according to the theorem, if the estimator $\bar{x}(y,z)$ is additively separable, there is no asymptotic loss in performance by not using the side information at the encoder.

Corollary 7. *Assume the hypotheses of Theorem 6, and that the optimal reconstruction levels $\hat{\bar{x}}(q,y)$ for each of the conditional quantizers $q(\bar{x},y)$ are simply the centroids of the quantization cells for a uniform distribution. Then, there is a WZ quantizer $q(\bar{x}_Z)$ that leads to no asymptotic loss in performance if the reconstruction function is $\hat{x}(q,y) = \hat{\bar{x}}_Z(q) + \bar{x}_Y(y)$, where $\hat{\bar{x}}_Z(q)$ are the centroids of $q(\bar{x}_Z)$.*

Proof: In the proof of Theorem 6, $q(\bar{x}_Z)$ is a uniform tessellating quantizer without index repetition, a translated copy of $q(\bar{x},y)$. □

Theorem 6 and Corollary 7 show that the WZ quantization setting of Fig. 3 can be implemented as depicted in Fig. 8, where $\hat{\bar{x}}_Z(q,y)$ can be made independent from y without asymptotic loss in performance, so that the pair $q(\bar{x}_Z)$, $\hat{\bar{x}}_Z(q)$ form a uniform tessellating quantizer and reconstructor for $\bar{X}_Z$.

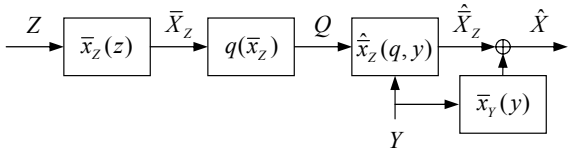


Fig. 8. Asymptotically optimal implementation of MSE WZ quantization of a noisy source with additively separable $\bar{x}(y,z)$.

Finally, if $\bar{x}_Z(z)$ is a bijective vector field, then, under mild conditions, including continuous differentiability of $\bar{x}_Z(z)$ and its inverse, it can be shown that $\mathrm{h}(\bar{X}\,|\,Y)$ in Theorem 6 satisfies

$$\mathrm{h}(\bar{X}\,|\,Y) = \mathrm{h}(Z\,|\,Y) + \mathrm{E}\log_2 \left|\det \tfrac{\mathrm{d}\bar{x}_Z}{\mathrm{d}z}(Z)\right|,$$

where $\mathrm{d}\bar{x}_Z(z)/\mathrm{d}z$ denotes the Jacobian matrix of $\bar{x}_Z(z)$.

4. WZ Transform Coding of Clean Sources

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

$$M_n\, 2^{\frac{2}{n}\mathrm{h}(X|Y)} = \epsilon^2_{X|Y}(y)\, \left(\det \Sigma_{X|Y}(y)\right)^{1/n},$$

where $\epsilon^2_{X|Y}(y)$ depends only on M_n and $p_{X|Y}(x\,|\,y)$, normalized with covariance identity. If $\mathrm{h}(X\,|\,y)$ is finite, then

$$\epsilon^2_{X|Y}(y)\, \left(\det \Sigma_{X|Y}(y)\right)^{1/n} > 0,$$

and since $\mathrm{G}_Y[2^{\frac{2}{n}[\mathrm{h}(X|y)]_{y=Y}}] = 2^{\frac{2}{n}\,\mathrm{h}(X|Y)}$, (6) is equivalent to

$$\mathcal{D} \simeq \mathrm{G}[\epsilon^2_{X|Y}(Y)]\ \mathrm{G}[\big(\det \Sigma_{X|Y}(Y)\big)^{1/n}]\, 2^{-2\mathcal{R}}.$$

We are now ready to consider the transform coding setting in Fig. 9. Let $X = (X_1, \ldots, X_n)$

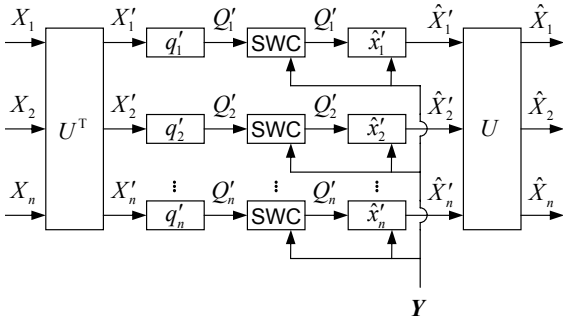


Fig. 9. Transformation of the source vector.

be a continuous random vector of finite dimension n, modeling source data, and let Y be an arbitrary random variable playing the role of side information available at the decoder, for instance, a random vector of dimension possibly different from n. The source data undergo an orthogonal transform represented by the matrix U, precisely, $X' = U^{\mathrm{T}}\, X$. Each transformed component X'_i is coded individually with a scalar WZ quantizer (represented in Fig. 1). The quantization index is assumed to be coded with an ideal Slepian-Wolf coder, abbreviated as SWC in Fig. 9. The (entire) side information Y is used for Slepian-Wolf decoding and reconstruction to obtain the transformed estimate $\hat{X}'$, which is inversely transformed to recover an estimate of the original source vector according to $\hat{X} = U\, \hat{X}'$.

The expected distortion in subband i is $\mathcal{D}_i = \mathrm{E}\,(X'_i - \hat{X}'_i)^2$. The rate required to code the quantization index Q'_i is $\mathcal{R}_i = \mathrm{H}(Q'_i|\, Y)$. Define the total expected distortion per sample as $\mathcal{D} = \frac{1}{n}\ \mathrm{E}\ \|X - \hat{X}\|^2$, 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 $\mathcal{C} = \mathcal{D} + \lambda\, \mathcal{R}$.

Define the expected conditional covariance $\bar{\Sigma}_{X|Y} = \mathrm{E}\,\Sigma_{X|Y}(Y) = \mathrm{E}_Y \mathrm{Cov}[X|\, Y]$. Note that $\bar{\Sigma}_{X|Y}$ is the covariance of the error of the best estimate of X given Y, i.e., $\mathrm{E}[X|\, Y]$. In fact, the orthogonality principle of conditional estimation implies

$$\bar{\Sigma}_{X|Y} + \mathrm{Cov}\,\mathrm{E}[X|\, Y] = \mathrm{Cov}\, X,$$

thus $\bar{\Sigma}_{X|Y} \preccurlyeq \mathrm{Cov}\, X$, with equality if and only if $\mathrm{E}[X|\, Y]$ is a constant with probability 1.

Theorem 8 (WZ transform coding)**.** *Assume R_i large so that the results for high-rate approximation of Theorem 1 can be applied to each subband in Fig. 9, i.e.,*

$$\mathcal{D}_i \simeq\ \tfrac{1}{12}\, 2^{2\,\mathrm{h}(X'_i|Y)}\, 2^{-2R_i}. \tag{11}$$

Suppose further that the change of the shape of the PDF of the transformed components with the choice of U is negligible so that $\prod_i \mathrm{G}\,\epsilon^2_{X'_i|Y}(Y)$ may be considered constant, and that $\mathrm{Var}\,\sigma^2_{X'_i|Y}(Y) \simeq 0$, which means that the variance of the conditional distribution does not change significantly with the side information.

Then, minimization of the overall Lagrangian cost $\mathcal{C}$ 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 Δ such that $\mathcal{D} \simeq \frac{1}{12}\Delta^2$.*

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

3. *An optimal choice of U is one that diagonalizes $\bar{\Sigma}_{X|Y}$, that is, it is the KLT for the expected conditional covariance matrix.*

4. *The transform coding gain δ_{T}, which we define as the inverse of the relative decrease of distortion due to the transform, satisfies*

$$\delta_{\mathrm{T}} \simeq \frac{\prod_i \mathrm{G}[\sigma^2_{X_i|Y}(Y)]^{1/n}}{\prod_i \mathrm{G}[\sigma^2_{X'_i|Y}(Y)]^{1/n}} \simeq \simeq \frac{\prod_i \mathrm{G}[\sigma^2_{X_i|Y}(Y)]^{1/n}}{\big(\det \bar{\Sigma}_{X|Y}\big)^{1/n}}.$$

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

$$\mathcal{C} = \tfrac{1}{n}\sum_i \mathcal{D}_i + \lambda\,\mathcal{R}_i$$

yields a common distortion condition, $\mathcal{D}_i \simeq \mathcal{D}$ (proportional to λ). Equation (11) is equivalent to

$$\mathcal{D}_i \simeq \mathrm{G}[\epsilon^2_{X_i'|Y}(Y)]\;\mathrm{G}[\sigma^2_{X_i'|Y}(Y)]\,2^{-2R_i}.$$

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

$$\mathcal{D} \simeq \prod_i \mathrm{G}[\epsilon^2_{X_i'|Y}(Y)]^{1/n}\cdot \cdot\prod_i \mathrm{G}[\sigma^2_{X_i'|Y}(Y)]^{1/n}\,2^{-2\mathcal{R}}, \quad (12)$$

which is equivalent to Item 2 in the statement of the theorem. 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 scalar S such that $\operatorname{Var} S \simeq 0$, it can be shown that $\mathrm{G}\,S \lesssim \mathrm{E}\,S$. It is assumed in the theorem that $\operatorname{Var}\sigma^2_{X_i'|Y}(Y) \simeq 0$, hence

$$\mathrm{G}\,\sigma^2_{X_i'|Y}(Y) \simeq \mathrm{E}\,\sigma^2_{X_i'|Y}(Y).$$

This, together with the assumption that $\prod_i \mathrm{G}\,\epsilon^2_{X_i'|Y}(Y)$ may be considered constant, implies that the choice of U that minimizes the distortion (12) is approximately equal to that minimizing $\prod_i \mathrm{E}\,\sigma^2_{X_i'|Y}(Y)$.

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

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

where a notation analogous to that of X is used for X'.

Finally, from Hadamard's inequality and the fact that U is orthogonal, it follows that

$$\prod_i \mathrm{E}\,\sigma^2_{X_i'|Y}(Y) \geqslant \det\bar{\Sigma}_{X'|Y} = \det\bar{\Sigma}_{X|Y}.$$

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

Corollary 9 (Gaussian case)**.** *If X and Y are jointly Gaussian random vectors, then it is only necessary to assume the high-rate approximation hypothesis of Theorem 8, in order for it to hold. Furthermore, if $\mathcal{D}_{\mathrm{VQ}}$ and $\mathcal{R}_{\mathrm{VQ}}$ denote the distortion and the rate when an optimal vector quantizer is used, then we have:*

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

Proof: Conditionals of Gaussian random vectors are Gaussian, and linear transforms preserve Gaussianity, thus $\prod_i \mathrm{G}\,\epsilon^2_{X_i'|Y}(Y)$, which depends only on the type of PDF, is constant with U. Furthermore,

$$\Sigma_{X|Y}(y) = \Sigma_X - \Sigma_{XY}\,\Sigma_Y^{-1}\,\Sigma_{XY}^{\mathrm{T}},$$

constant with y, hence $\operatorname{Var}\sigma^2_{X_i'|Y}(Y) = 0$. The differential entropy identity follows from the fact that for Gaussian random vectors (conditional) independence is equivalent to (conditional) uncorrelatedness, and that this is the case for each y. To complete the proof, apply Corollary 3. □

The conclusions and the proof of the previous corollary are equally valid if we only require that $X|\,y$ be Gaussian for every y, and $\Sigma_{X|Y}(y)$ be constant.

As an additional example with $\operatorname{Var}\sigma^2_{X_i'|Y}(Y) = 0$, consider $X = f(Y) + N$, for any (measurable) function f, and assume that N and Y are independent random vectors. $\Sigma_{X'|Y}(y) = U^{\mathrm{T}}\,\Sigma_N\,U$, constant with y. If in addition, N is Gaussian, then so is $X|\,y$.

Corollary 10 (DCT)**.** *Suppose that for each y, $\Sigma_{X|Y}(y)$ is Toeplitz with a square summable associated autocorrelation so that it is also asymptotically circulant as $n \to \infty$. In terms of the associated random process, this means that X_i is conditionally covariance stationary given Y, that is, $(X_i - \mathrm{E}[X_i|\, y]|\, y)_{i\in\mathbb{Z}}$ is second-order stationary for each y. Then, it is not necessary to assume that $\mathrm{Var}\, \sigma^2_{X_i'|Y}(Y) \simeq 0$ in Theorem 8 in order for it to hold, with the following modifications for U and δ_{T}:*

1. *The Discrete Cosine Transform (DCT) is an asymptotically optimal choice for U*[c]*.*

2. *The transform coding gain is given by*

$$\delta_{\mathrm{T}} \simeq \mathrm{G}\, \delta_{\mathrm{T}}(Y), \quad \delta_{\mathrm{T}}(Y) = \frac{\prod_i \sigma^2_{X_i|Y}(Y)^{1/n}}{\big(\det \Sigma_{X|Y}(Y)\big)^{1/n}}.$$

Proof: The proof proceeds along the same lines of that of Theorem 8, observing that the DCT matrix asymptotically diagonalizes $\Sigma_{X|Y}(y)$ *for each* y, since it is symmetric and asymptotically circulant [64, Chapter 3]. □

Observe that the coding performance of the cases considered in Corollaries 9 and 10 would be asymptotically the same if the transform U were allowed to be a function of y.

We would like to remark that there are several ways by which the transform coding gain in Item 4 of the statement of Theorem 8, and also in Item 2 of Corollary 10, can be manipulated to resemble an arithmetic-geometric mean ratio involving the variances of the transform coefficients. This is consistent with the fact that the transform coding gain is indeed a gain. The following corollary is an example.

Corollary 11. *Suppose, in addition to the hypotheses of Theorem 8, that $\sigma^2_{X_i|Y}(y) = \sigma^2_{X_0|Y}(y)$ for all $i = 1, \ldots, n$, and for all y. This can be understood as a weakened version of the conditional covariance stationarity assumption in Corollary 10. Then, the transform coding gain satisfies*

$$\delta_{\mathrm{T}} \simeq \mathrm{G}\, \delta_{\mathrm{T}}(Y), \quad \delta_{\mathrm{T}}(Y) = \frac{\frac{1}{n}\sum_i \sigma^2_{X_i'|Y}(Y)}{\prod_i \sigma^2_{X_i'|Y}(Y)^{1/n}}.$$

Proof: Define

$$\delta_{\mathrm{T}}(y) = \frac{\prod_i \sigma^2_{X_i|Y}(y)^{1/n}}{\prod_i \sigma^2_{X_i'|Y}(y)^{1/n}}.$$

According to Theorem 8, it is clear that $\delta_{\mathrm{T}} \simeq \mathrm{G}\, \delta_{\mathrm{T}}(Y)$. Now, for each y, since by assumption the conditional variances are constant with i, the numerator of $\delta_{\mathrm{T}}(y)$ satisfies

$$\prod_i \sigma^2_{X_i|Y}(y)^{1/n} = \sigma^2_{X_0|Y}(y) = \tfrac{1}{n}\sum_i \sigma^2_{X_i|Y}(y).$$

Finally, since $X' = U^{\mathrm{T}} X$ and U is orthonormal,

$$\sum_i \sigma^2_{X_i|Y}(y) = \mathrm{E}[\|X - \mathrm{E}[X|\, y]\|^2|\, y] =$$
$$= \mathrm{E}[\|X' - \mathrm{E}[X'|\, y]\|^2|\, y] = \sum_i \sigma^2_{X_i'|Y}(y). \quad \square$$

5. WZ Transform Coding of Noisy Sources

5.1. Fundamental Structure

If $\bar{x}(y, z)$ is additively separable, the asymptotically optimal implementation of a WZ quantizer established by Theorem 6 and Corollary 7, illustrated in Fig. 8, suggests the transform coding setting represented in Fig. 10. In this setting, the WZ uniform tessellating quantizer and reconstructor for $\bar{X}_Z$, regarded as a clean source, have been replaced by a WZ transform coder of clean sources, studied in Section 4. The transform coder is a rotated, scaled Z-lattice quantizer, and the translation argument used in the proof of Theorem 6 still applies. By this argument, an additively separable encoder estimator $\bar{x}(y, z)$ can be replaced by an encoder estimator $\bar{x}_Z(z)$ and a decoder estimator $\bar{x}_Y(y)$ with no loss in performance at high rates.

The transform coder acts now on $\bar{X}_Z$, which undergoes the orthonormal transformation $\bar{X}'_Z = U^{\mathrm{T}} \bar{X}_Z$. Each transformed coefficient $\bar{X}'_{Z\,i}$ is

[c] Precisely, U^{T} is the analysis DCT matrix, and U the synthesis DCT matrix.

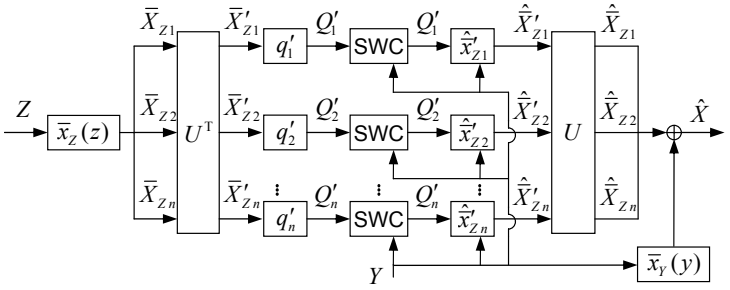


Fig. 10. WZ transform coding of a noisy source.

coded separately with a WZ scalar quantizer (for a clean source), followed by an ideal Slepian-Wolf coder (SWC), and reconstructed with the help of the (entire) side information Y. The reconstruction $\hat{\bar{X}}'_Z$ is inversely transformed to obtain $\hat{\bar{X}}_Z = U\hat{\bar{X}}'_Z$. The final estimate of X is $\hat{X} = \bar{x}_Y(Y) + \hat{\bar{X}}_Z$. Clearly, the last summation could be omitted by appropriately modifying the reconstruction functions of each subband. All the definitions of the previous section are maintained, except for the overall rate per sample, which is now $\mathcal{R} = \frac{1}{n}\sum \mathcal{R}_i$, where $\mathcal{R}_i$ is the rate of the i^{th} subband. $\bar{\mathcal{D}} = \frac{1}{n}\,\mathrm{E}\,\|\bar{X}_Z - \hat{\bar{X}}_Z\|^2$ denotes the distortion associated with the clean source $\bar{X}_Z$.

The decomposition of a WZ transform coder of a noisy source into an estimator and a WZ transform coder of a clean source allows the direct application of the results for WZ transform coding of clean sources in Section 4.

Theorem 12 (Noisy WZ transform coding)**.** *Suppose $\bar{x}(y,z)$ is additively separable. Assume the hypotheses of Theorem 8 for $\bar{X}_Z$. In summary, assume that the high-rate approximation hypotheses for WZ quantization of clean sources hold for each subband, the change in the shape of the PDF of the transformed components with the choice of the transform U is negligible, and the variance of the conditional distribution of the transformed coefficients given the side information does not change significantly with the values of the side information. Then, there exists a WZ transform coder, represented in Fig. 10, asymptotically optimal in Lagrangian cost, such that:*

1. *All bands introduce the same distortion $\bar{\mathcal{D}}$. All quantizers are uniform, without index repetition, and with a common interval width Δ such that $\bar{\mathcal{D}} \simeq \Delta^2/12$.*
2. *$\mathcal{D} = \mathcal{D}_\infty + \bar{\mathcal{D}}$, $\bar{\mathcal{D}} \simeq \frac{1}{12}\,2^{\frac{2}{n}\sum_i \mathrm{h}(\bar{X}'_{Zi}|Y)}\,2^{-2\mathcal{R}}$.*
3. *U diagonalizes $\mathrm{E}\,\mathrm{Cov}[\bar{X}_Z|\,Y]$, i.e., is the KLT for the expected conditional covariance matrix of $\bar{X}_Z$.*

Proof: Apply Theorem 8 to $\bar{X}_Z$. Note that since $\bar{X} = \bar{X}_Y + \bar{X}_Z$ and $\hat{X} = \bar{X}_Y + \hat{\bar{X}}_Z$, then $\bar{X}_Z - \hat{\bar{X}}_Z = \bar{X} - \hat{\bar{X}}$, and use (7) for (Y,Z) instead of Z to prove Item 2. □

Similarly to Theorem 6, since $\bar{X}|\,y = \bar{x}_Y(y) + \bar{X}_Z|\,y$, $\mathrm{h}(\bar{X}'_{Zi}|\,Y) = \mathrm{h}(\bar{X}'_i|\,Y)$. In addition, $\bar{\mathcal{D}} = \frac{1}{n}\,\mathrm{E}\,\|\bar{X} - \hat{\bar{X}}\|^2$ and $\mathrm{E}\,\mathrm{Cov}[\bar{X}_Z|\,Y] = \mathrm{E}\,\mathrm{Cov}[\bar{X}|\,Y] \preccurlyeq \mathrm{Cov}\,\bar{X}$.

Corollary 13 (Gaussian case)**.** *If X, Y and Z are jointly Gaussian random vectors, then it is only necessary to assume the high-rate approximation hypotheses of Theorem 12, in order for it to hold. Furthermore, if $\mathcal{D}_{\mathrm{VQ}}$ denotes the distortion when the optimal vector quantizer of Fig. 8 is used, then*

$$\frac{\mathcal{D} - \mathcal{D}_\infty}{\mathcal{D}_{\mathrm{VQ}} - \mathcal{D}_\infty} \simeq \frac{1/12}{M_n} \xrightarrow[n\to\infty]{} \frac{\pi e}{6} \simeq 1.53\,\mathrm{dB}.$$

Proof: $\bar{x}(y,z)$ is additively separable. Apply Corollary 9 to $\bar{X}_Z$ and Y, which are jointly Gaussian. □

Corollary 14 (DCT)**.** *Suppose that $\bar{x}(y,z)$ is additively separable and that for each y, $\mathrm{Cov}[\bar{X}|\,y] = \mathrm{Cov}[\bar{X}_Z|\,y]$ is Toeplitz with a square summable*

associated autocorrelation so that it is also asymptotically circulant as $n \to \infty$. In terms of the associated random processes, this means that $\bar{X}_i$ (equivalently, $\bar{X}_{Zi}$) is conditionally covariance stationary given Y, i.e., $((\bar{X}_i - \mathrm{E}[\bar{X}_i|\, y])|\, y)_{i\in\mathbb{Z}}$ is second-order stationary for each y.

Then, it is not necessary to assume in Theorem 12 that the conditional variance of the transformed coefficients is approximately constant with the values of the side information in order for it to hold, and the DCT is an asymptotically optimal choice for U.

Proof: Apply Corollary 10 to $\bar{X}_Z$ and Y. □

We remark that the coding performance of the cases considered in Corollaries 13 and 14 would be asymptotically the same if the transform U *and* the encoder estimator $\bar{x}_Z(z)$ were allowed to depend on y.

For any random vector Y, set $X = f(Y) + Z + N_X$ and $Z = g(Y) + N_Z$, where f, g are any (measurable) functions, N_X is a random vector such that $\mathrm{E}[N_X|\, y, z]$ is constant with (y, z), and N_Z is a random vector independent from Y such that $\mathrm{Cov}\, N_Z$ is Toeplitz. $\mathrm{Cov}[\bar{X}|\, y] = \mathrm{Cov}[Z|\, y] = \mathrm{Cov}\, N_Z$, thus this is an example of constant conditional variance of transformed coefficients which, in addition, satisfies the hypotheses of Corollary 14.

5.2. Variations on the Fundamental Structure

The fundamental structure of the noisy WZ transform coder analyzed can be modified in a number of ways. We now consider variations on the encoder estimation and transform for this structure, represented completely in Fig. 10, and partially in Fig. 11(a). Later, in Section 6, we shall focus on variations involving the side information.

A general variation consists of performing the encoder estimation in the transform domain. More precisely, define $Z' = U^{\mathrm{T}} Z$, $\bar{X}'_Z = U^{\mathrm{T}} \bar{X}_Z$ and $\bar{x}'_Z(z') = U^{\mathrm{T}} \bar{x}_Z(Uz')$ for all z'. Then, the encoder estimator satisfies $\bar{x}_Z(z) = U \bar{x}'_Z(U^{\mathrm{T}} z)$, as illustrated in Fig. 11(b). Since $UU^{\mathrm{T}} = I$, the estimation and transform $U^{\mathrm{T}} \bar{x}_Z(z)$ can be written simply as $\bar{x}'_Z(U^{\mathrm{T}} z)$, as shown in Fig. 11(c).

The following informal argument will suggest a convenient transform-domain estimation structure. Suppose that X, Y and Z are zero-mean, jointly wide-sense stationary random processes. Suppose further that they are jointly Gaussian, or, merely for simplicity, that a linear estimator of X given $\binom{Y}{Z}$ is required. Then, under certain regularity conditions, a vector Wiener filter $(h_Y\ h_Z)$ can be used to obtain the best linear estimate $\bar{X}$:

$$\bar{X}(n) = (h_Y\ h_Z)(n) * \tbinom{Y}{Z}(n) = \\ = h_Y(n) * Y(n) + h_Z(n) * Z(n).$$

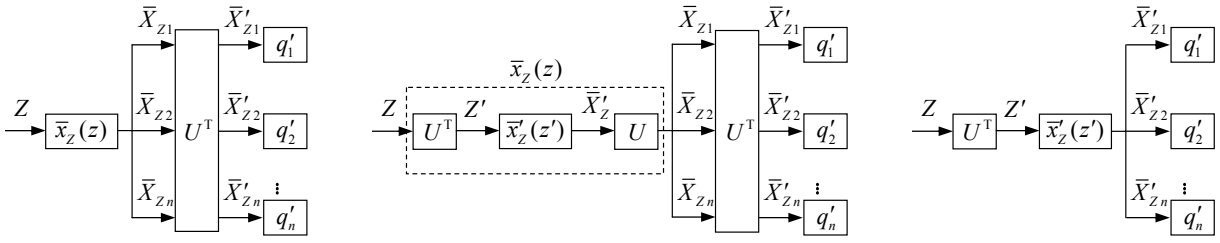
Observe that, in general, h_Y will differ from the individual Wiener filter to estimate X given Y, and similarly for h_Z. The Fourier transform of the Wiener filter is given by

$$(H_Y\ H_Z)(e^{j\omega}) = S_{X\binom{Y}{Z}}(e^{j\omega})\, S_{\binom{Y}{Z}}(e^{j\omega})^{-1}, \quad (13)$$

where S denotes a power spectral density matrix. For example, let N_Y, N_Z be zero-mean wide-sense stationary random processes, representing additive noise, uncorrelated with each other and with X, with a common power spectral density matrix S_N. Let $Y = X + N_Y$ and $Z = X + N_Z$ be noisy versions of X. Then, as an easy consequence of (13), we conclude

$$H_Y(e^{j\omega}) = H_Z(e^{j\omega}) = \\ = \frac{S_X(e^{j\omega})}{2S_X(e^{j\omega}) + S_N(e^{j\omega})}. \quad (14)$$

The factor 2 multiplying S_X in the denominator reflects the fact that 2 signals are using for denoising. Suppose now that X, Y and Z are instead blocks (of equal length) of consecutive samples of random processes. Recall that a block drawn from the convolution of a sequence with a filter can be represented as a product of a Toeplitz matrix h, with entries given by the impulse response of the filter, and a block x drawn from the input sequence. If the filter has finite energy, the Toeplitz matrix h is asymptotically circulant as the block length increases, so that it is asymptotically diagonalized by the Discrete Fourier Transform (DFT) matrix [65,66], denoted by U, as $h = UHU^{\mathrm{T}}$. The matrix multiplication



(a) Fundamental structure. (b) Estimation in the transformed domain. (c) Equivalent structure.

Fig. 11. Variations of the fundamental structure of a WZ transform coder of a noisy source.

$y = hx$, analogous to a convolution, is equivalent to $U^{\mathrm{T}}y = HU^{\mathrm{T}}x$, analogous to a spectral multiplication for each frequency, since H is diagonal. This suggests the following structure for

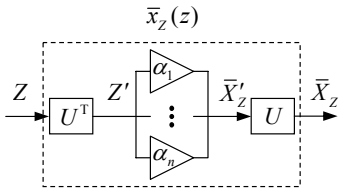


Fig. 12. Structure of the estimator $\bar{x}_Z(z)$ inspired by linear shift-invariant filtering. A similar structure may be used for $\bar{x}_Y(y)$.

the estimator used in the WZ transform coder, represented in Fig. 12: $\bar{x}(y,z) = \bar{x}_Y(y) + \bar{x}_Z(z)$, where $\bar{x}_Z(z) = UH_ZU^{\mathrm{T}}z$, for some diagonal matrix H_Z , and similarly for $\bar{x}_Y(y)$. The (diagonal) entries of H_Y and H_Z can be set according to the best linear estimate of X'_i given $\begin{pmatrix} Y'_i \\ Z'_i \end{pmatrix}$. For the previous example, in which Y and Z are noisy observations of X,

$$(H_{Y\,ii}\, H_{Z\,ii}) = \Sigma_{X'_i \begin{pmatrix} Y'_i \\ Z'_i \end{pmatrix}} \Sigma^{-1}_{\begin{pmatrix} Y'_i \\ Z'_i \end{pmatrix}} \quad \Rightarrow$$

$$\Rightarrow \quad H_{Y\,ii} = H_{Z\,ii} = \frac{\sigma^2_{X'_i}}{2\sigma^2_{X'_i} + \sigma^2_{N'_i}},$$

where $\sigma^2_{X'_i} = u_i^{\mathrm{T}}\Sigma_X u_i$ is the variance of the i^{th} transform coefficient of X, and u_i the corresponding (column) analysis vector of U, and similarly for $\sigma^2_{N'_i}$. Alternatively, $H_{Y\,ii}$ and $H_{Z\,ii}$ can be approximated by sampling the Wiener filter for the underlying processes (14) at the appropriate frequencies. Furthermore, if the Wiener filter component h_Z associated with $\bar{x}_Z(z)$ is even, as in the previous example, then the convolution matrix is not only Toeplitz but *also* symmetric, and the DCT can be used instead of the DFT as the transform U [64][(d)]. An efficient method for general DCT-domain filtering is presented in [67].

If the transform-domain estimator is of the form $\bar{x}'_Z(z') = H_Z z'$, for some diagonal matrix H_Z, as in the structure suggested above, or more generally, if $\bar{x}'_Z(z')$ operates individually on each transformed coefficient z'_i, then the equivalent structure in Fig. 11(c) can be further simplified to group each subband scalar estimation $\bar{x}'_{Z'\,i}(z'_i)$ and each scalar quantizer $q'_i(z'_i)$ as a single quantizer. The resulting structure transforms the noisy observation and then uses a scalar WZ quantizer of a *noisy* source for each subband. This is in general different from the fundamental structure in Figs. 10 or 11(a), in which an estimator was applied to the noisy observation, the estimation was transformed, and each transformed coefficient was quantized with a WZ quantizer for a clean source. Since this modified structure is more constrained than the general structure, its performance may be degraded. However, the design of the noisy WZ scalar quantizers at each subband, for instance using the extension of the Lloyd algorithm in [8], may be simpler than the implementation of a nonlinear vector estimator $\bar{x}_Z(z)$, or a noisy WZ vector quantizer oper-

[(d)] If a real Toeplitz matrix is *not* symmetric, there is no guarantee that the DCT will asymptotically diagonalize it, and the DFT may produce complex eigenvalues.

ating directly on the noisy observation vector.

6. Transformation of the Side Information

6.1. Linear Transformations

Suppose that the side information is a random vector of finite dimension k. A very convenient simplification in the setting of Figs. 9 and 10 would consist of using scalars, obtained by some transformation of the side information vector, in each of the Slepian-Wolf coders and in the reconstruction functions. This is represented in Fig. 13. Even more conveniently, we are interested in linear transforms $Y' = V^{\mathrm{T}} Y$ that lead to a small loss in terms of rate and distortion. It is not required for V to define an injective transform, since no inversion is needed.

Proposition 15. *Let X be a random scalar with mean μ_X, and let Y be a k-dimensional random vector with mean μ_Y. Suppose that X and Y are jointly Gaussian. Let $c \in \mathbb{R}^k$, which gives the linear estimate $\hat{X} = c^{\mathrm{T}} Y$. Then,*

$$\min_c \mathrm{h}(X|\,\hat{X}) = \mathrm{h}(X|\,Y),$$

and the minimum is achieved for c such that $\hat{X}$ is the best linear estimate of $X - \mu_X$ given $Y - \mu_Y$, in the MSE sense.

Proof: Set $c^* = \Sigma_{XY}\Sigma_Y^{-1}$, so that $c^{*\mathrm{T}}Y$ is the best linear estimate of $X-\mu_X$ given $Y-\mu_Y$. (The assumption that Y is Gaussian implies, by definition, the invertibility of Σ_Y, and therefore the existence of a unique estimate.) For each y, $X|\,y$ is a Gaussian random scalar with variance $\sigma^2_{X|Y}$, constant with y, equal to the MSE of the best *affine* estimate of X given Y. Since additive constants preserve variances, the MSE is equal to the *variance* of the error of the best *linear* estimate of $X-\mu_X$ given $Y-\mu_Y$, also equal to $\mathrm{Var}[X-c^{*\mathrm{T}}Y]$. On the other hand, for each c, $X|\,\hat{x}$ is a Gaussian random scalar with variance $\sigma^2_{X|\hat{X}}$ equal to the variance of the error of the best linear estimate of $X-\mu_X$ given $\hat{X}-\mu_{\hat{X}}$, denoted by $\mathrm{Var}[X-\alpha^* c^{\mathrm{T}}Y]$. Minimizing

$$\mathrm{h}(X|\,\hat{X}) = \tfrac{1}{2}\,\log_2(2\pi e\,\sigma^2_{X|\hat{X}})$$

is equivalent to minimizing $\sigma^2_{X|\hat{X}}$. Since

$$\sigma^2_{X|\hat{X}} = \mathrm{Var}[X-\alpha^* c^{\mathrm{T}}Y] \geqslant \mathrm{Var}[X-c^{*\mathrm{T}}Y] = \sigma^2_{X|Y},$$

the minimum is achieved, in particular, for $c = c^*$ and $\alpha^* = 1$ (and in general for any scaled c^*). □

The following theorems on transformation of the side information are given for the more general, noisy case, but are immediately applicable to the clean case by setting $Z = X = \bar{X} = \bar{X}_Z$.

Theorem 16 (Linear transformation of side information)**.** *Under the hypotheses of Corollary 13, for high rates, the transformation of the side information given by*

$$V^{\mathrm{T}} = U^{\mathrm{T}}\,\Sigma_{\bar{X}_Z Y}\,\Sigma_Y^{-1} \tag{15}$$

minimizes the total rate $\mathcal{R}$, with no performance loss in distortion or rate with respect to the transform coding setting of Fig. 10 (and in particular Fig. 9), in which the entire vector Y is used for decoding and reconstruction. Precisely, reconstruction functions defined by $\mathrm{E}[\bar{X}'_{Zi}|\,q, y]$ and by $\mathrm{E}[\bar{X}'_{Zi}|\,q, y'_i]$ give approximately the same distortion $\mathcal{D}_i$, and $\mathcal{R}_i = \mathrm{H}(\bar{X}'_{Zi}|\,Y'_i) \simeq \mathrm{H}(\bar{X}'_{Zi}|\,Y)$.

Proof: Theorems 6 and 12 imply

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

thus the minimization of $\mathcal{R}_i$ is approximately equivalent to the minimization of $\mathrm{h}(\bar{X}'_{Zi}|\,Y)$. Since linear transforms preserve Gaussianity, $\bar{X}'_Z$ and Y are jointly Gaussian, and Proposition 15 applies to each $\bar{X}'_{Zi}$. V is determined by the best linear estimate of $\bar{X}'_Z$ given Y, once the means have been removed. This proves that there is no loss in rate. Corollary 2 implies that a suboptimal reconstruction is asymptotically as efficient, thus there is no loss in distortion either. □

Observe that $\Sigma_{\bar{X}_Z Y}\,\Sigma_Y^{-1}$ in (15) corresponds to the best linear estimate of $\bar{X}_Z$ from Y, disregarding their means. This estimate is transformed according to the same transform applied to X, yielding an estimate of $\bar{X}'_Z$. In addition, joint Gaussianity implies the existence of a matrix B such that $\bar{X}_Z = BZ$. Consequently, $\Sigma_{\bar{X}_Z Y} = B\Sigma_{ZY}$.

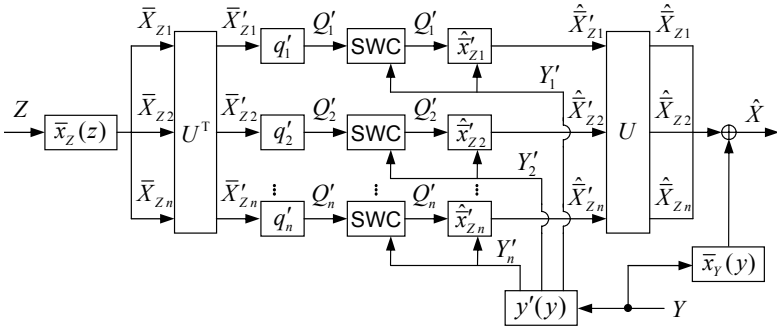


Fig. 13. WZ transform coding of a noisy source with transformed side information.

6.2. General Transformations

Theorem 16 shows that under the hypotheses of high-rate approximation, for jointly Gaussian statistics, the side information could be linearly transformed and a scalar estimate used for Slepian-Wolf decoding and reconstruction in each subband, instead of the entire vector Y, with no asymptotic loss in performance. Here we extend this result to general statistics, connecting WZ coding and statistical inference.

Let X and Θ be random variables, representing, respectively, an observation and some data we wish to estimate. A statistic for Θ from X is a random variable T such that $\Theta \leftrightarrow X \leftrightarrow T$, for instance, any function of X. A statistic is sufficient if and only if $\Theta \leftrightarrow T \leftrightarrow X$.

Proposition 17. *A statistic T for a continuous random variable Θ from an observation X satisfies $\mathrm{h}(\Theta|\,T) \geqslant \mathrm{h}(\Theta|\,X)$, with equality if and only if T is sufficient.*

Proof: Use the data processing inequality to write $\mathrm{I}(\Theta;T) \leqslant \mathrm{I}(\Theta;X)$, with equality if and only if T is sufficient [68], and express the mutual information as a difference of entropies. □

Theorem 18 (Reduction of side information)**.** *Under the hypotheses of Theorem 12 (or Corollaries 13 or 14), a sufficient statistic Y'_i for $\bar{X}'_{Zi}$ from Y can be used instead of Y for Slepian-Wolf decoding and reconstruction, for each subband i in the WZ transform coding setting of Fig. 10, with no asymptotic loss in performance.*

Proof: Theorems 6 and 12 imply $\mathcal{R}_i = \mathrm{H}(\bar{X}'_{Zi}|\,Y) \simeq \mathrm{h}(\bar{X}'_{Zi}|\,Y) - \log_2 \Delta$. Proposition 17 ensures that $\mathrm{h}(\bar{X}'_{Zi}|\,Y) = \mathrm{h}(\bar{X}'_{Zi}|\,Y'_i)$, and Corollary 7 that a suboptimal reconstruction is asymptotically as efficient if Y'_i is used instead of Y. □

In view of these results, Theorem 16 incidentally shows that in the Gaussian case, the best linear MSE estimate is a sufficient statistic, which can also be proven directly (for instance combining Propositions 15 and 17). The obtention of (minimal) sufficient statistics has been studied in the field of statistical inference, and the Lehmann-Scheffé method is particularly useful (e.g. [69]).

Many of the ideas on the structure of the estimator $\bar{x}(y,z)$ presented in Section 5.2 can be applied to the transformation of the side information $y'(y)$. For instance, it could be carried out in the domain of the data transform U. If, in addition, $\bar{x}_Y(y)$ is also implemented in the transform domain, for example in the form of Fig. 12, then, in view of Fig. 13, a single transformation can be shared as the first step of both $y'(y)$ and $\bar{x}_Y(y)$. Furthermore, the summation $\bar{x}_Y(Y) + \hat{\bar{X}}_Z$ can be carried out in the transform domain, since $\hat{\bar{X}}'_Z$ is available, eliminating the need to undo the transform as the last step of $\bar{x}_Y(y)$.

Finally, suppose that the linear transform in (15) is used, and that U (asymptotically) diagonalizes both $\Sigma_{\bar{X}_Z Y}$ and Σ_Y. Then, since U is orthonormal, it is easy to see that $y'(y) = V^{\mathrm{T}} y = \Lambda_{\bar{X}_Z Y} \Lambda_Y^{-1} U^{\mathrm{T}} y$, where Λ denotes the corresponding diagonal matrices and $U^{\mathrm{T}} y$ is the transformed

side information. Of course, the scalar multiplications for each subband may be suppressed by designing the Slepian-Wolf coders and the reconstruction functions accordingly, and, if $\bar{x}_Y(y)$ is of the form of Fig. 12, the additions in the transform domain can be incorporated into the reconstruction functions.

7. Experimental Results

7.1. Transform WZ Coding of Clean Video

In [11], we apply WZ 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 WZ decoding is performed by exploiting the temporal similarities between the current frame and the side information.

In the following experiments, we extend the WZ video codec, outlined in [11], to a transform domain WZ 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 WZ frames which are intraframe encoded but interframe decoded, adopting the WZ transform coding setup for clean sources and transformed side information, described in Sections 4 and 6.

To encode a WZ frame X, we first apply a blockwise DCT to generate X'. As stated in Corollary 10, the DCT is an asymptotically optimal choice for the orthonormal transform. Although the system is not necessarily high rate, we follow Theorem 8 in the quantizer design. Each transform coefficient subband is independently quantized using uniform scalar quantizers with no index repetitions and similar step sizes across bands[(e)]. This allows for a approximately uniform distortion across bands. Rate-compatible punctured turbo codes are used for Slepian-Wolf coding in each subband. 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. The rate-compatible punctured turbo code for each band can come close to achieving the ideal Slepian-Wolf rate for the given transform band.

At the decoder, we take previously reconstructed frames to generate side information Y, which is used to decode X. In the first setup (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 WZ frame. The DCT is applied to 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 subband is then reconstructed as the best estimate given the previously reconstructed symbols and the side information. Note that unlike in Fig. 9, where the entire Y is used as side information for decoding and reconstructing each transform band, in our simplified implementation, we only use the corresponding side information subband. More details of the proposed scheme and extended results can be found in [70].

The compression results for the first 100 frames of *Mother & Daughter* are shown in Fig. 14. MSE has been used as a distortion measure, expressed as Peak Signal-To-Noise Ratio (PSNR) in dB, defined as $10\log_{10}(255^2/\text{MSE})$. 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:

1. DCT-based intraframe coding: the even frames are encoded as Intracoded (I) frames.
2. H.263+ interframe coding with an I-B-I-B predictive structure, counting only the rate

[(e)] Since we use fixed-length codes for Slepian-Wolf coding and each transform band has different dynamic range, it is not possible to have exactly the same step sizes.

and distortion of the Bidirectionally predicted (B) frames.

We also plot the compression results of the pixel-domain WZ codec. For the pixel-domain WZ codec, we quantize the pixels using the optimal scalar quantizers suggested by Theorem 1, that is, uniform quantizers with no index repetition.

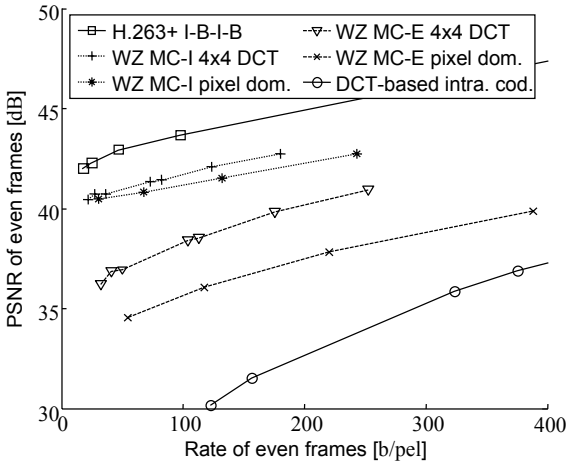


Fig. 14. Rate and PSNR comparison of WZ codec vs. DCT-based intraframe coding and H.263+ I-B-I-B coding. *Mother & Daughter* sequence.

As 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 WZ codec. With the less reliable MC-E, using a transform before encoding results in a 2 to 2.5 dB improvement. These improvements over the pixel-domain system demonstrate the transform gain in a practical Wyner-Ziv coding setup.

Compared to conventional DCT-based intraframe coding, the WZ 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 approaching to the compression efficiency of interframe video coders.

7.2. WZ Transform Coding of Noisy Images

We implement various cases of WZ transform coding of noisy images to confirm the theoretic results of Sections 3, 5 and 6. The source data X consists of all 8×8 blocks of the first 25 frames of the *Foreman* Quarter Common Intermediate Format (QCIF) video sequence, with the mean removed. Assume that the encoder does not know X, but has access to $Z = X + V$, where V is a block of white Gaussian noise of variance σ_V^2. The decoder has access to side information $Y = X + W$, where W is white Gaussian noise of variance σ_W^2. Note that this experimental setup reproduces our original statement of the problem of WZ quantization of noisy sources, drawn in Fig. 3. In our experiments, V, W and X are statistically independent. In this case, $\mathrm{E}[X \,|\, y, z]$ is not additively separable. However, since we wish to test our theoretic results which apply only to separable estimates, we constrain our estimators to be linear. Thus, in all the experiments of this section, the estimate of X, given Y and Z is defined as

$$\bar{x}(y,z) = \Sigma_{X\binom{Y}{Z}} \Sigma^{-1}_{\binom{Y}{Z}} \begin{pmatrix} y \\ z \end{pmatrix} = \bar{x}_Y(y) + \bar{x}_Z(z).$$

We now consider the following cases, all constructed using linear estimators and WZ 2D-DCT coders of clean sources:

1. Assume that Y is made available to the encoder estimator, perform conditional linear estimation of X given Y and Z, followed by WZ transform coding of the estimate. This corresponds to a conditional coding scenario where both the encoder and the decoder have access to the side information Y. This experiment is carried out for the purpose of comparing its performance with that of true WZ transform coding of a noisy source. Since we are concerned with the performance of uniform quantization at high rates, the quantizers q_i' are all chosen to be uniform, with the same step-size for all transform sub-bands. We assume ideal entropy coding of the quantization indices conditioned on the side-information.

2. Perform noisy WZ transform coding of Z exactly as shown in Fig. 13. As mentioned above, the orthonormal transform

being used is the 2D-DCT applied to 8×8 pixel blocks of $\bar{X}_Z$. As before, all quantizers are uniform with the same step-size and ideal Slepian-Wolf coding is assumed, i.e., the Slepian-Wolf rate required to encode the quantization indices in the i^{th} sub-band is simply the conditional entropy $\mathrm{H}(Q'_i|Y'_i)$. As seen in Fig. 13, the decoder recovers the estimate $\hat{\bar{X}}_Z$, and obtains the final estimate as $\hat{X} = \bar{x}_Y(Y) + \hat{\bar{X}}_Z$.

3. Perform WZ transform coding directly on Z, reconstruct $\hat{Z}$ at the decoder and obtain $\hat{X} = \bar{x}(Y, \hat{Z})$. This experiment is performed in order to investigate the penalty incurred, when the the noisy input Z is treated as a clean source for WZ transform coding.

4. Perform noisy WZ transform coding of Z as in Case 2, except that $\hat{\bar{x}}'_{Zi}(q'_i, y'_i) = \mathrm{E}[\bar{X}'_i | \, q'_i]$, i.e., the reconstruction function does not use the side information Y. This experiment is performed in order to examine the penalty incurred at high rates for the situation described in Corollary 7, where the side-information is used for Slepian-Wolf encoding but ignored in the reconstruction function.

Fig. 15 plots rate vs. PSNR for the above cases, with $\sigma_V^2 = \sigma_W^2 = 25$, and $\sigma_X^2 = 2730$ (measured). The performance of conditional estimation (Case 1) and WZ transform coding (Case 2) are in close agreement at high rates as predicted by Theorem 12. Our theory does not explain the behavior at low rates. Experimentally, we observed that Case 2 slightly outperforms Case 1 at lower rates. Both cases show superior rate-distortion performance than direct WZ coding of Z (Case 3). Neglecting the side-information in the reconstruction function (Case 4) is inefficient at low rates, but at high rates, this simpler scheme approaches the performance of Case 2 with the ideal reconstruction function, thus confirming Corollary 7.

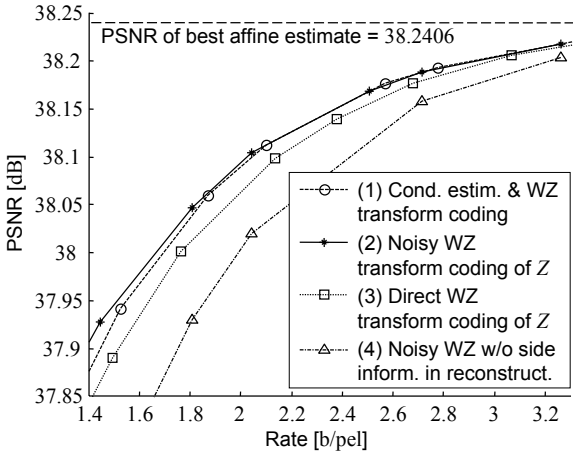


Fig. 15. WZ transform coding of a noisy image is asymptotically equivalent to the conditional case. *Foreman* sequence.

8. Conclusions

If ideal Slepian-Wolf coders are used, uniform tessellating quantizers without index repetition are asymptotically optimal at high rates. It is known [5] that the rate loss in the WZ problem for smooth continuous sources and quadratic distortion vanishes as $\mathcal{D} \to 0$. Our work shows that this is true also for the operational rate loss and for each finite dimension n.

The theoretic 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.

If the conditional expectation of the unseen source data X given the side information Y and the noisy observation Z is additively separable, then, at high rates, optimal WZ quantizers of Z can be decomposed into estimators and uniform tessellating quantizers for clean sources, achieving the same rate-distortion performance as if the side information were available at the encoder. This is consistent with the experimental results of the application of the Lloyd algorithm for noisy WZ quantization design in [54].

The additive separability condition for high-

rate WZ quantization of noisy sources, albeit less restrictive, is similar to the condition required for zero rate loss in the quadratic Gaussian noisy Wyner-Ziv problem [8], which applies exactly for any rate but requires arbitrarily large dimension.

We propose a WZ transform coder of noisy sources consisting of an estimator and a WZ transform coder for clean sources. Under certain conditions, in particular if the encoder estimate is conditionally covariance stationary given Y, the DCT is an asymptotically optimal transform. The side information for the Slepian-Wolf decoder and the reconstruction function in each subband can be replaced by a sufficient statistic with no asymptotic loss in performance.

A. Appendix: Asymmetric Slepian-Wolf Coding with Arbitrary Side Information

We show that the asymmetric Slepian-Wolf coding result extends to sources in countable alphabets and arbitrary, possibly continuous, side information. Let U, V be arbitrary random variables. The general definition of conditional entropy used here is $\mathrm{H}(U|V) = \mathrm{I}(U; U|V)$, with the conditional mutual information defined in [71], consistent with [12].

Proposition 19 (Asymmetric Slepian-Wolf Coding)**.** *Let Q be a discrete random variable representing source data to be encoded, and let Y be an arbitrary random variable acting as side information in an asymmetric Slepian-Wolf coder of Q. Any rate satisfying $\mathcal{R} > \mathrm{H}(Q|Y)$ is admissible in the sense that a code exists with arbitrarily low probability of decoding error, and any rate $\mathcal{R} < \mathrm{H}(Q|Y)$ is not admissible.*

Proof: The statement follows from the Wyner-Ziv theorem for general sources [12], that is, the one-letter characterization of the Wyner-Ziv rate-distortion function for source data and side information distributed in arbitrary alphabets. Suppose first that the alphabet of Q is finite, and the alphabet of Y is arbitrary. The reconstruction alphabet in the Wyner-Ziv coding setting is chosen to be the same as that of Q, and the distortion measure is the Hamming distance $d(q, \hat{q}) = 1\{q \neq \hat{q}\}$. Then, $\mathrm{I}(Q; Y) = \mathrm{H}(Q) - \mathrm{H}(Q|Y) < \infty$ [71, Lemma 2.1], as assumed in the Wyner-Ziv theorem, and the two regularity conditions required by theorem are clearly satisfied [12, Equation (2.2),Theorems 2.1,2.2]. Proceeding as in [4, Remark 3, p.3], and realizing that Fano's inequality is still valid for arbitrary Y, the asymmetric Slepian-Wolf result follows immediately from the the Wyner-Ziv theorem. This proves the case of finite Q and arbitrary Y.

In the case when the alphabet of Q is countably infinite, observe that for any rate $\mathcal{R} > \mathrm{H}(Q|Y)$, there exists a finite quantization $\tilde{Q}$ of Q such that $\mathcal{R} > \mathrm{H}(Q|Y) \geqslant \mathrm{H}(\tilde{Q}|Y)$, and $\tilde{Q}$ can be coded with arbitrarily low probability of decoding error between the original Q and the decoded $\tilde{Q}$. This shows achievability. To prove the converse, suppose that a rate $\mathcal{R} < \mathrm{H}(Q|Y)$ is admissible. For any such $\mathcal{R}$, simply by the definition of conditional entropy for general alphabets in terms of finite measurable partitions, there exists a finite quantization $\tilde{Q}$ of Q such that $\mathcal{R} < \mathrm{H}(\tilde{Q}|Y) \leqslant \mathrm{H}(Q|Y)$. Since the same code used for Q could be used for $\tilde{Q}$, with trivial modifications, with arbitrarily small probability of error, this would contradict the converse for the finite-alphabet case. □

The proposition can also be proven directly from the Slepian-Wolf result for finite alphabets, without invoking the Wyner-Ziv theorem for general sources. The arguments used are somewhat lengthier, but similarly to the previous proof, exploit the definitions of general information-theoretic quantities in terms of finite measurable partitions.

Acknowledgment

The authors would like to thank the anonymous reviewers for their helpful comments, which motivated a number improvements in this paper.

References

[1] D. Rebollo-Monedero, A. Aaron, B. Girod, Transforms for high-rate distributed source coding, in: Proc. Asilomar Conf. Signals, Syst., Comput., Vol. 1, Pacific Grove, CA, 2003, pp. 850–854, invited paper.

[2] D. Rebollo-Monedero, S. Rane, B. Girod, Wyner-Ziv quantization and transform coding of noisy sources at

high rates, in: Proc. Asilomar Conf. Signals, Syst., Comput., Vol. 2, Pacific Grove, CA, 2004, pp. 2084 – 2088.

[3] J. D. Slepian, J. K. Wolf, Noiseless coding of correlated information sources, IEEE Trans. Inform. Theory IT-19 (1973) 471–480.

[4] A. D. Wyner, J. Ziv, The rate-distortion function for source coding with side information at the decoder, IEEE Trans. Inform. Theory IT-22 (1) (1976) 1–10.

[5] R. Zamir, The rate loss in the Wyner-Ziv problem, IEEE Trans. Inform. Theory 42 (6) (1996) 2073–2084.

[6] T. Linder, R. Zamir, K. Zeger, On source coding with side-information-dependent distortion measures, IEEE Trans. Inform. Theory 46 (7) (2000) 2697–2704.

[7] H. Yamamoto, K. Itoh, Source coding theory for multiterminal communication systems with a remote source, Trans. IECE Japan E63 (1980) 700–706.

[8] D. Rebollo-Monedero, B. Girod, A generalization of the rate-distortion function for Wyner-Ziv coding of noisy sources in the quadratic-Gaussian case, in: Proc. IEEE Data Compression Conf. (DCC), Snowbird, UT, 2005, pp. 23–32.

[9] H. S. Witsenhausen, A. D. Wyner, Interframe coder for video signals, U.S. Patent 4191970 (Nov. 1980).

[10] R. Puri, K. Ramchandran, PRISM: A new robust video coding architecture based on distributed compression principles, in: Proc. Allerton Conf. Commun., Contr., Comput., Allerton, IL, 2002.

[11] A. Aaron, R. Zhang, B. Girod, Wyner-Ziv coding of motion video, in: Proc. Asilomar Conf. Signals, Syst., Comput., Pacific Grove, CA, 2002, pp. 240–244.

[12] A. D. Wyner, The rate-distortion function for source coding with side information at the decoder—II: General sources, Inform., Contr. 38 (1) (1978) 60–80.

[13] R. Zamir, T. Berger, Multiterminal source coding with high resolution, IEEE Trans. Inform. Theory 45 (1) (1999) 106–117.

[14] H. Viswanathan, Entropy coded tesselating quantization of correlated sources is asymptotically optimal, unpublished (1996).

[15] M. E. Hellman, Convolutional source encoding, IEEE Trans. Inform. Theory IT-21 (6) (1975) 651–656.

[16] S. S. Pradhan, K. Ramchandran, Distributed source coding using syndromes (DISCUS): Design and construction, in: Proc. IEEE Data Compression Conf. (DCC), Snowbird, UT, 1999, pp. 158–167.

[17] J. García-Frías, Y. Zhao, Compression of correlated binary sources using turbo codes, IEEE Commun. Lett. 5 (10) (2001) 417–419.

[18] J. Bajcsy, P. Mitran, Coding for the Slepian Wolf problem with turbo codes, in: Proc. IEEE Global Telecomm. Conf. (GLOBECOM), Vol. 2, 2001, pp. 1400–1404.

[19] G.-C. Zhu, F. Alajaji, Turbo codes for nonuniform memoryless sources over noisy channels, IEEE Commun. Lett. 6 (2) (2002) 64–66.

[20] A. Aaron, B. Girod, Compression with side information using turbo codes, in: Proc. IEEE Data Compression Conf. (DCC), Snowbird, UT, 2002, pp. 252–261.

[21] A. D. Liveris, Z. Xiong, C. N. Georghiades, A distributed source coding technique for correlated images using turbo-codes, IEEE Commun. Lett. 6 (9) (2002) 379–381.

[22] A. D. Liveris, Z. Xiong, C. N. Georghiades, Compression of binary sources with side information at the decoder using LDPC codes, IEEE Commun. Lett. 6 (10) (2002) 440–442.

[23] D. Schonberg, S. S. Pradhan, K. Ramchandran, LDPC codes can approach the Slepian Wolf bound for general binary sources, in: Proc. Allerton Conf. Commun., Contr., Comput., Allerton, IL, 2002.

[24] B. Girod, A. Aaron, S. Rane, D. Rebollo-Monedero, Distributed video coding, in: Proc. IEEE, Special Issue Advances Video Coding, Delivery, Vol. 93, 2005, pp. 71–83, invited paper.

[25] R. Zamir, S. Shamai, Nested linear/lattice codes for Wyner-Ziv encoding, in: Proc. IEEE Inform. Theory Workshop (ITW), Killarney, Ireland, 1998, pp. 92–93.

[26] R. Zamir, S. Shamai, U. Erez, Nested linear/lattice codes for structured multiterminal binning, IEEE Trans. Inform. Theory 48 (6) (2002) 1250–1276.

[27] S. D. Servetto, Lattice quantization with side information, in: Proc. IEEE Data Compression Conf. (DCC), Snowbird, UT, 2000, pp. 510–519.

[28] J. Kusuma, L. Doherty, K. Ramchandran, Distributed compression for sensor networks, in: Proc. IEEE Int. Conf. Image Processing (ICIP), Vol. 1, Thessaloniki, Greece, 2001, pp. 82–85.

[29] M. Fleming, M. Effros, Network vector quantization, in: Proc. IEEE Data Compression Conf. (DCC), Snowbird, UT, 2001, pp. 13–22.

[30] M. Fleming, Q. Zhao, M. Effros, Network vector quantization, IEEE Trans. Inform. Theory 50 (8) (2004) 1584–1604.

[31] J. Cardinal, G. V. Asche, Joint entropy-constrained multiterminal quantization, in: Proc. IEEE Int. Symp. Inform. Theory (ISIT), Lausanne, Switzerland, 2002, p. 63.

[32] D. Rebollo-Monedero, R. Zhang, B. Girod, Design of optimal quantizers for distributed source coding, in: Proc. IEEE Data Compression Conf. (DCC), Snowbird, UT, 2003, pp. 13–22.

[33] S. P. Lloyd, Least squares quantization in PCM, IEEE Trans. Inform. Theory IT-28 (1982) 129–1373.

[34] Z. Xiong, A. Liveris, S. Cheng, Z. Liu, Nested quantization and Slepian-Wolf coding: A Wyner-Ziv coding paradigm for i.i.d. sources, in: Proc. IEEE Workshop Stat. Signal Processing (SSP), St. Louis, MO, 2003, pp. 399–402.

[35] Y. Yang, S. Cheng, Z. Xiong, W. Zhao, Wyner-Ziv coding based on TCQ and LDPC codes, in: Proc. Asilomar Conf. Signals, Syst., Comput., Vol. 1, Pacific Grove, CA, 2003, pp. 825–829.

[36] Z. Xiong, A. D. Liveris, S. Cheng, Distributed source coding for sensor networks, IEEE Signal Processing Mag. 21 (5) (2004) 80–94.

[37] H. Hotelling, Analysis of a complex of statistical variables into principal components, J. Educ. Psychol. 24 (1933) 417–441, 498–520.

[38] K. Karhunen, Über lineare methoden in der

wahrscheinlichkeitsrechnung, Ann. Acad. Sci. Fenn., Ser. A I Math.-Phys. 37 (1947) 3–79.

[39] M. Loève, Fonctions aléatoires du second ordre, in: P. Lévy (Ed.), Processus stochastiques et mouvement Brownien, Gauthier-Villars, Paris, France, 1948.

[40] M. Gastpar, P. L. Dragotti, M. Vetterli, The distributed Karhunen-Loève transform, in: Proc. IEEE Int. Workshop Multimedia Signal Processing (MMSP), St. Thomas, US Virgin Islands, 2002, pp. 57–60.

[41] M. Gastpar, P. L. Dragotti, M. Vetterli, The distributed, partial, and conditional Karhunen-Loève transforms, in: Proc. IEEE Data Compression Conf. (DCC), Snowbird, UT, 2003, pp. 283–292.

[42] M. Gastpar, P. L. Dragotti, M. Vetterli, On compression using the distributed Karhunen-Loève transform, in: Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing (ICASSP), Vol. 3, Philadelphia, PA, 2004, pp. 901–904.

[43] M. Gastpar, P. L. Dragotti, M. Vetterli, The distributed Karhunen-Loève transform, IEEE Trans. Inform. TheorySubmitted.

[44] S. S. Pradhan, K. Ramchandran, Enhancing analog image transmission systems using digital side information: A new wavelet-based image coding paradigm, in: Proc. IEEE Data Compression Conf. (DCC), Snowbird, UT, 2001, pp. 63–72.

[45] R. L. Dobrushin, B. S. Tsybakov, Information transmission with additional noise, IRE Trans. Inform. Theory IT-8 (1962) S293–S304.

[46] J. K. Wolf, J. Ziv, Transmission of noisy information to a noisy receiver with minimum distortion, IEEE Trans. Inform. Theory IT-16 (4) (1970) 406–411.

[47] Y. Ephraim, R. M. Gray, A unified approach for encoding clean and noisy sources by means of waveform and autoregressive vector quantization, IEEE Trans. Inform. Theory IT-34 (1988) 826–834.

[48] T. Flynn, R. Gray, Encoding of correlated observations, IEEE Trans. Inform. Theory 33 (6) (1987) 773–787.

[49] H. S. Witsenhausen, Indirect rate-distortion problems, IEEE Trans. Inform. Theory IT-26 (1980) 518–521.

[50] S. C. Draper, G. W. Wornell, Side information aware coding strategies for sensor networks, IEEE J. Select. Areas Commun. 22 (6) (2004) 966–976.

[51] W. M. Lam, A. R. Reibman, Quantizer design for decentralized estimation systems with communication constraints, in: Proc. Conf. Inform. Sci. Syst., Baltimore, MD, 1989.

[52] W. M. Lam, A. R. Reibman, Design of quantizers for decentralized estimation systems, IEEE Trans. Inform. Theory 41 (11) (1993) 1602–1605.

[53] J. A. Gubner, Distributed estimation and quantization, IEEE Trans. Inform. Theory 39 (4) (1993) 1456–1459.

[54] D. Rebollo-Monedero, B. Girod, Design of optimal quantizers for distributed coding of noisy sources, in: Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing (ICASSP), Vol. 5, Philadelphia, PA, 2005, pp. 1097–1100, invited paper.

[55] R. M. Gray, D. L. Neuhoff, Quantization, IEEE Trans. Inform. Theory 44 (1998) 2325–2383.

[56] V. K. Goyal, Theoretical foundations of transform coding, IEEE Signal Processing Mag. 18 (5) (2001) 9–21.

[57] T. M. Cover, A proof of the data compression theorem of Slepian and Wolf for ergodic sources, IEEE Trans. Inform. Theory 21 (2) (1975) 226–228, (Corresp.).

[58] W. R. Bennett, Spectra of quantized signals, Tech. J. 27, Bell Syst. (Jul. 1948).

[59] S. Na, D. L. Neuhoff, Bennett's integral for vector quantizers, IEEE Trans. Inform. Theory 41 (1995) 886–900.

[60] A. Gersho, Asymptotically optimal block quantization, IEEE Trans. Inform. Theory IT-25 (1979) 373–380.

[61] P. L. Zador, Topics in the asymptotic quantization of continuous random variables, Tech. Memo. 1966, Bell Lab., unpublished.

[62] R. M. Gray, T. Linder, J. Li, A Lagrangian formulation of Zador's entropy-constrained quantization theorem, IEEE Trans. Inform. Theory 48 (3) (2002) 695–707.

[63] R. Zamir, M. Feder, On lattice quantization noise, IEEE Trans. Inform. Theory 42 (4) (1996) 1152–1159.

[64] K. R. Rao, P. Yip, Discrete cosine transform: Algorithms, advantages, applications, Academic Press, San Diego, CA, 1990.

[65] U. Grenander, G. Szegö, Toeplitz forms and their applications, University of California Press, Berkeley, CA, 1958.

[66] R. M. Gray, Toeplitz and circulant matrices: A review (2002).
URL `http://ee.stanford.edu/~gray/toeplitz.pdf`

[67] R. Kresch, N. Merhav, Fast DCT domain filtering using the DCT and the DST, IEEE Trans. Image Processing 8 (1999) 821–833.

[68] T. M. Cover, J. A. Thomas, Elements of information theory, Wiley, New York, 1991.

[69] G. Casella, R. L. Berger, Statistical Inference, 2nd Edition, Thomson Learning, Australia, 2002.

[70] A. Aaron, S. Rane, E. Setton, B. Girod, Transform-domain Wyner-Ziv codec for video, in: Proc. IT&S/SPIE Conf. Visual Commun., Image Processing (VCIP), San Jose, CA, 2004.

[71] A. D. Wyner, A definition of conditional mutual information for arbitrary ensembles, Inform., Contr. 38 (1) (1978) 51–59.