Wyner-Ziv Quantization and Transform Coding of Noisy Sources at High Rates

David Rebollo-Monedero, Shantanu Rane and Bernd Girod
Information Systems Lab., Dept. of Electrical Eng.
Stanford University, Stanford, CA 94305
{drebollo,srane,bgirod}@stanford.edu

***Abstract*— We extend Wyner-Ziv high-rate quantization and transform coding theory to the case in which a noisy observation of some source data is available at the encoder, but we are interested in estimating the unseen source data at the decoder, with the help of side information. Ideal Slepian-Wolf coders are assumed, thus rates are conditional entropies of quantization indices given the side information. Transform coders of noisy images for different communication constraints are compared. Experimental results show that the Wyner-Ziv transform coder achieves a performance close to the case in which the side information is also available at the encoder.**

I. Introduction

Consider an image sensor sending a noisy reading to a central station which has access to a similar, local noisy image. If both noisy images were available at the remote sensor, efficient joint denoising could be carried out and the result could be sent to the central station. In most cases, reducing the noise would also help reduce the amount of bits required to encode the information sent. In addition, the remote sensor could exploit the statistical dependence with the local information, available to the central station, to further reduce the rate. However, if the noisy local image is *not* available at the encoder, we wish to know if it is possible to maintain the same rate-distortion performance, and how to build a coder capable of such performance.

Source coding with side information at the decoder, also known as Wyner-Ziv (WZ) coding in the lossy case, has been extensively studied. The information-theoretical work [1]–[3] establishes reasonable conditions under which the rate-distortion performance is similar to the case in which the side information is also available at the encoder. This has been confirmed by studies aimed at practical implementations, WZ quantization and transform coding [4]–[8].

There is also extensive literature on source coding of a noisy observation of an unseen source without side information [9]–[11]. However, most of the work on distributed coding of noisy sources is information-theoretical [12]–[14], or when operational, is based on fixed-rate coding [15], [16] and does not consider high-rate quantization or transforms.

In this paper, we extend the theory on high-rate quantization and transform coding with side information in [7] to coding of noisy sources, assuming the availability of lossless coders with efficiency close to ideal Slepian-Wolf coding [4].

Section II contains the theoretical results for high-rate quantization, and Section III, for transforms coding. Experimental results on image denoising are shown in Section IV.

Throughout the paper, we follow the convention of using uppercase letters for random vectors, and lowercase letters for particular values they take on. We shall use the operator name Cov for the covariance matrix of random vectors, and the lowercase version cov for its trace.

II. High-Rate WZ Quantization of Noisy Sources

We study the properties of high-rate quantizers of a noisy source with side information at the decoder, as illustrated in Fig. 1, which we shall refer to as WZ quantizers of a noisy source. A noisy observation Z of some unseen source

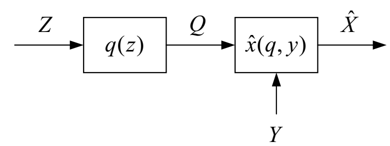

Fig. 1. WZ quantization of a noisy source.

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 dimension $n \in \mathbb{Z}^+$ (no restrictions on the alphabets of Y and Z are imposed).

Mean-squared error is used as a distortion measure, thus the expected distortion *per sample of the unseen source* is $\mathcal{D} = \frac{1}{n} \operatorname{E}[\|X - \hat{X}\|^2]$. The formulation in this work assumes that the coding of the index Q is carried out by an ideal Slepian-Wolf coder. The expected rate per sample is defined accordingly as $\mathcal{R} = \frac{1}{n} \operatorname{H}(Q|Y)$ [4]. 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 $\mathcal{R}$.

We start by considering the simpler case of quantization of a noisy source *without* side information, depicted in Fig. 2. The following theorem extends the main result of [10], [11] to entropy-constrained quantization, valid for any rate $\mathcal{R} =$

0-7803-8622-1/04/$20.00

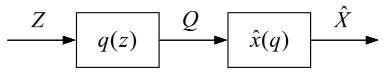

Fig. 2. Quantization of a noisy source without side information.

$\mathrm{H}(Q)$, 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 1 (MSE quantization of a noisy source): For any nonnegative λ and any Lagrangian-cost optimal quantizer of a noisy source *without* side information (Fig. 2), there exists an implementation with the same cost in two steps:

1) Obtention of the estimate $\bar{X}$.
2) Quantization of $\bar{X}$ 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. 3. Furthermore, the total distortion per sample is

$$\mathcal{D} = \tfrac{1}{n}(\mathrm{E}\operatorname{cov}[X|Z] + \mathrm{E}[\|\bar{X} - \hat{X}\|^2]), \qquad (1)$$

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

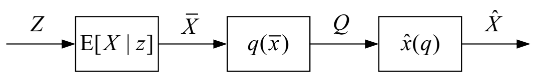

Fig. 3. Optimal implementation of MSE quantization of a noisy source without side information.

Proof: The proof is a modification of that in [11], 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 (1). Note that the first term of (1) 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, i. e., satisfying $\sum_q 2^{-r(q)} \leq 1$, with $\mathcal{R} = \mathrm{E}\, r(Q)$. The Lagrangian cost of the setting in Fig. 2 can be written as

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

and the cost of the setting in Fig. 3 as

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

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

The hypotheses of the next theorem are believed to hold if the Bennett assumptions [17], [18] apply to the PDF $p(\bar{x})$ of the MSE estimate, and if Gersho's conjecture [19] is true (known to be the case for $n = 1$), among other technical conditions, mentioned in [20]. M_n denotes the minimum normalized moment of inertia of the convex polytopes tessellating $\mathbb{R}^n$ (e. g., $M_1 = 1/12$).

Theorem 2 (High-rate quantization of a noisy source): Assume that $\mathrm{h}(\bar{X}) < \infty$ and that there exists a lattice 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. 2 such that:

1) An asymptotically optimal implementation of $q(z)$ is that of Theorem 1, represented in Fig. 3, with a lattice quantizer $q(\bar{x})$ having cell volume V.
2) The rate and the distortion per sample satisfy

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

Proof: Immediate from Theorem 1 and conventional theory of high-rate quantization of clean sources. ∎

We are now ready to consider the WZ quantization of a noisy source in Fig. 1. 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 $\mathcal{D}_\infty = \frac{1}{n}\,\mathrm{E}\operatorname{cov}[X|Y, Z]$. The following theorem extends the results on high-rate WZ quantization in [7] to noisy sources. The remark on the hypotheses of Theorem 2 also applies here, where the Bennett assumptions apply instead to the conditional PDF $p(\bar{x}|y)$ *for each* y.

Theorem 3 (High-rate WZ quantization of a noisy source): 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 support set of Y, $\mathrm{h}(\bar{X}|y) < \infty$, and that there exists a lattice 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 $\mathcal{R}$, or more precisely, minimizing $\mathcal{C}$ as $\lambda \to 0^+$, for the WZ quantization setting represented in Fig. 1 such that:

1) $q(z)$ can be implemented as an estimator $\bar{x}_Z(z)$ followed by a lattice 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

$$\mathcal{D} \simeq \mathcal{D}_\infty + M_n\, V^{\frac{2}{n}}, \qquad (2)$$
$$\mathcal{R} \simeq \tfrac{1}{n}(\mathrm{h}(\bar{X}|Y) - \log_2 V), \qquad (3)$$
$$\mathcal{D} \simeq \mathcal{D}_\infty + M_n\, 2^{\frac{2}{n}\,\mathrm{h}(\bar{X}|Y)}\, 2^{-2\mathcal{R}}. \qquad (4)$$

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

Proof: The proof is similar to that for clean sources [7, 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. 4. An entirely analogous argument using conditional costs, as

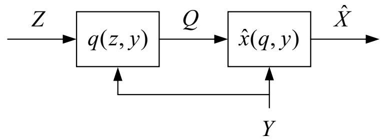

Fig. 4. Conditional quantization of a noisy source.

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 everywhere, by Theorem 1, the optimal conditional quantizer can be implemented as in Fig. 5, with conditional costs

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

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

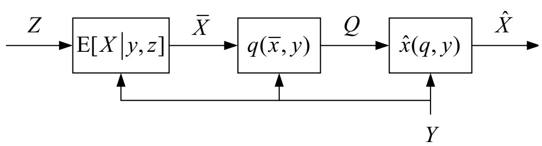

Fig. 5. Optimal implementation of MSE conditional quantization of a noisy source.

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 (2) and (3) are valid for the conditional quantizer of Fig. 5. The validity of (4) for the conditional quantizer can be shown by solving for V in (3) and substituting the result into (2).

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 lattice 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 (2)-(4), 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 $\mathrm{h}(\bar{x}_Y(y) + \bar{X}_Z|y) = \mathrm{h}(\bar{X}_Z|y)$. ■

The case in which X can be written as $X = f(Y) + g(Z) + N$, for any functions $f(y)$ and $g(z)$ 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 transformation and $q(\bar{x}_Z)$ is a lattice quantizer, the overall quantizer $q(\bar{x}_Z(z))$ is also a lattice 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 4: Assume the hypotheses of Theorem 3, 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 3, $q(\bar{x}_Z)$ is a lattice quantizer without index repetition, a translated copy of $q(\bar{x},y)$. ■

Theorem 3 and Corollary 4 show that the WZ quantization setting of Fig. 1 can be implemented as depicted in Fig. 6, 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 lattice quantizer and reconstructor for $\bar{X}_Z$.

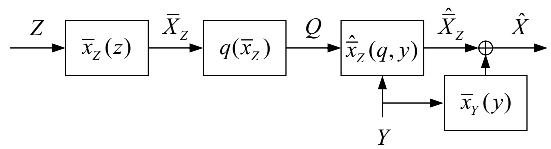

Fig. 6. Asymptotically optimal implementation of MSE WZ quantization of a noisy source with additively separable $\bar{x}(y,z)$.

III. WZ Transform Coding of Noisy Sources

If $\bar{x}(y,z)$ is additively separable, the asymptotically optimal implementation of a WZ quantizer established by Theorem 3 and Corollary 4, illustrated in Fig. 6, suggests the transform coding setting represented in Fig. 7. In this setting, the WZ

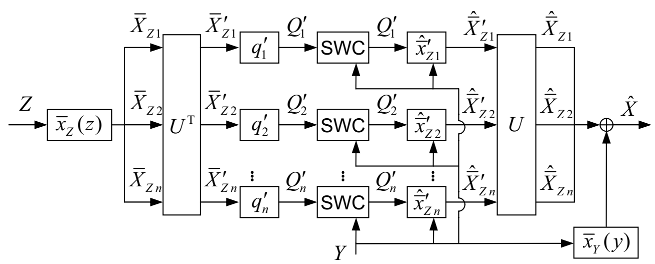

Fig. 7. WZ transform coding of a noisy source.

lattice 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 [7]. The transform coder is a rotated, scaled Z-lattice quantizer, and the translation argument used in the proof of Theorem 3 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 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 vector 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 band. 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} band. $\bar{\mathcal{D}} = \frac{1}{n}\,\mathrm{E}[\|\bar{X}_Z - \hat{\bar{X}}_Z\|^2]$ 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 [7].

Theorem 5 (WZ Transform Coding of Noisy Sources): Suppose $\bar{x}(y,z)$ is additively separable. Assume the hypotheses of [7, Theorem 4] for $\bar{X}_Z$. In summary, assume the high-rate approximation hypotheses for WZ quantization of clean sources hold for each band, 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. 7, 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 [7, Theorem 4] 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 (1) for (Y,Z) instead of Z to prove 2). ∎

Similar to Theorem 3, 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 $\mathrm{E}\,\mathrm{Cov}[\bar{X}_Z|Y] = \mathrm{E}\,\mathrm{Cov}[\bar{X}|Y] \preceq \mathrm{Cov}\,\bar{X}$.

Corollary 6 (Gaussian case): If X, Y and Z are jointly Gaussian, then it is only necessary to assume the high-rate approximation hypotheses of Theorem 5, in order for it to hold. Furthermore, if $\mathcal{D}_{VQ}$ denotes the distortion when the optimal vector quantizer of Fig. 6 is used, then

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

Proof: $\bar{x}(y,z)$ is additively separable. Apply [7, Corollary 5] to $\bar{X}_Z$ and Y, which are jointly Gaussian. ∎

Corollary 7 (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., $((\bar{X}_i - \mathrm{E}[\bar{X}_i|y])|y)_i$ is autocorrelation stationary for each y. Then, it is not necessary to assume in Theorem 5 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 [7, Corollary 6] to $\bar{X}_Z$ and Y. ∎

Observe that the coding performance of the cases considered in Corollaries 6 and 7 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 $f(y)$, $g(y)$ are 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. $\mathrm{Cov}[\bar{X}|y] = \mathrm{Cov}\,N_Z$, thus this is an example of constant conditional variance of transformed coefficients which in addition satisfies the hypotheses of Corollary 7.

It was shown in [7] 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 band, 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 8: 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 [21], and express the mutual information as a difference of entropies. ∎

Theorem 9 (Reduction of side information): Under the hypotheses of Theorem 5 (or Corollaries 6 or 7), a sufficient statistic Y'_i for $\bar{X}'_{Z_i}$ from Y can be used instead of Y for Slepian-Wolf decoding and reconstruction, for each band i in the WZ transform coding setting of Fig. 7, with no asymptotic loss in performance.

Proof: Theorems 3 and 5 imply $\mathcal{R}_i = \mathrm{H}(\bar{X}'_{Zi}|Y) \simeq \mathrm{h}(\bar{X}'_{Zi}|Y) - \log_2 \Delta$. Proposition 8 ensures that $\mathrm{h}(\bar{X}'_{Zi}|Y) = \mathrm{h}(\bar{X}'_{Zi}|Y'_i)$, and Corollary 4 that a suboptimal reconstruction is asymptotically as efficient if Y'_i is used instead of Y. ∎

In view of these results, [7] incidentally shows that in the Gaussian case, the best linear MSE estimate is a sufficient statistic, which can also be proven directly. 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. [22]).

IV. Experimental Results

We implement various cases of WZ transform coding of a noisy image to confirm the theoretical results of Sections II and III. The source data X consists of the first 25 frames of the foreman QCIF video sequence, with the mean removed. Assume that the encoder does not know X, but has access to $Z = X + V$, where $V \sim \mathcal{N}(0, \sigma_V^2)$. The decoder has side information $Y = X + W$, where $W \sim \mathcal{N}(0, \sigma_W^2)$. V and W are independent of each other and of X. In this case, $\mathrm{E}[X|y,z]$ is not additively separable. However, since our theoretical results apply to separable estimates, the estimators are constrained to be linear, and therefore we define $\bar{x}(y,z) = \mathrm{Cov}[X, (Y\ Z)^{\mathrm{T}}]\mathrm{Cov}[(Y\ Z)^{\mathrm{T}}]^{-1}(y\ z)^{\mathrm{T}} = \bar{x}_Y(y) + \bar{x}_Z(z)$.

We consider the following cases, all using estimators and WZ transform coders of clean sources:

1) Assume that Y is made available to the encoder estimator, perform conditional linear estimation of X followed by WZ transform coder of the estimate.
2) Noisy WZ transform coding of Z as shown in Fig. 7.
3) Perform WZ transform coding directly on Z, reconstruct $\hat{Z}$ at the decoder and obtain $\hat{X} = \bar{x}(Y, \hat{Z})$.
4) 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.

Fig. 8 plots rate vs. PSNR for the above cases, with $\sigma_V^2 = \sigma_W^2 = 25$, and $\sigma_X^2 = 2730$ (measured). The performance

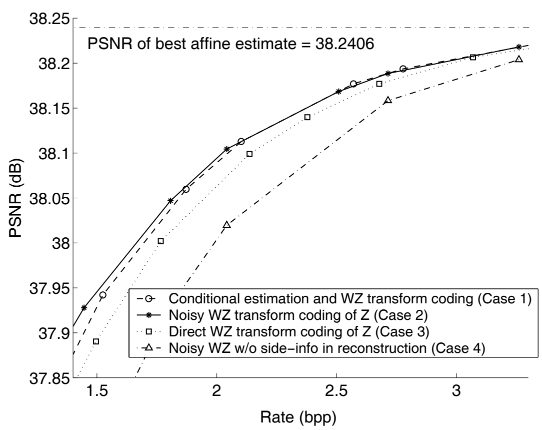

Fig. 8. WZ transform coding of a noisy image is asymptotically equivalent to the conditional case.

of conditional estimation (case 1) and WZ transform coding (case 2) are in close agreement at high rates as predicted by Theorem 5. Our theory does not explain the behavior at low rates. Experimentally, we observed that case 2 slightly outperforms case 1 at small positive rates. Both these cases show better 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 4.

V. Conclusions

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 lattice quantizers for clean sources, achieving the same rate-distortion performance as if the side information where available at the encoder.

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 can be replaced by a sufficient statistic for each of the Slepian-Wolf decoders and reconstruction functions in each band, with no asymptotic loss in performance.

References

[1] J. D. Slepian and J. K. Wolf, "Noiseless coding of correlated information sources," *IEEE Trans. Inform. Theory*, vol. IT-19, pp. 471–480, July 1973.

[2] A. D. Wyner and J. Ziv, "The rate-distortion function for source coding with side information at the decoder," *IEEE Trans. Inform. Theory*, vol. IT-22, no. 1, pp. 1–10, Jan. 1976.

[3] R. Zamir, "The rate loss in the Wyner-Ziv problem," *IEEE Trans. Inform. Theory*, vol. 42, no. 6, pp. 2073–2084, Nov. 1996.

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

[5] S. S. Pradhan and 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, Mar. 2001, pp. 63–72.

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

[7] D. Rebollo-Monedero, A. Aaron, and B. Girod, "Transforms for high-rate distributed source coding," in *Proc. Asilomar Conf. Signals, Syst., Comput.*, Pacific Grove, CA, Nov. 2003, invited paper.

[8] B. Girod, A. Aaron, S. Rane, and D. Rebollo-Monedero, "Distributed video coding," in *Proc. IEEE, Special Issue Advances Video Coding, Delivery*, 2003, invited paper, submitted.

[9] R. L. Dobrushin and B. S. Tsybakov, "Information transmission with additional noise," *IRE Trans. Inform. Theory*, vol. IT-8, pp. S293–S304, 1962.

[10] J. K. Wolf and J. Ziv, "Transmission of noisy information to a noisy receiver with minimum distortion," *IEEE Trans. Inform. Theory*, vol. IT-16, no. 4, pp. 406–411, July 1970.

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

[12] H. Yamamoto and K. Itoh, "Source coding theory for multiterminal communication systems with a remote source," *Trans. IECE Japan*, vol. E63, pp. 700–706, Oct. 1980.

[13] T. Flynn and R. Gray, "Encoding of correlated observations," *IEEE Trans. Inform. Theory*, vol. 33, no. 6, pp. 773–787, Nov. 1987.

[14] H. S. Witsenhausen, "Indirect rate-distortion problems," *IEEE Trans. Inform. Theory*, vol. IT-26, pp. 518–521, Sept. 1980.

[15] J. Gubner, "Distributed estimation and quantization," *IEEE Trans. Inform. Theory*, vol. 39, no. 4, pp. 1456–1459, July 1993.

[16] A. R. W.M. Lam, "Design of quantizers for decentralized estimation systems," *IEEE Trans. Inform. Theory*, vol. 41, no. 11, pp. 1602–1605, Nov. 1993.

[17] W. R. Bennett, "Spectra of quantized signals," Bell Syst., Tech. J. 27, July 1948.

[18] S. Na and D. L. Neuhoff, "Bennett's integral for vector quantizers," *IEEE Trans. Inform. Theory*, vol. 41, pp. 886–900, July 1995.

[19] A. Gersho, "Asymptotically optimal block quantization," *IEEE Trans. Inform. Theory*, vol. IT-25, pp. 373–380, July 1979.

[20] R. M. Gray and D. L. Neuhoff, "Quantization," *IEEE Trans. Inform. Theory*, vol. 44, pp. 2325–2383, Oct. 1998.

[21] T. M. Cover and J. A. Thomas, *Elements of information theory*. New York: Wiley, 1991.

[22] G. Casella and R. L. Berger, *Statistical Inference*, 2nd ed. Australia: Thomson Learning, 2002.