SOURCE CODING WITH INTERMITTENT AND DEGRADED SIDE INFORMATION AT THE DECODER

Francesca Bassi, Michel Kieffer[*]

LSS – CNRS – SUPELEC – Univ Paris-Sud
91192 Gif-sur-Yvette, France

Claudio Weidmann[†]

ftw. Telecommunications Research Center
A-1220 Vienna, Austria

ABSTRACT

Practical schemes for distributed video coding with side information at the decoder need to consider non-standard correlation models in order to take non-stationarities into account. In this paper we introduce two correlation models for Gaussian sources, the Gaussian-Bernoulli-Gaussian (GBG) and the Gaussian-Erasure (GE) models, and evaluate lower and upper bounds on their rate-distortion functions. Provided that the probability of impulse noise or of erasures remains small, these bounds remain close to the rate-distortion function for Gaussian correlation. Two practical schemes for the GE correlation model are also presented, with performance about 1.5 dB away from the upper bound.

Index Terms— Source coding, rate-distortion theory, Wyner-Ziv coding

1. INTRODUCTION

Lossy source coding with side information available only at the decoder [1, Sec. 14.9], also known as Wyner-Ziv (WZ) coding, has received growing attention in the past years, as shown by the number of practical schemes that have been proposed. These schemes are mainly designed either for binary source X and side information Y, or for Gaussian correlation between continuous source and side information, see, *e.g.*, [2, 3, 4]. In practical schemes as video coders exploiting WZ concepts [5, 6] the decoder builds the side information using, *e.g.*, extrapolation of previously decoded frames or interpolation of key frames. As a result, the Gaussian correlation model in the pixel or in the transform domain is not too accurate when, *e.g.*, occlusions occurs, or when new objects appear. This may lead to an increase in the data rate needed to recover X given Y at the decoder.

In this paper we consider two alternative correlation models, the Gaussian Bernoulli-Gaussian (GBG) and the Gaussian-Erasure (GE). GBG correlation models the difference $Y - X$ as a mixture of a small-variance Gaussian noise and a large-variance Gaussian noise, the latter being switched on and off by a Bernoulli process [7]. GBG correlation is a good candidate to represent the correlation between the frame being encoded and the side information obtained by interpolation: some pixels may be well predicted, whereas others have poor prediction, *e.g.*, when an occlusion occurs.

GE correlation is a standard Gaussian correlation model where the side information may be absent for some source samples. It is a good candidate to model the correlation between the frame being encoded and the side information extrapolated (with motion estimation) from previous frames: areas where no side information is available may appear, for example when the camera is moving.

These models bear a similarity with the Heegard and Berger problem [8, 9], which considers efficient coding when side information may or may not be available at the decoder. The main difference

and difficulty here is the fact that the sample positions for which the side information is degraded or absent are random. For GE correlation the problem is simpler, since the decoder knows the erasure positions.

GBG and GE correlation models are formally described in Section 2, before computing lower and upper bounds on the rate-distortion function. Two practical schemes for the GE case are presented in Section 3, with simulation results in Section 4. The connection to distributed compressed sensing is detailed in Section 5, before ending with some conclusions and perspectives.

2. MODELS AND BOUNDS

Let $\{X_j, Y_j\}_{j=1}^{\infty}$ be a sequence of independent drawings of a pair of correlated continuous random variables (X, Y). Assume $X \sim \mathcal{N}(0, \sigma_x^2)$. The correlation is defined assuming a memoryless channel between X and Y.

In the proposed GBG model the correlation channel is affected by the sum of Gaussian noise $G_g \sim \mathcal{N}(0, \sigma_g^2)$ and of Bernoulli-Gaussian noise. Bernoulli-Gaussian noise consists of the product of a Bernoulli process B with parameter p and Gaussian noise $G_i \sim \mathcal{N}(0, \sigma_i^2)$. It is assumed that $\sigma_g^2 \ll \sigma_i^2$. The GBG correlation channel is then defined by

$$Y_{\text{GBG}} = \begin{cases} X + G_g & \text{with probability } (1-p), \\ X + (G_g + G_i) & \text{with probability } p. \end{cases} \quad (1)$$

In the proposed GE model, the correlation channel is still affected by Gaussian noise $G_g \sim \mathcal{N}(0, \sigma_g^2)$. The Bernoulli process now models the occurrence of erasures on the correlation channel. The GE correlation channel is defined by

$$Y_{\text{GE}} = \begin{cases} X + G_g & \text{with probability } (1-p), \\ \Delta & \text{with probability } p. \end{cases} \quad (2)$$

The next two sections present lower and upper bounds on the WZ rate-distortion function for both models.

2.1. Lower bounds

Consider the MSE distortion measure, and let D be the target distortion between the source X and its reconstruction $\widehat{X}$. Intuitively, the rate required by the system can be lower-bounded by assuming that the encoder knows the locations of the samples affected by impulse noise (GBG model) or erased (GE model). This results in a setup similar to [9], with partial side information B (the positions of the erasures or impulses) available at both encoder and decoder. The encoder works in a time-division regime governed by B, employing in each fraction of the time the minimum rate required to match the global distortion constraint D.

For $D \le \sigma_g^2$ the optimal rate allocation results in equal distortions for both fractions of time. Thus for the GBG correlation model the optimal encoder operates on the Gaussian WZ rate-distortion

function $R_{X|Y,\sigma_{\mathrm{g}}^2}(D)$ for noise of variance σ_{g}^2 for a fraction $(1-p)$ of the time, and on the WZ rate distortion function $R_{X|Y,\sigma_{\mathrm{g}}^2+\sigma_{\mathrm{i}}^2}(D)$ for a noise of variance $\sigma_{\mathrm{g}}^2 + \sigma_{\mathrm{i}}^2$ for the remaining:

$$\underline{R}_{\mathrm{GBG}}(D) = (1-p)\, R_{X|Y,\sigma_{\mathrm{g}}^2}(D) + p\, R_{X|Y,\sigma_{\mathrm{g}}^2+\sigma_{\mathrm{i}}^2}(D). \quad (3)$$

A tighter lower bound may be obtained using the approach in [10], where it is assumed that only the receiver knows B (it will only be a lower bound, since in the GBG correlation model neither the sender nor the receiver know B).

Similarly, for the GE correlation model the encoder operates on the Gaussian WZ rate-distortion function for noise of variance σ_g^2 for a fraction $(1-p)$ of the time, while it sends full rate for the description of X for the remaining:

$$\underline{R}_{\mathrm{GE}}(D) = (1-p)\, R_{X|Y,\sigma_{\mathrm{g}}^2}(D) + p\, R_X(D). \quad (4)$$

For $D > \sigma_g^2$ some rate needs to be spent only where impulses or erasures are present in the correlation. The minimum rate required is then $\underline{R}_{\mathrm{GBG}} = p\, R_{X|Y,\sigma_{\mathrm{g}}^2+\sigma_{\mathrm{i}}^2}(D_i)$ for the GBG model, and $\underline{R}_{\mathrm{GE}} = p\, R_X(D_i)$ for the GE model, where D_i has to satisfy the constraint $D = pD_i + (1-p)\sigma_g^2$. For target distortion $D > \sigma_g^2$ the lower bound in figure 1 was determined numerically using reverse water-filling.

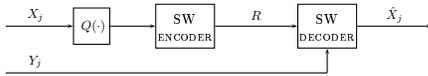## 2.2. Asymptotic upper bound (high-rate analysis)

Fig. 1. Theoretical scheme.

In order to obtain an upper bound on the rate-distortion function for a given target distortion $D \leq \sigma_g^2$ one may consider the theoretical scheme in Figure 1 and assume high-rate regime. The source is scalar-quantized with step ψ, then Slepian-Wolf encoded [1, Sec. 14.8] and sent to the receiver. We assume an ideal Slepian-Wolf encoding scheme with vanishing probability of decoding error. Thus for the GBG correlation model an upper (achievable) bound is, for $D = \psi^2/12$,

$$\overline{R}_{\mathrm{GBG}} = H(U|Y) = h(X|Y) - \log_2 \psi$$
$$\leq (1-p)h(X|Y, B{=}0) + p\, h(X|Y, B{=}1) + H(B) - \log_2 \psi$$

where $h(X|Y)$ denotes the conditional differential entropy. Similarly, for the GE model the upper bound for $D = \psi^2/12$ is

$$\overline{R}_{\mathrm{GE}} = H(U|Y) = h(X|Y) - \log_2 \psi$$
$$= (1-p)h(X|Y{\neq}\Delta) + p\, h(X|Y{=}\Delta) - \log_2 \psi.$$

Figure 2 shows upper and lower bounds obtained for the GBG and GE correlation models for $p = 0.05$, $\sigma_x = 1$, $\sigma_{\mathrm{g}} = 0.2$, $\sigma_{\mathrm{i}} = 1$. The figure also shows that for a GBG or a GE correlation model with small p a performance comparable to standard Gaussian correlation may be obtained.

3. PRACTICAL SCHEMES FOR GE CORRELATION

The remainder of the paper focuses on practical schemes for the GE model. They may be extended to the GBG model, with more sophisticated impulse correction tools, see Section 6.
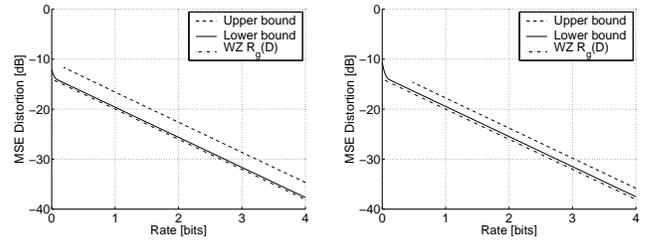

(a) GBG model   (b) GE model.

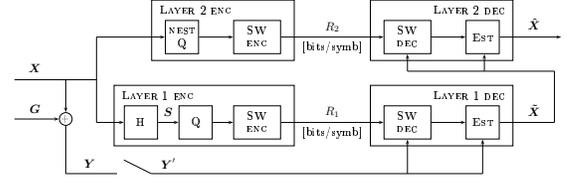**Fig. 2**. Lower and upper bounds of the rate-distortion functions

Fig. 3. Practical scheme for GE model

3.1. Single-layer scheme

A single-layer scheme for both correlation models may be obtained following the ideas presented in Section 2.2 for evaluating the rate-distortion upper bound. By scalar quantization of the source and the side information with the same quantization step ψ one obtains discrete versions of the source and the side information, which are still correlated. The correlation may be represented by a discrete memoryless channel. For large values of ψ many quantization indexes will be equal. Only the entries affected by impulse noise or erased will be different. The indexes may then be efficiently encoded using a Slepian-Wolf scheme, such as [11].

The main disadvantage of this scheme is that it is intrinsically linked to scalar quantization. Using a vector quantizer would, for a cell size comparable to the one of the cubic lattice, increase the number of indexes which are different, and thus increase the rate on the channel without decreasing the distortion.

3.2. Two-layer scheme

The two-layer encoding scheme for GE correlation depicted in Figure 3 is proposed as an alternative to the single-layer scheme. Both layers operate on blocks of N i.i.d. source samples. Layer 1 aims to produce a coarse estimation of the source, $\tilde{X}$, compensating for the entries erased by the correlation channel. Layer 2 regards $\tilde{X}$ as side information, employing standard Wyner-Ziv coding techniques to provide the finer estimate $\hat{X}$ within target distortion per source symbol D. The rate required by the system is $R = R_1 + R_2$ bits per source symbol. Let $X \sim \mathcal{N}(0, \sigma_x^2 I_N)$ be the source sequence, $G \sim \mathcal{N}(0, \sigma_g^2 I_N)$ the Gaussian noise sequence, and Y' the partially erased side information sequence.

Layer 1 encoder consists of a precoding block followed by a quantizer. The aim of precoding is to reduce the dimension of the sequence to be conveyed by a factor $\frac{N-K}{N}$, while preserving the information needed to estimate the erased entries. Precoding consists in applying some linear transform $H_{(N-K)\times N}$, whose rows are required to be orthonormal. The output of the precoder, $S = H \cdot X$, is distributed as $\mathcal{N}(0, \sigma_x^2 I_{N-K})$; each sample is quantized using a uniform quantizer. For values $R_1 > \frac{(N-K)}{N} \cdot 2.5$ bits, the rate on Layer 1 for target distortion D_q per source sample is well approximated by the high-rate expression

$$R_1(D_q) = \frac{N-K}{N}\left(\frac{1}{2}\log_2\left(2\pi e \sigma_x^2\right) - \frac{1}{2}\log_2\left(12 D_q\right) \right) \quad (5)$$

bits per source sample, see [12]. The quantization indexes are then Slepian-Wolf encoded.

The same transformation and quantization are applied to the side information at the decoder to reconstruct the quantized transformed samples. Using the transmitted bits it is possible to recover an estimate of S. The decoder exploits all the information available on the source sequence to produce the estimate $\tilde{X}$. The observation vector is expressed as

$$A = \begin{bmatrix} S \\ V \cdot Y \end{bmatrix} = \begin{bmatrix} H \\ V \end{bmatrix} \cdot X + \begin{bmatrix} Q \\ V \cdot G \end{bmatrix} = F \cdot X + Z, \quad (6)$$

where $V_{(N-\nu)\times N}$ is obtained from I_N by eliminating the ν lines corresponding to the erased positions in Y', and Q is the quantization noise vector; the covariance matrix of Z is

$$\Gamma_z = \begin{bmatrix} D_q I_{N-K} & 0 \\ 0 & \sigma_g^2 I_{N-\nu} \end{bmatrix}. \quad (7)$$

The distortion per source symbol on Layer 1 output is

$$D_1 = \frac{1}{N} E[\| X - \tilde{X} \|^2] = \frac{1}{N} \operatorname{trace}(\Gamma_{x-\tilde{x}}), \quad (8)$$

where $\Gamma_{x-\tilde{x}} = E[(X - \tilde{X})(X - \tilde{X})^T]$ is the covariance matrix of the estimation error. Layer 2 may then use any Wyner-Ziv coding scheme for Gaussian correlation. Here we consider nested scalar quantization followed by a Slepian-Wolf coder, described in [4]. The output of the nested quantizer is the coset leader of the source X, whose correlation with the side information $\tilde{X}$ is exploited to achieve compression. As in [4] we assume theoretical Slepian-Wolf rate. In order to match the distortion constraint D, Layer 2 requires rate

$$R_2 = \frac{1}{2} \log_2(2\pi e D_1) - \frac{1}{2} \log_2(12D) \quad (9)$$

bits per source sample.

The overall distortion per source symbol of the estimate $\hat{X}$ is derived from (5), (9), and (12) as

$$D(R_1, R_2) = \frac{1}{N} \frac{2\pi e}{12} \operatorname{trace}(\Gamma_{x-\tilde{x}}(R_1)) \, 2^{-2R_2}. \quad (10)$$

A MAP estimator is used, since it allows the estimation of $\tilde{X}$ even when more than $N - K$ samples are erased. Its expression is

$$\tilde{X} = \tilde{X}_{\text{MAP}} = \left(F^T \Gamma_z^{-1} F + K_x^{-1}\right)^{-1} F^T \cdot A; \quad (11)$$

the covariance matrix of the estimation error is

$$\Gamma_{\text{MAP}} = E[(\tilde{X}_{\text{MAP}} - X)(\tilde{X}_{\text{MAP}} - X)^T] \quad (12)$$
$$= \left(D_q^{-1}(R_1)H^T H + \sigma_g^{-2} V^T V + K_x^{-1}\right)^{-1}.$$

3.3. High-rate regime

The overall distortion (10) decreases as R_2 increases; the dependence on the rate R_1 is related to (8) through the covariance matrix $\Gamma_{x-\hat{x}}$. Using (5) in (12) one obtains

$$\Gamma_{\text{MAP}} = \left(\alpha^{-1} 2^{\beta R_1} H^T I_{N-K} H^T + \Delta\right)^{-1}, \quad (13)$$

where $\alpha = \frac{2\pi e \sigma_x^2}{12}$, $\beta = 2\frac{N}{N-K} > 0$ and $\Delta = \sigma_g^{-2} V^T V + \sigma_x^{-2} I_N$. Using the matrix inversion lemma (13) becomes

$$\Gamma_{\text{MAP}} = \Delta^{-1} - \Delta^{-1} H^T \left(\alpha 2^{-\beta R_1} I + H\Delta^{-1} H^T\right)^{-1} H\Delta^{-1}. \quad (14)$$
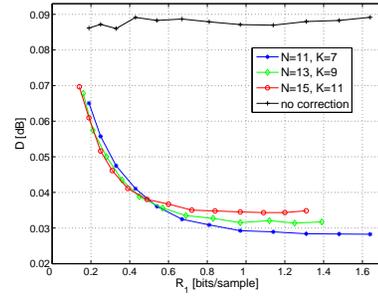

Fig. 4. Performance of Layer 1 (GE correlation)

For increasing R_1, the term depending on R_1 in (14) becomes negligible and the expression of Γ_{MAP} becomes

$$\Gamma_{\text{MAP}s} = \lim_{R_1 \to +\infty} \Gamma_{\text{MAP}}$$
$$= \Delta^{-1} - \Delta^{-1} H^T \left(H\Delta^{-1} H^T\right)^{-1} H\Delta^{-1}. \quad (15)$$

Thus, for R_1 sufficiently large, but still finite, the asymptotic overall distortion depends only on R_2:

$$D(R_1, R_2) = D(R_2) = \frac{1}{N} \frac{2\pi e}{12} \operatorname{trace}(\Gamma_{\text{MAP}s}) \, 2^{-2R_2}. \quad (16)$$

4. SIMULATIONS

We consider here the GE correlation model, with the same parameters as in Section 2.2. The estimate $\tilde{X}$ for Layer 1 is obtained using (11). H is obtained by extracting $N - K$ rows of a size N discrete Fourier transform matrix. The size N and the $N - K$ rows are chosen such that row i and row $N - K + 1 - i$ are hermitian symmetric. Thus, for a real-valued vector, after a multiplication by H, only $N - K$ real entries need to be transmitted.

Figure 4 shows the performance of Layer 1 of the coding scheme in Figure 3, for codes with fixed $N - K$, and varying N. For $p = 0.05$ the choice of $N - K = 4$ provides enough information to estimate the erased entries of the side information (the probability of having more than 4 erased entries per block is about 10^{-4}). The influence of the increase of R_1 becomes negligible for $R_1 > 0.7$ bits per source sample (after ideal Slepian-Wolf encoding), as expected from (14).

Layer 2 is now appended to Layer 1 (R_1 is fixed at 0.75 bits per source sample). The performance of the entire coding scheme is represented in Figure 5. The width of the coarse quantization cell is taken as 4, in order to guarantee a vanishing probability of Wyner-Ziv decoding error. The rate is the sum of R_1 and the rate at the output of the ideal Slepian-Wolf coder of Layer 2.

Figure 5 also shows the performance of the single-layer coding scheme (ideal Slepian-Wolf encoding is also assumed in this case).
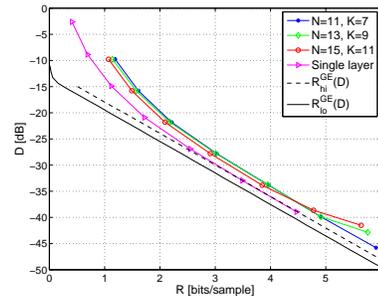

Fig. 5. Comparison of the performance of the two proposed coding schemes (GE correlation)

At high rates, the single-layer scheme reaches the upper bound. The two-layer scheme performs about 1.5 dB away from the upper bound of the rate-distortion curve (code $N = 15$, $K = 11$). A floor can be observed at high rates, for longer codes ($N = 13$, $N = 15$). This is due to the increased probability of more than $N - K$ erasures per sequence. The performance of the two-layer scheme may be improved by using a more efficient Wyner-Ziv encoder in Layer 2 (vector quantization), which is not possible for the single-layer scheme.

5. CONNECTIONS WITH COMPRESSED SENSING

We briefly outline how WZ schemes for GE and GBG correlation can be used for compressed sensing (CS) [13, 14]. A bare bones example of a CS problem is to compress a signal vector $\mathbf{x} \in \mathbb{R}^N$ of the form $\mathbf{x} = \Psi \mathbf{u}$, where Ψ is an orthonormal N-by-N matrix and $\mathbf{u} \in \mathbb{R}^N$ has at most K non-zero components (we say that $\mathbf{x}$ is K-sparse with respect to Ψ). The problem is to determine a compression mechanism for $\mathbf{x}$ when only the decoder knows the sparsifying basis Ψ. Distributed CS (DCS) extends this paradigm to multiple correlated signals, which may be composed of sparse and non-sparse components, and have to be encoded independently [15].

In the simple case $\Psi = I_N$, when $\mathbf{x} = \mathbf{u}$ is directly sparse, there is an immediate connection with the GBG model: model $\mathbf{x}$ as a vector of i.i.d. samples of GBG noise with $p = K/N$ and σ_i^2 the variance of the sparse signal components, while $\sigma_g^2 << \sigma_i^2$ is some background noise (this actually generalizes the concept of K-sparsity). Then a CS scheme can be built using a GBG-WZ encoder for i.i.d. $X \sim \mathcal{N}(0, \sigma_g^2 + p\sigma_i^2)$. The decoder simply uses the trivial side information $Y = 0$ to decode the quantized value $\hat{\mathbf{x}}$ (notice that in this scheme, the roles of source and side information are reversed).

An extension to distributed CS is also possible, e.g., for model JSM-3 in [15], which considers M distributed sources $X_j = Z_C + Z_j$, where Z_C is a possibly non-sparse common component and Z_j is an independent (form other sources) sparse innovation component. Each sensor (source) then computes a two-part code: first, it applies a standard quantizer of rate R_1 to obtain $\hat{X}_j$. Second, it uses a rate R_2 WZ code for X_j given side information $\hat{Z}_{C,j}$, which the decoder computes from the $\hat{X}_j$ as $\hat{Z}_{C,j} = \frac{1}{M-1} \sum_{i=1, i \neq j}^{M} \hat{X}_i$. If component Z_j is modeled by independent realizations of GBG noise, a GBG-WZ encoder can be used to encode X_j given side information $\hat{Z}_{C,j}$. The larger the number of sources M, the smaller the rate R_1 needed to reduce the noise on the side information, $\hat{Z}_{C,j} - Z_C$. If the Z_j are i.i.d. with the same GBG distribution, this noise will also be GBG, with the same p and correspondingly scaled variances.

In the more general case, when $\mathbf{x}$ is sparse in a non-trivial basis $\Psi \neq I_N$, this GBG-WZ approach can only be used if the codebook is invariant under the rotation Ψ, which allows to encode without knowing Ψ. This will be the case at high rates ($D < \sigma_g^2$), when the optimal codebook is Gaussian and thus spherically symmetric. The simple scheme with scalar nested quantization followed by SW coding does not have this symmetry property; however, it is a useful tool to overbound the optimal WZ rate-distortion performance.

6. CONCLUSIONS

We considered two new correlation models (GBG and GE) for the problem of source coding with side information at the decoder, and provided bounds on the rate-distortion function. Two practical encoding schemes have been proposed for the GE correlation model. At high rate, the two schemes perform quite similarly and are about 1.5 dB away from the upper bound on rate-distortion.

The main advantage of employing a two-layer scheme compared to a single-layer scheme is that it allows to use more sophisticated Wyner-Ziv coding techniques, e.g., TCQ followed by LDPC-based stage. Moreover, it may be much more profitable in the case of a colored correlation noise G: the single-layer scheme is restricted to a scalar quantizer, whereas the second layer of the two-layer scheme may use a vector quantizer to efficiently encode G. Finally, the two-layer scheme is easily adaptable to varying erasure probability p, since the second layer will not be affected.

A similar two-layer scheme may be used for the GBG correlation model. The first layer, at decoder side, needs to estimate the number, locations and amplitudes of the impulse noise spikes. Techniques used in compressed sampling may be used for that purpose [13, 14]; decoding techniques for BCH codes on the real field may also be efficiently put at work [7].

References

[1] T. M. Cover and J. M. Thomas, *Elements of Information Theory*, Wiley, New–York, 1991.

[2] A.D. Liveris, Z. Xiong, and C.N. Georghiades, "Nested convolutional/turbo codes for the binary Wyner-Ziv problem," in *Proc. IEEE ICIP*, Sept. 2003, pp. 601–604.

[3] Y. Yang, S. Cheng, Z. Xiong, and W. Zhao, "Wyner-Ziv coding based on TCQ and LDPC codes," *37th Asilomar Conf. on Signals, Systems and Computers*, pp. 825–829, Nov. 2003.

[4] Z. Liu, S. Cheng, A.D. Liveris, and Z. Xiong, "Slepian-Wolf coded nested lattice quantization for Wyner-Ziv coding: High-rate performance analysis and code design," *IEEE Trans. Inform. Theory*, vol. 52, no. 10, pp. 4358–4379, Oct. 2006.

[5] B. Girod, A. Aaron, S. Rane, and D. Rebollo-Monedero, "Distributed video coding," *Proc. of the IEEE*, vol. 1, no. 71, pp. 93, Jan. 2005.

[6] R. Puri and K. Ramchandran, "PRISM: A video coding paradigm based on motion–compensated prediction at the decoder," *IEEE Trans. Image Processing*, vol. 16, no. 10, pp. 2436–2448, 2007.

[7] A. Gabay, M. Kieffer, and P. Duhamel, "Joint source-channel coding using real BCH codes for robust image transmission," *IEEE Trans. Signal Processing*, vol. 16, no. 6, pp. 1568–1583, 2007.

[8] C. Heegard and T. Berger, "Rate distortion when side information may be absent," *IEEE Trans. Inform. Theory*, vol. 31, no. 6, pp. 727–734, Nov. 1985.

[9] M. Fleming and M. Effros, "On rate–distortion with mixed types of side information," *IEEE Trans. Inform. Theory*, vol. 52, no. 4, pp. 1698–1705, Apr. 2006.

[10] C. T. K. Ng, C. Tian, A. J. Goldsmith, and S. Shamai, "Minimum expected distortion in Gaussian source coding with uncertain side information," in *Proc. Information Theory Workshop*, Lake Tahoe, CA, USA, Sept. 2–6 2007, pp. 454–459.

[11] Z. Tu, J. Li, and R. S. Blum, "An efficient SF–ISF approach for the Slepian–Wolf source coding problem," *Eurasip Journal on Applied Signal Processing*, vol. 6, pp. 961–971, May 2005.

[12] J. Zhuang, V. K. Goyal, and M. Vetterli, "Transform coding with backward adaptive updates," *IEEE Trans. on Information Theory*, vol. 46, no. 4, pp. 1623–1633, 2000.

[13] D. L. Donoho, "Compressed sensing," *IEEE Trans. Inform. Theory*, vol. 52, no. 4, pp. 1289–1306, April 2006.

[14] E. Candès and T. Tao, "Near-optimal signal recovery from random projections and universal encoding strategies?," *IEEE Trans. Inform. Theory*, vol. 52, pp. 5406–5425, Dec. 2006.

[15] M. Wakin, M. Duarte, S. Sarvotham, D. Baron, and R. Baraniuk, "Recovery of jointly sparse signals from few random projections," in *Proc. NIPS*, Dec. 2005.