WYNER-ZIV CODING WITH UNCERTAIN SIDE INFORMATION QUALITY

Francesca Bassi[1], *Michel Kieffer*[1,2], *Claudio Weidmann*[3]

[1]LSS – CNRS – SUPELEC – Univ Paris-Sud, 3 rue Joliot-Curie, 91192 Gif-sur-Yvette, France
[2] on sabbatical leave at LTCI - CNRS - Télécom ParisTech, 75014 Paris, France
[3] INTHFT, Vienna University of Technology, Gusshausstrasse 25/389, 1040 Vienna, Austria

ABSTRACT

Wyner-Ziv coding of continuous sources with uncertain side information quality is defined by modeling the correlation noise as a Gaussian-mixture. The analysis of the theoretical rate-distortion performance is presented, along with a coding solution not relying on the presence of a feedback channel. The attainable performance of the coding scheme is derived, and a brief discussion on implementation issues concludes the paper.

1. INTRODUCTION

In recent years Wyner-Ziv coding has captured major attention due to its suitability to emerging applications such as wireless sensor networks [1] and distributed video coding [2]. Investigation on practical solutions began with the contribution of DISCUS [3], and spread in the last decade producing innumerable works (see [4] for a survey). As emphasized in [5], the main issue in practical design is the robustness of the scheme against the variations in time of the side information quality, which are usually not captured by the underlying theoretical model, but strongly affect the performance. The classical approach to the problem is the attempt to track the instantaneous quality of the correlation channel: either introducing a feed-back link, thus enabling rate adaptation [6], or allowing the encoder to process previously transmitted data to facilitate the estimation of the correlation noise level [7].

Relatively few works investigate the information theoretical consequences of distributed compression with uncertain side information quality. Heegard and Berger [8] and Kaspi [9] considered the event that the side information might be absent at the decoder, without the encoder being aware. Recent works by Verdú and Weissman [10] and Perron *et al.* [11] address the evaluation of the Wyner-Ziv rate-distortion function when the side information either is a noiseless copy of the source or is erased. The authors introduced a different correlation model in [12, 13], where the side information is a copy of the source affected by background noise and, occasionally, by additive noise impulses. The decoder is uninformed of the state of the correlation channel, whereas in the Wyner-Ziv setup for possibly erased side information [10–12] the receiver can detect the erasures.

This work generalizes the correlation model presented in [12, 13] and outlines a theoretical framework for the analysis of Wyner-Ziv coding with uncertain side information quality. The instantaneous state of the correlation channel is described by a hidden random variable; the probability density function of the correlation noise results in a Gaussian mixture, with weighting coefficients determined by the state probability distribution.

After derivation of lower and upper bounds to the Wyner-Ziv rate-distortion function, a coding architecture not relying on the presence of a feedback channel is proposed. Its attainable performance is characterized, and implementation details are discussed in the concluding section.

F. Bassi is also with DEIS, Università di Bologna, Italy. This work was partly supported by the European Commission in the framework of the FP7 Network of Excellence in Wireless COMmunications NEWCOM++.

2. MODEL AND THEORETICAL BOUNDS

The standard Gaussian Wyner-Ziv setup [14] considers the compression of a memoryless source $X \sim \mathcal{N}(0, \sigma_x^2)$, when the side information $Y \sim \mathcal{N}(0, \sigma_Y^2)$ is available at the decoder. The correlation of the jointly Gaussian pair (X, Y) is modeled by a virtual channel $Y = X + Z$, where $Z \sim \mathcal{N}(0, \sigma_z^2)$ is independent additive noise. The model can be generalized by assuming the variance σ_z^2 as varying with time. The instantaneous state of the correlation channel is described by a random variable S, taking values in the alphabet $\mathcal{S} = \{0, 1, \dots, T-1\}$. The set $\mathcal{S}$ is mapped one-to-one onto $\mathcal{R} = \{\sigma_0^2, \sigma_1^2, \dots, \sigma_{T-1}^2\}$, where $\sigma_s^2 < \sigma_{s+1}^2$, so that the realization $S = s$ represents the event of correlation noise with variance σ_s^2. Let $p_s = \Pr(S = s)$. The probability density function $f_Z(z)$ of the correlation noise Z is the Gaussian mixture

$$f_Z(z) = \sum_{s=0}^{T-1} p_s \, f_{Z|S=s}(z) = \sum_{s=0}^{T-1} \frac{p_s}{\sqrt{2\pi\sigma_s^2}} \exp\left(-\frac{z^2}{2\sigma_s^2}\right). \quad (1)$$

Gaussian-mixture models are a widely deployed technique for the estimation of the probability density function of natural signals [15]. A wide range of correlation channels can be approximated by a suitable choice of the number of components, of their variances, and of their mixing weights p_s. Without loss of generality, examples and coding schemes will be discussed in detail only for a two-component mixture, introduced in [12] as the Gaussian-Bernoulli-Gaussian (GBG) model, given by $\mathcal{S} = \{0, 1\}$, $\mathcal{R} = \{\sigma_0^2, \sigma_1^2\}$, $p_1 = p$ and $p_0 = (1 - p)$. Section 2.1 explores quantization with decoder side information; a limit-achieving quantizer is derived, which will be used for bounding purposes in Section 2.2.

2.1 High-rate quantization with decoder side information

Wyner-Ziv coding can be understood as a quantization problem. The decoder side information helps in the reconstruction of the source, and allows to compress the quantization indexes through Slepian-Wolf coding, which plays the same role as entropy coding in the standard quantization problem. Necessary conditions for the existence of limit-achieving Wyner-Ziv quantizers are derived in [4, 16]. We provide here an incremental result, along with a brief sketch of the proof. A system composed by a dithered lattice quantizer [17] followed by an ideal Slepian-Wolf chain satisfies the conditions in [4, 16], and hence is Wyner-Ziv optimal, as the length n of the source sequence $\boldsymbol{X}$ grows towards infinity. The rate of the dithered lattice quantizer with side information at the decoder is derived mirroring the proof for standard quantization in [17], and is expressed, in bits per sample, as

$$R_q = \frac{1}{n} h(\boldsymbol{X}|\boldsymbol{Y}) - \frac{1}{n} \log_2 \left(V(\Lambda)\right), \quad (2)$$

where $V(\Lambda)$ is the volume of the fundamental Voronoi cell of the lattice Λ. As $n \to \infty$ the shape of the best tessellation associated with Λ approaches an Euclidean ball, and the second moment approaches $\sigma^2(\Lambda) = V(\Lambda)^{\frac{2}{n}}/2\pi e$. Substitution into (2) yields

$$R_q = R_q\left(\sigma^2(\Lambda)\right) = h(X|Y) - \frac{1}{2}\log_2\left(2\pi e\sigma^2(\Lambda)\right). \qquad (3)$$

The second moment $\sigma^2(\Lambda)$ represents the quantization noise variance per sample. Under high-resolution assumption (the probability density function of the source can be assumed constant over the quantization cell) it is $D = \sigma^2(\Lambda)$. Substitution in (3) completes the proof.

2.2 Rate-distortion bounds

The Wyner-Ziv rate-distortion function is defined as

$$R^{\mathrm{WZ}}_{X|Y}(D) = \min_{(f_{U|X}, F_D)\in\mathcal{M}(D)} \Big(I(X;U) - I(Y;U)\Big) \qquad (4)$$

where U is an auxiliary random variable, and $\mathcal{M}(D)$ is the set of all probability density functions $f_{U|X}$ and matched reconstruction functions $F_D : \mathcal{U}\times\mathcal{Y} \to \hat{\mathcal{X}}$ satisfying the distortion constraint $E[(X-\widehat{X})^2] \le D$. The Wyner-Ziv rate-distortion function is known in closed form only in very few cases: for binary source with binary symmetric correlation channel and for Gaussian source with Gaussian additive correlation channel [14], and for discrete sources with erasure correlation channel [10, 11]. In order to characterize the analytical form of (4) for uncertain side information quality we introduce the bounds

$$\underline{R}^{\mathrm{WZ}}_{X|Y}(D) \le R^{\mathrm{WZ}}_{X|Y}(D) \le \overline{R}^{\mathrm{WZ}}_{X|Y}(D). \qquad (5)$$

The lower bound $\underline{R}^{\mathrm{WZ}}_{X|Y}(D)$ is determined as the best performance achievable by a genie-aided system, where the genie informs both the encoder and the decoder of the realization of S. The genie-aided system works in time-division regime, matching the best achievable performance associated with the event $S = s$, given by the Gaussian Wyner-Ziv rate-distortion function for correlation noise of variance σ_s^2. Like for the compression of parallel Gaussian sources [18], the system has to satisfy the same distortion constraint D for each transmission instant, under the high-rate assumption

$$D < \frac{\sigma_x^2\,\sigma_0^2}{(\sigma_x^2+\sigma_0^2)}. \qquad (6)$$

If (6) holds it is

$$\underline{R}^{\mathrm{WZ}}_{X|Y}(D) = \sum_{s=0}^{T-1} p_s\,\frac{1}{2}\log_2\left(\frac{\sigma_x^2\,\sigma_s^2}{(\sigma_x^2+\sigma_s^2)D}\right). \qquad (7)$$

The upper bound $\overline{R}^{\mathrm{WZ}}_{X|Y}(D)$ under the assumption (6) is determined as the achievable performance of a system composed by a lattice dithered quantizer Λ followed by an ideal Slepian-Wolf coding chain. Under the hypothesis $n\to\infty$ the rate R_q of the lattice quantizer with side information at the decoder is obtained from (3)

$$\begin{aligned} R_q &= h(X|Y) - \frac{1}{n}\log_2\left(2\pi e\sigma^2(\Lambda)\right)\\ &\le h(X|Y,S) - \frac{1}{2}\log_2\left(2\pi e\sigma^2(\Lambda)\right) + H(S), \end{aligned} \qquad (8)$$

where (8) is obtained developing $I(X;S|Y)$. Expressing the differential entropy in (8) as $h(X|Y,S) = \sum_{s=0}^{T-1} p_s\, h(X|Y,S=s)$ and replacing $\sigma^2(\Lambda)$ with D finally yields the upper bound

$$\overline{R}^{\mathrm{WZ}}_{X|Y}(D) = \sum_{s=0}^{T-1} p_s\,\frac{1}{2}\log_2\left(\frac{\sigma_x^2\,\sigma_s^2}{(\sigma_x^2+\sigma_s^2)D}\right) + H(S). \qquad (9)$$

3. THEORETICAL CODING SCHEME

The blindness of both the encoder and the decoder to the realization of S prevents, in the uncertain side information quality setup, all attempts to perform rate-adaptation: the system is forced to operate at constant rate. The simplest solution is obtained employing a standard Gaussian Wyner-Ziv coding scheme, dimensioned for the worst-case quality over the correlation channel, *i.e.* $S = T-1$. The system is clearly suboptimal, since rate loss is experienced for any event $S \ne T-1$. The loss becomes smaller as the variances of the components in (1) tend to be similar, vanishing as $\sigma_s^2 \to \sigma_{s+1}^2$, $\forall s \in \mathcal{S}$, while it dramatically impacts the performance when the correlation noise presents occasional high-variance impulses, so that $\sigma_{T-1}^2 \gg \sigma_s^2$ and $p_{T-1} < p_s, \forall s \in \{0,\dots,T-2\}$.

The alternative coding solution proposed here still relies on fundamental coding blocks optimized for the Gaussian correlation model, but allows to restrain the rate loss also in the most penalizing case. The coding architecture is based on two layers, where suboptimality is confined to Layer 1. The Layer 1 coding chain is exploited to produce a first, rough estimate of the source, regarded as Gaussian side information for the optimal (in the Wyner-Ziv sense) transmission on Layer 2. The two-layer coding approach has been first proposed in [12] for the case of possibly erased side information. In [12] Layer 1 sends the syndrome of a real-field code, used to correct the degraded side information sequence; the decoding process relies on the availability of the erasure pattern at the receiver side. The same solution cannot be employed here, due to the blindness of the decoder to S: an alternative coding scheme (depicted in Figure 1) needs to be designed. The remainder of this section derives its asymptotically achievable rate-distortion performance.

3.1 Encoding

The coding scheme works on a sequence $\boldsymbol{X}$ of n source symbols. Let L be an $[n\times n]$ orthonormal matrix. The rank m matrix Φ consists of a selection of $m \le n$ rows of L; similarly, the rank $(n-m)$ matrix $\overline{\Phi}$ is obtained from the remaining $(n-m)$ rows, so that

$$\mathrm{L} = \mathrm{P}\begin{pmatrix}\overline{\Phi}\\ \Phi\end{pmatrix} \qquad (10)$$

with P a permutation matrix. The source $\boldsymbol{X}$ is multiplied at the encoder side by Φ and $\overline{\Phi}$, generating the Layer 1 sequence $\Phi\boldsymbol{X}$ and the Layer 2 sequence $\overline{\Phi}\boldsymbol{X}$. Since $\boldsymbol{X}$ is i.i.d. Gaussian, it has a spherically symmetric distribution, which is not affected by the rotation through L. The sequences $\Phi\boldsymbol{X}$ and $\overline{\Phi}\boldsymbol{X}$ are selections of m and $(n-m)$ components of $\mathrm{L}\boldsymbol{X}$ respectively, hence the distribution of their symbols can be modeled as Gaussian with variance σ_x^2.

The encoder performs vector quantization of the Layer 1 and Layer 2 sequences by means of the dithered lattice quantizers $\Lambda_1 \in \mathbb{R}^m$ and $\Lambda_2 \in \mathbb{R}^{(n-m)}$, generating $\Phi\boldsymbol{X} + \boldsymbol{N}_1$ and $\overline{\Phi}\boldsymbol{X} + \boldsymbol{N}_2$. The quantization indexes are ideally Slepian-Wolf compressed and sent to the receiver. The per source sample rates R_1 and R_2 are evaluated, using (2), as

$$R_1 = \frac{1}{n}\,h(\Phi\boldsymbol{X}|\Phi\boldsymbol{Y}) - \frac{1}{n}\log_2\Big(V(\Lambda_1)\Big) \qquad (11)$$

$$R_2 = \frac{1}{n}\,h(\overline{\Phi}\boldsymbol{X}|\overline{\Phi}\widetilde{\boldsymbol{X}}) - \frac{1}{n}\log_2\Big(V(\Lambda_2)\Big), \qquad (12)$$

where $\widetilde{\boldsymbol{X}}$ is the output of the Layer 1 decoder, and $\Phi\boldsymbol{Y}$ and $\overline{\Phi}\widetilde{\boldsymbol{X}}$ are the Slepian-Wolf side information sequences for Layer 1 and Layer 2, respectively. Since the sequences $\Phi\boldsymbol{X}$ and $\overline{\Phi}\boldsymbol{X}$ have correlated components, in general it is $h(\Phi\boldsymbol{X}|\boldsymbol{Y}) \le h(\Phi\boldsymbol{X}|\Phi\boldsymbol{Y})$ and $h(\overline{\Phi}\boldsymbol{X}|\widetilde{\boldsymbol{X}}) \le h(\overline{\Phi}\boldsymbol{X}|\overline{\Phi}\widetilde{\boldsymbol{X}})$. It can be proved, however, that equality holds if the pair $(\boldsymbol{X}, \widetilde{\boldsymbol{X}})$ is jointly Gaussian, hence no loss is introduced on Layer 2 employing $\overline{\Phi}\widetilde{\boldsymbol{X}}$ as Slepian-Wolf side information (see [19]). Assuming Layer 1 decoder output $\widetilde{\boldsymbol{X}}$ as composed by identically distributed Gaussian symbols of variance $\sigma_x^2 - D'$,

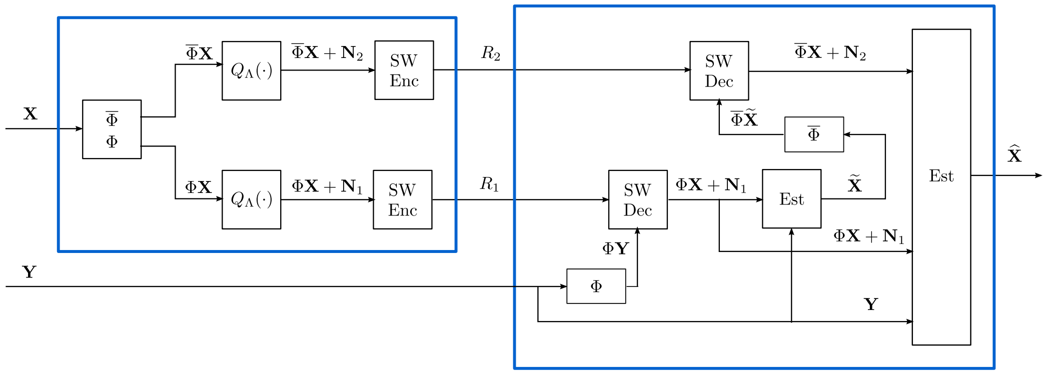

Figure 1: Two-layer coding scheme for Wyner-Ziv coding with uncertain side information quality.

and $(\boldsymbol{X}, \widetilde{\boldsymbol{X}})$ a jointly Gaussian pair, (12) yields the Layer 2 transmission rate as

$$R_2 = \frac{(n-m)}{2n} \log_2 \left(\frac{\sigma_x^2 \, D'}{(\sigma_x^2 + D')\sigma^2(\Lambda_2)} \right). \tag{13}$$

In order to determine R_1 characterization of the noise $\Phi \boldsymbol{Z} = \Phi \boldsymbol{Y} - \Phi \boldsymbol{X}$ is needed. The sequence $\boldsymbol{\zeta} = \Phi \boldsymbol{Z}$ is composed of m statistically dependent symbols

$$\zeta_j = \boldsymbol{\phi_j} \, \boldsymbol{Z} = \sum_{i=1}^{n} \phi_{j,i} \, z_i, \qquad \forall j \in \{1, 2, \ldots, m\}. \tag{14}$$

Denote $\pi_{\boldsymbol{s}} = \Pr(\boldsymbol{S} = \boldsymbol{s}) = \prod_{i=1}^{n} p_{s_i}$. The marginal probability density function $f_j(\zeta_j)$ is given by

$$\begin{aligned} f_j(\zeta_j) &= \sum_{\boldsymbol{s}} \pi_{\boldsymbol{s}} \, f(\zeta_j | \boldsymbol{s}) \qquad (15) \\ &= \sum_{\boldsymbol{s}} \pi_{\boldsymbol{s}} \big(f(\phi_{j,1} z_1 | s_1) * \ldots * f(\phi_{j,n} z_n | s_n) \big), \end{aligned}$$

where the convolution product is possible because the random variables $(\phi_{j,i} z_i | s_i)$ are independent. Since $(\phi_{j,i} z_i | s_i)$ is normally distributed, (15) results in

$$f_j(\zeta_j) = \sum_{\boldsymbol{s}} \pi_{\boldsymbol{s}} \, \frac{1}{\sqrt{2\pi\sigma_{\boldsymbol{s}}^2}} \exp\left(-\frac{\zeta_j^2}{2\sigma_{\boldsymbol{s}}^2} \right), \tag{16}$$

where the variance $\sigma_{\boldsymbol{s}}^2$ is given by

$$\sigma_{\boldsymbol{s}}^2 = \sum_{i=1}^{n} \phi_{ji}^2 \, \sigma_{z_i|\boldsymbol{s}}^2. \tag{17}$$

The j-th element of the noise sequence $\Phi \boldsymbol{Z}$ is marginally distributed as a Gaussian mixture, where the variances of the components depend on the associated rows $\boldsymbol{\phi}_j$ of Φ. This implies that some rate loss is induced by the use of $\Phi \boldsymbol{Y}$ as the Slepian-Wolf side information at the Layer 1 decoder. Assuming the linear transform matrix L in (10) to be a Discrete Cosine Transform (DCT) matrix, it is

$$\phi_{i,j} = \sqrt{\frac{2}{n}} \cos\left(\frac{(2j-1)(i-1)}{n} \frac{\pi}{2} \right) \le \sqrt{\frac{2}{n}}, \quad \forall (j,i). \tag{18}$$

Define $\sigma_z^2 = \sum_{s=0}^{T-1} p_s \, \sigma_s^2$. As $n \to \infty$ the properties of typical sequences assure that $\boldsymbol{S}$ contains close to $n \, p_s$ elements $S = s$, $\forall s \in \{0, 1, \ldots, T-1\}$. Using (18) in (17) allows to bound $\sigma_{\boldsymbol{s}}^2$, as $n \to \infty$, with

$$\sigma_{\boldsymbol{s}}^2 \le \frac{2}{n} \sum_{i=1}^{n} \sigma_{z_i|\boldsymbol{s}}^2 \to \frac{2}{n} \sum_{s=0}^{T-1} n \, p_s \, \sigma_s^2 = 2\sigma_z^2. \tag{19}$$

The ideal Slepian-Wolf chain in Layer 1 is designed to guarantee vanishing probability of error in the worst-case scenario, that is when the variance of the correlation noise between $\Phi \boldsymbol{X}$ and $\Phi \boldsymbol{Y}$ equals $2\sigma_z^2$. Thus (11) becomes

$$R_1 = \frac{m}{2n} \log_2 \left(\frac{\sigma_x^2 \, 2\sigma_z^2}{(\sigma_x^2 + 2\sigma_z^2) \, \sigma^2(\Lambda_1)} \right). \tag{20}$$

The rate loss induced on Layer 1 transmission chain is thus due to the worst-case policy in protection against Slepian-Wolf decoding errors, and to the choice of $\Phi \boldsymbol{Y}$ as Slepian-Wolf side information sequence. The penalty, however, is limited thanks to the effect of the linear transformation Φ, which induces similar probability density functions (16) for each ζ_j, with components characterized by comparable variances.

3.2 Decoding

The decoder operates in two steps. The Layer 1 Slepian-Wolf index is decoded to recover the quantized sequence $\Phi \boldsymbol{X} + \boldsymbol{N}_1$, where $\boldsymbol{N}_1$ is the quantization noise composed, under the assumption $n \to \infty$, by independent Gaussian symbols with variance $\sigma^2(\Lambda_1)$. The Layer 1 decoder outputs the estimate $\widetilde{\boldsymbol{X}}$ of the source, to be employed in the second step to decode the Layer 2 Slepian-Wolf index. The estimate $\widehat{\boldsymbol{X}}$ output of the overall system is derived from the joint observation of $\Phi \boldsymbol{X} + \boldsymbol{N}_1$, $\overline{\Phi} \boldsymbol{X} + \boldsymbol{N}_2$ and $\boldsymbol{Y}$.

The Layer 1 estimator operates on the observed sequences $\Phi \boldsymbol{X} + \boldsymbol{N}_1$ and $\boldsymbol{Y}$. The system is described as

$$\boldsymbol{\Omega} = \begin{pmatrix} \Phi \\ \mathrm{I}_n \end{pmatrix} \boldsymbol{X} + \begin{pmatrix} \boldsymbol{N}_1 \\ \boldsymbol{Z} \end{pmatrix} = \mathrm{A} \boldsymbol{X} + \boldsymbol{W}. \tag{21}$$

The MMSE estimator of $\boldsymbol{X}$ from the observation $\boldsymbol{\Omega} = \boldsymbol{\omega}$ is

$$\widetilde{\boldsymbol{x}}(\boldsymbol{\omega}) = E\big[\boldsymbol{X} | \boldsymbol{\omega}\big] = \int_{\mathbb{R}^n} \boldsymbol{x} \, f_{\boldsymbol{X}|\boldsymbol{\omega}}(\boldsymbol{x}) \, \mathrm{d}\boldsymbol{x}, \tag{22}$$

where the conditional density $f_{\boldsymbol{X}|\boldsymbol{\omega}}$ can be developed as

$$f_{\boldsymbol{X}|\boldsymbol{\omega}} = \sum_{\boldsymbol{s}} \Pr(\boldsymbol{S} = \boldsymbol{s} | \boldsymbol{\omega}) \, f_{\boldsymbol{X}|\boldsymbol{s},\boldsymbol{\omega}}. \tag{23}$$

Using (23) in (22) yields the expression of the estimate as

$$\widetilde{\boldsymbol{x}}(\boldsymbol{\omega}) = \sum_{\boldsymbol{s}} \Pr(\boldsymbol{S} = \boldsymbol{s} | \boldsymbol{\omega}) \, \widetilde{\boldsymbol{x}}_{\boldsymbol{s}}(\boldsymbol{\omega}, \boldsymbol{s}), \tag{24}$$

where $\widetilde{\boldsymbol{x}}_{\boldsymbol{s}}(\boldsymbol{\omega}, \boldsymbol{s})$ is, by definition, the MMSE estimate of the source sequence obtained from the observation vector $\boldsymbol{\Omega} = \boldsymbol{\omega}$, when the realization of the state sequence $\boldsymbol{S} = \boldsymbol{s}$ is available at the decoder. Since $(\boldsymbol{X}, \boldsymbol{\Omega}|\boldsymbol{S} = \boldsymbol{s})$ are jointly Gaussian by construction, the MMSE estimate $\widetilde{\boldsymbol{x}}_{\boldsymbol{s}}(\boldsymbol{\omega}, \boldsymbol{s})$ has linear form

$$\widetilde{\boldsymbol{x}}_{\boldsymbol{s}}(\boldsymbol{\omega}, \boldsymbol{s}) = \left(\mathrm{K}_{\boldsymbol{X}}^{-1} + \mathrm{A}^T \mathrm{K}_{\boldsymbol{W}|\boldsymbol{s}}^{-1} \mathrm{A}\right)^{-1} \mathrm{A}^T \mathrm{K}_{\boldsymbol{W}|\boldsymbol{s}}^{-1}\, \boldsymbol{\omega}, \tag{25}$$

where $\mathrm{K}_{\boldsymbol{W}|\boldsymbol{s}}$ is the covariance matrix of the observation noise when $\boldsymbol{s}$ is available. The estimate $\widetilde{\boldsymbol{x}}(\boldsymbol{\omega})$, although expressed as a combination of the linear estimates $\widetilde{\boldsymbol{x}}_{\boldsymbol{s}}(\boldsymbol{\omega}, \boldsymbol{s})$, is not linear on $\boldsymbol{\omega}$, since the weighting coefficients in (24) have expression

$$\Pr\left(\boldsymbol{S} = \boldsymbol{s}|\boldsymbol{\omega}\right) = \frac{f_{\boldsymbol{\Omega}|\boldsymbol{s}}(\boldsymbol{\omega}|\boldsymbol{s}) \Pr(\boldsymbol{S} = \boldsymbol{s})}{\sum_{\boldsymbol{s}} f_{\boldsymbol{\Omega}|\boldsymbol{s}}(\boldsymbol{\omega}|\boldsymbol{s}) \Pr(\boldsymbol{S} = \boldsymbol{s})}. \tag{26}$$

The MSE vector $\boldsymbol{M}$ at the output of Layer 1 estimator is defined by

$$\begin{aligned}\boldsymbol{M} &= \mathrm{diag}\left(\mathrm{Cov}\left[\boldsymbol{X} - \widetilde{\boldsymbol{x}}(\boldsymbol{\Omega})\right]\right) \\ &= \mathrm{diag}\left(\mathrm{K}_{\boldsymbol{X}} - E_{\boldsymbol{\Omega}}\left[\widetilde{\boldsymbol{x}}(\boldsymbol{\Omega})\, \widetilde{\boldsymbol{x}}(\boldsymbol{\Omega})^T\right]\right), \end{aligned}\tag{27}$$

whose analytical expression cannot be evaluated. The performance of the Layer 1 estimator is upper bounded by the performance of the suboptimal LMMSE estimator, whose expression is

$$\begin{aligned}\boldsymbol{M}_{\mathrm{L}} &= \mathrm{diag}\Big(\mathrm{K}_{\boldsymbol{X}} - \mathrm{K}_{\boldsymbol{X}\boldsymbol{\Omega}}\, \mathrm{K}_{\boldsymbol{\Omega}}^{-1}\, \mathrm{K}_{\boldsymbol{\Omega}\boldsymbol{X}}\Big) \\ &= \mathrm{diag}\left(\frac{\sigma_x^2\, \sigma_z^2}{\sigma_x^2 + \sigma_z^2}\left(\mathrm{I}_n - \frac{\sigma_x^2\, \sigma_z^2}{\Delta_1}\Phi^T\Phi\right)\right), \end{aligned}\tag{28}$$

with $\Delta_1 = \sigma_x^2\big(\sigma_z^2 + \sigma^2(\Lambda_1)\big) + \sigma_z^2\sigma^2\left(\Lambda_1\right)$. Consider, as before, the linear transform matrix L to be a DCT matrix, and assume n odd; the rows $(n-1)/2 - j$ and $(n+1)/2 + 1 + j$ of L are symmetric. Proper choice of the pairs of rows for the construction of Φ guarantees that $\Phi^T\Phi$ has equal diagonal elements, of value m/n. The expression of the per sample distortion D' on the output of the Layer 1 decoder is then, from (28),

$$D' = \frac{\sigma_x^2\, \sigma_z^2}{\sigma_x^2 + \sigma_z^2}\left(1 - \frac{m}{n}\frac{\sigma_x^2\, \sigma_z^2}{\Delta_1}\right), \tag{29}$$

and depends on the length n only through the ratio m/n.

The Layer 2 estimator operates on the observed sequences $\Phi\boldsymbol{X} + \boldsymbol{N}_1$, $\overline{\Phi}\boldsymbol{X} + \boldsymbol{N}_2$ and $\boldsymbol{Y}$. The system is described as

$$\boldsymbol{\Omega}' = \begin{pmatrix}\Phi \\ \overline{\Phi} \\ \mathrm{I}_n\end{pmatrix}\boldsymbol{X} + \begin{pmatrix}\boldsymbol{N}_1 \\ \boldsymbol{N}_2 \\ \boldsymbol{Z}\end{pmatrix} = \mathrm{A}'\boldsymbol{X} + \boldsymbol{W}'. \tag{30}$$

As for the Layer 1 estimator the optimal reconstruction is the weighted sum of the MMSE estimates $\widehat{\boldsymbol{x}}_{\boldsymbol{s}}(\boldsymbol{\omega}', \boldsymbol{s})$, obtained from the observation $\boldsymbol{\Omega}' = \boldsymbol{\omega}'$ when the realization $\boldsymbol{S} = \boldsymbol{s}$ is available at the decoder side

$$\widehat{\boldsymbol{x}}(\boldsymbol{\omega}') = \sum_{\boldsymbol{s}} \Pr(\boldsymbol{S} = \boldsymbol{s}|\boldsymbol{\omega}')\, \widehat{\boldsymbol{x}}_{\boldsymbol{s}}(\boldsymbol{\omega}', \boldsymbol{s}), \tag{31}$$

with $\widehat{\boldsymbol{x}}_{\boldsymbol{s}}(\boldsymbol{\omega}', \boldsymbol{s})$ linear function of $\boldsymbol{\omega}'$, since $(\boldsymbol{X}, \boldsymbol{\Omega}'|\boldsymbol{S} = \boldsymbol{s})$ are jointly Gaussian. The MSE vector $\boldsymbol{M}'$ is obtained as

$$\boldsymbol{M}' = \mathrm{diag}\left(\mathrm{Cov}\left[\boldsymbol{X} - \widehat{\boldsymbol{x}}(\boldsymbol{\Omega}')\right]\right) \tag{32}$$

and can be upper bounded by the performance of the suboptimal LMMSE estimator

$$\begin{aligned}\boldsymbol{M}_{\mathrm{L}}' &= \mathrm{diag}\Big(\mathrm{K}_{\boldsymbol{X}} - \mathrm{K}_{\boldsymbol{X}\boldsymbol{\Omega}'}\, \mathrm{K}_{\boldsymbol{\Omega}'}^{-1}\, \mathrm{K}_{\boldsymbol{\Omega}'\boldsymbol{X}}\Big) \\ &= \mathrm{diag}\left(\frac{\sigma_x^2\, \sigma_z^2}{\sigma_x^2 + \sigma_z^2}\left(\mathrm{I}_n - \frac{\sigma_x^2\, \sigma_z^2}{\Delta_1}\Phi^T\Phi - \frac{\sigma_x^2\, \sigma_z^2}{\Delta_2}\overline{\Phi}^T\overline{\Phi}\right)\right), \end{aligned}\tag{33}$$

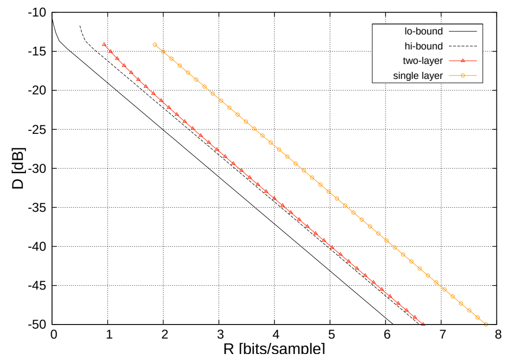

Figure 2: Performance of the coding scheme - GBG correlation model, $\sigma_x^2 = 1$, $p = 0.1$, $\sigma_0^2 = 0.04$, $\sigma_1^2 = 1$.

where $\Delta_2 = \sigma_x^2\big(\sigma_z^2 + \sigma^2(\Lambda_2)\big) + \sigma_z^2\sigma^2\left(\Lambda_2\right)$. The per sample distortion D at the output of the system is evaluated as

$$D = \frac{\sigma_x^2\, \sigma_z^2}{\sigma_x^2 + \sigma_z^2}\left(1 - \frac{m}{n}\frac{\sigma_x^2\, \sigma_z^2}{\Delta_1} - \frac{(n-m)}{n}\frac{\sigma_x^2\, \sigma_z^2}{\Delta_2}\right), \tag{34}$$

and depends only on the ratio m/n.

Remark that (29) and (34) express only upper bounds to the minimum distortions D' and D achievable by the scheme. They allow, however, to describe the operational rate-distortion performance $(R_1 + R_2, D)$ as a function of the design parameters m/n, $\sigma^2(\Lambda_1)$ and $\sigma^2(\Lambda_2)$, thus enabling design parameter optimization. The key feature of the two-layer architecture is the shift of (nearly) all suboptimality arising from the uncertainty of the side information quality to Layer 1. Transmission on the Layer 1 is expensive, so that it is convenient to restrain the ratio m/n. On the other hand the quality of the Layer 1 estimate, which provides consistent rate savings on Layer 2, increases with m/n as well: a trade off between the two conflicting requirements exists, which determines the optimum performance. Figure 2 shows, in red, the asymptotic (for $n \to \infty$) achievable performance of the coding scheme, evaluated for the GBG correlation model of parameters $\sigma_x^2 = 1$, $p = 0.1$, $\sigma_0^2 = 0.04$, $\sigma_1^2 = 1$. The optimum value of the ratio m/n is, in high rate regime, constant, and equals $m/n = 0.216$. The achievable performance is compared with the lower (solid line) and upper (dashed line) bounds to $R_{X|Y}^{\mathrm{WZ}}(D)$, and with the performance of the standard Wyner-Ziv coding scheme, dimensioned for the worst correlation noise (yellow line).

4. TOWARDS PRACTICAL IMPLEMENTATION

The two-layer coding architecture here proposed can be understood as a theoretical tool for the design of Wyner-Ziv coding schemes for uncertain side information quality. Remark that the fundamental components of the coding scheme are represented by quantizers, and Slepian-Wolf coding chains for *constant side information quality*. Highly efficient Slepian-Wolf coding implementations are by now available, as the result of the extensive research effort of the last decade (a state-of-the-art survey can be found in [4]). The purpose of this section is to focus on the implementation issues related to Layer 1 and Layer 2 estimators only.

The optimum Layer 1 (24) and Layer 2 (31) estimators are defined, as detailed in the previous section, as weighted sums of the marginal (with respect to the realization of $\boldsymbol{S}$) estimates. This clearly represents an obstacle to implementation, since the number of marginal estimates $\widetilde{\boldsymbol{x}}_{\boldsymbol{s}}$ and $\widehat{\boldsymbol{x}}_{\boldsymbol{s}}$ to be evaluated grows exponentially with the size n of the source block. The weighting coefficients $\Pr(\boldsymbol{S} = \boldsymbol{s}|\boldsymbol{\omega})$ and $\Pr(\boldsymbol{S} = \boldsymbol{s}|\boldsymbol{\omega}')$ associated with each marginal estimate pair $\widetilde{\boldsymbol{x}}_{\boldsymbol{s}}$ and $\widehat{\boldsymbol{x}}_{\boldsymbol{s}}$ represent the likelihoods of the state sequence realization $\boldsymbol{s}$, given the observation vectors $\boldsymbol{\omega}$ and $\boldsymbol{\omega}'$: among them, only a small fraction have a significant value, and concentrates (almost) the totality of the *a posteriori* probability. A

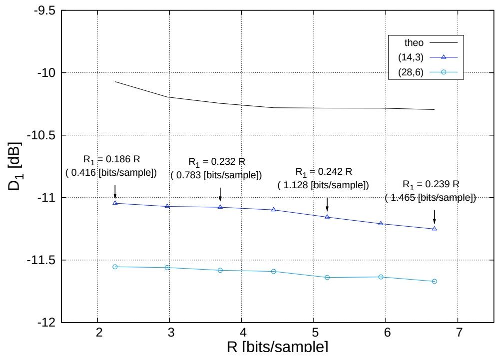

Figure 3: Layer 1 estimator, BMP implementation - GBG correlation model, $\sigma_x^2 = 1, p = 0.1, \sigma_0^2 = 0.04, \sigma_1^2 = 1$.

reliable approximation of the estimates $\widetilde{\boldsymbol{x}}$ (24) and $\widehat{\boldsymbol{x}}$ (31) can thus be obtained by identifying a set $\mathcal{W}^*$ of the sequences $\boldsymbol{S}|\boldsymbol{\omega}$ such as

$$\sum_{\boldsymbol{s} \in \mathcal{W}^*} \Pr(\boldsymbol{S} = \boldsymbol{s}|\boldsymbol{\omega}) \approx 1 \tag{35}$$

and performing the sums in (24) and (31) over the subspace $\mathcal{W}^*$, instead that over all possible $\boldsymbol{S}$.

The solution adopted to estimate the set $\mathcal{W}^*$ is derived adapting the Bayesian Matching Pursuit algorithm [20], originally proposed to solve the problem of compressed sensing of sparse sources. The algorithm performs a tree search in order to identify the elements $\boldsymbol{s}^* \in \mathcal{W}^*$; on each exploration level the surviving leaves (the state sequences $\boldsymbol{s}^*$) are selected as the ones maximizing $\Pr(\boldsymbol{S} = \boldsymbol{s}|\boldsymbol{\omega})$. The root of the tree is represented by the sequence $\boldsymbol{s}$ composed by all zero elements. The descent to the following level of the tree is performed activating one impulse position at a time. The $M(d)$ elements surviving at depth d and are employed to generate the following level.

The algorithm has been implemented and tested for the GBG correlation model, for the same mixture parameters considered before. Figure 3 presents the simulation results obtained on Layer 1 estimator, where the choice of the design parameters m/n, $\sigma^2(\Lambda_1)$ and $\sigma^2(\Lambda_2)$ results from the performance optimization procedure for $n \to \infty$. The black curve in Figure 3 represents the value of the distortion D' at the output of Layer 1 estimator obtained with the optimization procedure. The light-blue and blue curves depict the distortions obtained using the Bayesian Matching Pursuit algorithm, for block lengths $(n = 14, m = 3)$ and $(n = 28, m = 6)$ respectively: they consistently improve the theoretical distortion value D'. This should not surprise: the analytical expression (29), in fact, represent the attainable, but non minimum, distortion D'. The simulation results enable the conclusion that the optimization of the design parameters obtained for $n \to \infty$ is to be considered effective for finite-length design as well.

5. CONCLUSIONS

This paper outlines a theoretical framework for Wyner-Ziv coding with uncertain side information quality. The proposed coding scheme uses the two-layer approach presented in [12] for the case of possibly erased side information. The blindness of the decoder to the actual state over the correlation channel, however, makes impossible to adopt the coding strategy devised in [12]. The Bayesian approach proposed here allows to capture the effect of the most likely state realizations at a reasonable complexity, thus solving the decoding problem. The attainable performance of the scheme can be analytically characterized and compared to the theoretical bounds, showing that the coding solution is effective.

Acknowledgments

The authors would like to thank Phil Schniter for the fruitful discussions on the Fast Bayesian Matching Pursuit algorithm and its application to Wyner-Ziv coding with uncertain side information quality.

References

[1] Z. Xiong, A. Liveris, and S. Cheng, “Distributed source coding for sensor networks,” *IEEE Signal Processing Magazine*, vol. 21, no. 5, pp. 80– 94, Sept. 2004.

[2] B. Girod, A. Aaron, S. Rane, and D. Rebollo-Monedero, “Distributed video coding,” *Proceedings of the IEEE*, vol. 93, no. 1, pp. 71–83, Jan. 2005.

[3] S. Pradhan and K. Ramchandran, “Distributed source coding using syndromes (DISCUS): design and construction,” in *Proc. DCC*, Snowbird, Utah, March 1999, pp. 158–167.

[4] P. Dragotti and M. Gastpar, *Distributed Source Coding, Theory, Algorithms and Applications*. Elsevier, 2009.

[5] R. Westerlaken and R. Gunnewiek, R.K.and Lagendijk, “The role of the virtual channel in distributed source coding of video,” in *Proc. IEEE ICIP*, Sept. 2005, pp. 581–584.

[6] A. Aaron, S. Rane, and B. Girod, “Wyner-Ziv video coding with hash-based motion compensation at the receiver,” in *Proc. IEEE ICIP*, Oct. 2004, pp. 3097–3100.

[7] R. Puri, A. Majumdar, and K. Ramchandran, “PRISM: a video coding paradigm with motion estimation at the decoder,” *IEEE Transactions on Image Processing*, vol. 16, no. 10, pp. 2436–2448, Oct. 2007.

[8] C. Heegard and T. Berger, “Rate distortion when side information may be absent,” *IEEE Transactions on Information Theory*, vol. 31, pp. 727 – 734, Nov. 1985.

[9] A. Kaspi, “Rate-distortion function when side-information may be present at the decoder,” *IEEE Transactions on Information Theory*, vol. 40, no. 6, pp. 2031 – 2034, Nov. 1994.

[10] S. Verdú and T. Weissman, “The information lost in erasures,” *IEEE Trans. on Information Theory*, vol. 54, no. 11, pp. 5030–5058, Nov. 2008.

[11] E. Perron, S. Diggavi, and E. Telatar, “Lossy source coding with Gaussian or erased side-information,” in *Proc. IEEE ISIT*, 2009, pp. 1035–1039.

[12] F. Bassi, M. Kieffer, and C. Weidmann, “Source coding with intermittent and degraded side information at the decoder,” in *Proc. IEEE ICASSP*, April 2008, pp. 2941 – 2944.

[13] C. Weidmann, F. Bassi, and M. Kieffer, “Practical distributed source coding with impulse-noise degraded side information at the decoder,” in *Proc. EUSIPCO*, Aug. 2008.

[14] A. Wyner and J. Ziv, “The rate-distortion function for source coding with side information at the decoder,” *IEEE Transactions on Information Theory*, vol. 22, no. 1, pp. 1–10, Jan. 1976.

[15] M. Escobar and M. West, “Bayesian density estimation and inference using mixtures,” *Journal of the American Statistical Association*, vol. 90, no. 430, pp. 577–588, June 1995.

[16] D. Rebollo-Monedero, S. Rane, A. Aaron, and B. Girod, “High-rate quantization and transform coding with side information at the decoder,” *EURASIP Journal on Signal Processing*, vol. 86, no. 11, pp. 3160–3179, 2006.

[17] R. Zamir and M. Feder, “On universal quantization by randomized uniform/lattice quantizers,” *IEEE Transactions on Information Theory*, vol. 38, no. 2, pp. 428–436, March 1992.

[18] T. Cover and J. Thomas, *Elements of Information Theory*. Wiley, 2006.

[19] F. Bassi, M. Kieffer, and C. Weidmann, “Wyner-Ziv coding with uncertain side information quality,” LSS - CNRS - SUPELEC - Univ Paris-Sud, Tech. Rep., 2010. [Online]. Available: http://www.lss.supelec.fr/perso/francesca.bassi/pub/techrep.pdf

[20] P. Schniter and J. Potter, L.C. Ziniel, “Fast Bayesian matching pursuit: model uncertainty and parameter estimation for sparse linear models,” *IEEE Transactions on Signal Processing*, March 2009.