16th European Signal Processing Conference (EUSIPCO 2008), Lausanne, Switzerland, August 25-29, 2008, copyright by EURASIP

STATISTICAL MODEL-AIDED DECODING OF CONTINUOUS-VALUED SYNDROMES FOR SOURCE CODING WITH SIDE INFORMATION

Lorenzo Cappellari

Department of Information Engineering, University of Padova
via Gradenigo, 6/B - 35131 Padova - Italy
email: lorenzo.cappellari@dei.unipd.it

ABSTRACT

The concept of syndrome plays undoubtedly a central role in distributed source coding. With known source-side correlation, systems based on continuous-valued syndromes have indeed been shown to perform close to the Wyner-Ziv bound, both in theory and in practice. This paper investigates the application of the continuous-valued syndrome-based approach to the real case, where little or no knowledge regarding the source-side correlation is available at the encoder. Since in this case the encoder cannot operate at his best, traditional maximum likelihood decoding do not perform well. Iterative, factor graph based, statistical model-aided decoding is instead able to provide more accurate results. The experiments show in particular that model-aided decoding leads to about one order of magnitude less reconstruction errors within a few decoder iterations, which amounts to an increase of the signal-to-noise ratio of up to 3 dB.

1. INTRODUCTION

In the problem of *coding with side information at the decoder*, a source must be encoded and decoded within a given distortion under the condition that another source, correlated with the former, is only available at the decoder. Despite the theoretical performance bounds of this scenario were found in the seventies [1], practical applications have appeared only recently. Most of them rely on the *syndrome* concept borrowed from channel coding [2] or on some similar concepts, like the one proposed in [3] to embody quality scalability features as well.

To maximize the coding performance, the correlation between the two sources must be known during encoding. But, in real scenarios, the correlation is neither known nor time-invariant and must be estimated. In distributed video coding applications [4] where a frame and a motion-compensated estimate are considered as the two correlated sources, for example, such correlation is either coarsely approximated at the encoder [5] or learned using feedback from the decoder [4].

While feedback messages permit to obtain reasonable correlation estimates, coarse estimates at the encoder may be wrong and lead to decoding errors. However, without any feedback, the decoder can still try to minimize the errors by assuming a statistical model ([6, 7]) for the source [8].

This work proposes an iterative, factor graph based [9] solution to the problem of optimal syndrome decoding for the scenarios where the correlation structure can be modeled by a hidden Markov model [6] and the encoder relies only on coarse correlation estimates. Section 2 reviews the distributed source coding algorithm based on continuous-valued syndromes [3]. Section 3 illustrates the proposed decoding algorithm. The results of the simulations are presented in Section 4. Section 5 outlines the main results and discusses some future lines of research.

2. CONTINUOUS-VALUED SYNDROME-BASED DISTRIBUTED SOURCE CODING

2.1 Definition and Properties

Consider a real m-dimensional *lattice* $\Lambda \subset \mathbb{R}^m$. Since Λ is a subgroup of the additive group $\mathbb{R}^m$, it induces the partition of $\mathbb{R}^m$ into the cosets that belong to the quotient group $\mathbb{R}^m/\Lambda$. Choose a *labeling function* $l : \mathbb{R}^m/\Lambda \to \mathbb{R}^m$ such that $l(A) \in A$, $\forall A \in \mathbb{R}^m/\Lambda$, and call *fundamental region induced by* l the image of l, i.e. $R_l(\Lambda) \triangleq l(\mathbb{R}^m/\Lambda) \subset \mathbb{R}^m$. Note that the set of translates $\{R_l(\Lambda)+b : b \in \Lambda\}$ forms a regular tessellation of $\mathbb{R}^m$, and hence the volume of $R_l(\Lambda)$ equals $V(\Lambda)$, the volume of n-space per point of Λ.

If the sum operation induced by l on $R_l(\Lambda)$ is considered, l is an isomorphism. Denote with ν the natural homomorphism $\nu : \mathbb{R}^m \to \mathbb{R}^m/\Lambda$, and define the function $s_l \triangleq l \circ \nu : \mathbb{R}^m \to R_l(\Lambda)$. Since s_l identifies indeed the coset to which any element of $\mathbb{R}^m$ belongs, $s_l(a)$ represents the *continuous-valued syndrome* of $a \in \mathbb{R}^m$, by analogy with the role of the traditional syndrome in linear codes.

It can be easily shown that the continuous-valued syndrome satisfies the following properties [10]:

$$s_l(a+\lambda) = s_l(a),\ \forall a \in \mathbb{R}^m, \lambda \in \Lambda\,; \quad (1)$$

$$s_l(a) = a,\ \forall a \in R_l(\Lambda)\,; \quad (2)$$

$$s_l(a+b) = s_l(a)+s_l(b),\ \forall a,b \in \mathbb{R}^m\,, \quad (3)$$

with the sum on the right-hand side of (3) being the sum operation induced by l. As a remark, $\tilde{s}_l : \mathbb{R}^m \to (\mathbb{R}^m,+)$ is not a homomorphism[1] since $R_l(\Lambda)$ is not a subgroup of $(\mathbb{R}^m,+)$. However, it is straightforward to show that

$$\tilde{s}_l(a+b) = \tilde{s}_l\left(\tilde{s}_l(a)+\tilde{s}_l(b)\right),\ \forall a,b \in \mathbb{R}^m\,. \quad (4)$$

Since $\mathbb{R}^m$ is a normed space under the L^p-norm ($p \geq 1$), it is common to take elements with minimum norm as coset representatives. A labeling function V_p such that

$$\|V_p(A)\|_p \leq \|a\|_p,\ \forall A \in \mathbb{R}^m/\Lambda, a \in A$$

defines then a *convex* fundamental region $R_{V_p}(\Lambda)$ known as *fundamental Voronoi region*. As a remark, note that when $p=2$ the corresponding tessellation of $\mathbb{R}^m$ consists of decision regions for a minimum-distance quantizer (or decoder)

[1] In this section, for the sake of clarity, the continuous-valued syndrome $s_l(\cdot)$ is denoted as $\tilde{s}_l(\cdot)$ when intended as belonging to $(\mathbb{R}^m,+)$ rather than $(R_l(\Lambda),+)$.

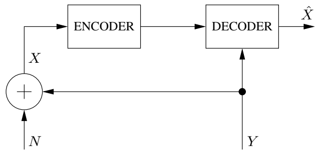


Figure 1: Coding with side information at the decoder, with a backward additive correlation channel (N is independent of Y).

that uses Λ as codebook. In order to evaluate the continuous-valued syndrome $s_{V_p}(a) \in R_{V_p}(\Lambda)$, it is immediate to verify that it can be obtained as *quantization error* of a minimum-norm quantizer that uses Λ as codebook. In particular, defining

$$Q_\Lambda^{(p)}(a) \triangleq \lambda \in \Lambda : \|\lambda - a\|_p \leq \|\gamma - a\|_p, \, \forall \gamma \in \Lambda$$

as the closest lattice point to a (with the further condition $(a-\lambda) \in R_{V_p}(\Lambda)$ in case of ambiguity), we have

$$\tilde{s}_{V_p}(a) = a - Q_\Lambda^{(p)}(a) \,. \tag{5}$$

As the traditional syndrome, this particular continuous-valued syndrome of a identifies hence the minimum-norm element that, subtracted to a, leads to an element (the closest) of Λ.

2.2 Coding with Side Information (CSI) at the Decoder

Consider now two continuous random sources X and Y whose correlation is described by means of a *backward* additive channel, i.e. such that $X = Y + N$, where N is a continuous random source independent of Y, *memoryless*, and *stationary*. Assume that we want to efficiently encode any m-dimensional realization x of X allowing only the decoder to access the corresponding realization y of Y, as in Fig. 1.

If each m-dimensional realization n of N belongs to $R_{V_p}(\Lambda)$, then the continuous-valued syndrome $s_{V_p}(x)$ is sufficient for perfect reconstruction at the decoder. In fact,

$$\begin{aligned} x &= y + n = y + \tilde{s}_{V_p}(n) & (6)\\ &= y + \tilde{s}_{V_p}(x-y) = y + \tilde{s}_{V_p}\left(\tilde{s}_{V_p}(x) + \tilde{s}_{V_p}(-y)\right) \,, & (7) \end{aligned}$$

where (6) follows from (2), and (7) follows from the property of the sum (4). From an operative point of view it is useful to note that

$$\begin{aligned} x &= y + \tilde{s}_{V_p}\left(\tilde{s}_{V_p}(x) - y - Q_\Lambda^{(p)}(-y)\right) & (8)\\ &= y + \tilde{s}_{V_p}\left(\tilde{s}_{V_p}(x) - y\right) & (9)\\ &= \tilde{s}_{V_p}(x) - Q_\Lambda^{(p)}\left(\tilde{s}_{V_p}(x) - y\right) \,, & (10) \end{aligned}$$

where (8) follows from the operative definition of the continuous-valued syndrome (5), (9) follows from the periodicity (1) of $s_{V_p}(\cdot)$, and (10), again, from the definition (5). Then, a single quantization operation is sufficient at the decoder for perfect reconstruction.

Since $s_{V_p}(X)$ is a continuous random variable, a real channel can actually convey only a quantized (*noisy*) version of the continuous-valued syndrome to the decoder. Then the reconstruction satisfies

$$\begin{aligned} \hat{x} &= y + \tilde{s}_{V_p}\left(\tilde{s}_{V_p}(x) + q + \tilde{s}_{V_p}(-y)\right) & (11)\\ &= y + \tilde{s}_{V_p}\left(\tilde{s}_{V_p}(x) + \tilde{s}_{V_p}(q) + \tilde{s}_{V_p}(-y)\right) & (12)\\ &= y + \tilde{s}_{V_p}(x+q-y) = y + \tilde{s}_{V_p}(n+q) & (13)\\ &= x + \left(q - Q_\Lambda^{(p)}(n+q)\right) = x + (q + q_{ol}) \,, & (14) \end{aligned}$$

where: (11) is obtained from (7) by substitution of $\tilde{s}_{V_p}(x)$ with $\tilde{s}_{V_p}(x) + q$ (q is an m-dimensional realization of a random variable Q), (12) and (13) follow from (4), and (14) follows, again, from the definition (5).

The total reconstruction error $q_t \triangleq q + q_{ol}$ is hence the sum of the *granular* error q, and of the *overload* error $q_{ol} \triangleq -Q_\Lambda^{(p)}(n+q)$. If q is negligible w.r.t. n, and it is reasonable that this happens at high transmission rates, then the probability of q_{ol} being not zero (that is called *error probability*[2] P_e) is negligible and $q_t \simeq q$. Hence, the same additive *error Q* would impair both the syndrome and the source at the decoder (as shown in a slightly different way in [3]).

3. STATISTICAL MODEL-AIDED DECODING

The traditional decoding algorithm (10), that simply amounts to *maximum-likelihood* (ML) decoding, gives perfect reconstruction if Λ is such that $n \in R_{V_p}(\Lambda)$, and is asymptotically optimal in the rate-distortion sense if $V(\Lambda) \simeq V(A_\varepsilon^{(m)})$ [10], where $V(A_\varepsilon^{(m)})$ is the volume of the *typical set* [11] of N. Hence, for optimal performance, the encoder (and the decoder) must know the statistics of N and choose accordingly a good lattice Λ.

In the case where N is stationary except for the mean and the variance, but they are known at both the encoder and the decoder, it is still possible to achieve optimality by coding an opportune affine-transformed version of X. While such a model may be a good interpretative model for real correlation channels, only partial knowledge about the source-side correlation is usually available at the encoder. If the syndrome formation algorithm (5) is tailored to this wrong correlation information, then ML decoding is suboptimal.

In order to ease the investigation for an optimal decoding algorithm, the actual encoding procedure shown in Fig. 2(a) is approximated as in Fig. 2(b) (which in fact describes the same operation if $s + q \in R_{V_p}(\Lambda)$) and Fig. 2(c) (where $\tilde{n} = n + q$). Considering that $Y \rightarrow \tilde{X} \rightarrow \tilde{S}$ is a Markov chain [11], optimal *maximum a posteriori probability* (MAP) decoding amounts to maximizing

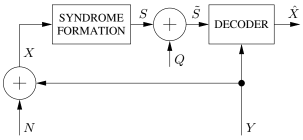
$$f(\tilde{x}|y,\tilde{s}) \propto f(\tilde{x}|y) f(\tilde{s}|\tilde{x}) \,, \tag{15}$$

where the terms $f(\tilde{x}|y)$ and $f(\tilde{s}|\tilde{x})$ take account of the correlation channel structure and of the syndrome formation algorithm respectively.

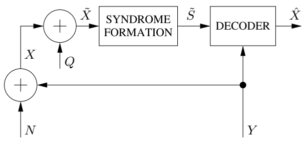
3.1 Statistical Modeling of the Correlation Channel

Due to the independence between Y and $\tilde{N}$, $f(\tilde{x}|y) = f_{\tilde{N}}(\tilde{x} - y|y) = f_{\tilde{N}}(\tilde{x} - y)$, where $f_{\tilde{N}}(\cdot)$ is the probability distribution function (*pdf*) of $\tilde{N}$. In order to model a possibly time-varying correlation channel whose actual statistics is not exactly known at any time instant, a doubly stochastic Hidden Markov Model (HMM) [6] is employed. In particular, it is assumed that the model has L states and that the distribution

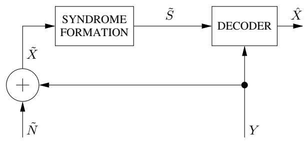
[2] An error occurs each time some coordinate of the m-dimensional vector $Q_\Lambda^{(p)}(n+q)$ is not 0.



(a) with the actual encoder



(b) approximated with an equivalent encoder



(c) approximated with an equivalent correlation channel

Figure 2: Syndrome formation and transmission.

corresponding to the j-th state is the generalized Gaussian distribution (GGD) $\mathscr{G}_\alpha(\mu_j, \sigma_j^2)$, $j = 0, 1, \ldots, L-1$.

The reasons for these choices are essentially two. Firstly, HMMs are very powerful tools for real signal modeling and lead to efficient signal processing algorithms (e.g. denoising [7]); hence, their utilization appears very suitable in the context of CSI as well. Secondly, GGDs with $\alpha < 2$ (e.g. Laplacian distributions) are more adequate for modeling the statistics of the difference between highly correlated data.

Once the hidden state variables $\sigma_i^{\tilde{N}}$, $i = 0, 1, \ldots, m-1$ are introduced, $f_{\tilde{N}}$ is found by marginalizing

$$\begin{aligned} f_{\tilde{N}}\left(n, \sigma^{\tilde{N}}\right) &= p\left(\sigma_0^{\tilde{N}}\right) f_{\tilde{N}}\left(n_0 \middle| \sigma_0^{\tilde{N}}\right) \cdot \\ &\cdot \prod_{i=1}^{m-1} p\left(\sigma_i^{\tilde{N}} \middle| \sigma_{i-1}^{\tilde{N}}\right) f_{\tilde{N}}\left(n_i \middle| \sigma_i^{\tilde{N}}\right) , \end{aligned}$$

where, with $\beta_j^2 = \sigma_j^2 \Gamma(1/\alpha)/2\Gamma(3/\alpha)$,

$$f_{\tilde{N}}(a|j) = \frac{\alpha/2}{\sqrt{2\Gamma(1/\alpha)^2 \beta_j^2}} e^{-\left(\frac{|a-\mu_j|}{\sqrt{2\beta_j^2}}\right)^\alpha} .$$

3.2 Modeling of Trellis-Based Syndrome Formation

Since the syndrome formation is a deterministic transformation, $f(\tilde{s}|\tilde{x})$ is a Dirac's delta function that, given $\tilde{x}$, reveals its syndrome. Under the condition $\tilde{s} \in R_{V_p}(\Lambda)$, $f(\tilde{s}|\tilde{x})$ is equivalently a delta function that reveals the event $\{\tilde{x} - \tilde{s} \in \Lambda\}$. If the quantizer used in (5) is an M-state trellis-coded quantizer [12] and the trellis state variables σ_k^C, $k = -1, 0, \ldots, m-1$ are introduced, $f(\tilde{s}|\tilde{x})$ is found marginalizing

$$f\left(\tilde{s}, \sigma^C | \tilde{x}\right) = \prod_{k=0}^{m-1} \chi_{\sigma_{k-1}^C}^{\sigma_k^C} \sum_{b \in \mathscr{B}_{\sigma_{k-1}^C}^{\sigma_k^C}} \delta\left(\tilde{x}_k - \tilde{s}_k - b\right)$$

where χ_j^l and $\mathscr{B}_j^l$ are the indicator function (unitary if j is connected to l, zero otherwise) and the set of reconstruction values relative to the transition from state j to l respectively.

3.3 The Factor Graph Approach to Decoding

Despite these factorizations, it seems still difficult to maximize (15). But once a *factor graph* [9] is available, it is

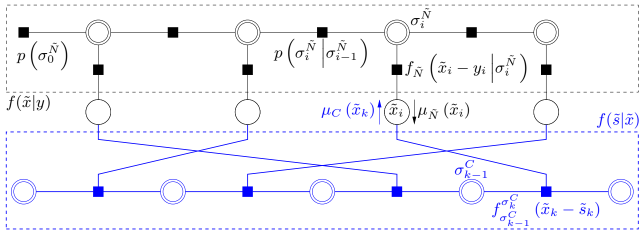


Figure 3: Example factor graph representing the right hand side of (15) with $m = 4$, and $f_{\sigma_{k-1}^C}^{\sigma_k^C}(x) \triangleq \chi_{\sigma_{k-1}^C}^{\sigma_k^C} \sum_{b \in \mathscr{B}_{\sigma_{k-1}^C}^{\sigma_k^C}} \delta(x - b)$.

straightforward to find the best MAP estimate (at each instant i) through standard message passing algorithms, i.e. to evaluate

$$\hat{x}_i = \arg\max_{\tilde{x}_i} f(\tilde{x}_i|y) f(\tilde{s}|\tilde{x}_i) .$$

In particular, to increase the independence between the information about $\tilde{x}$ brought by y and $\tilde{s}$, the syndrome formation algorithm is operated on a randomly scrambled version of x, such that the actual factor graph is similar to the one sketched in Fig. 3. Since the factor graph is not cycle-free, decoding is done according to the following iterative process.

1. Initialize the *model* messages $\mu_{\tilde{N}}^{(0)}(\tilde{x}_i)$, $i = 0, 1, \ldots, m-1$, with the *pdf*s of $\tilde{n}_i$ assumed during encoding, centered around y_i (i.e. take account of the partial knowledge on the source-side correlation); using these messages, run the standard *forward/backward algorithm* (FBA) on the factor graph representing $f(\tilde{s}|\tilde{x})$ to obtain the *code* messages $\mu_C^{(0)}(\tilde{x}_k)$, $k = 0, 1, \ldots, m-1$; set $j = 1$.
2. (*Model-based processing*) Using the messages $\mu_C^{(j-1)}(\tilde{x}_k)$, run the FBA on the factor graph representing $f(\tilde{x}|y)$ and obtain the messages $\mu_{\tilde{N}}^{(j)}(\tilde{x}_i)$.
3. (*Code-based processing*) Using the messages $\mu_{\tilde{N}}^{(j)}(\tilde{x}_i)$, run the FBA on the factor graph representing $f(\tilde{s}|\tilde{x})$ and obtain the messages $\mu_C^{(j)}(\tilde{x}_k)$.
4. Check if the required number of iterations has been done; if not, set $j \leftarrow j+1$ and go back to step 2.
5. Compute the marginals $f(\tilde{x}_i|y) f(\tilde{s}|\tilde{x}_i)$ as $\mu_{\tilde{N}}^{(j)}(\tilde{x}_i) \mu_C^{(j)}(\tilde{x}_k)$ and maximize them to obtain the MAP estimates.

Note that all *code* messages $\mu_C^{(j)}(\tilde{x}_k)$, $j = 0, 1, \ldots,$ consist of a sum of delta functions centered on the elements of the set $\mathscr{A}_k \triangleq \tilde{s}_k + \bigcup_{j,l=0}^{M-1} \mathscr{B}_j^l$, that is the $\tilde{s}_k$-translated extended reconstruction set [12] of the k-th trellis section. Similarly, in step 3, only the values of the *model* messages $\mu_{\tilde{N}}^{(j)}(\tilde{x}_i)$, $j = 0, 1, \ldots,$ taken on $\mathscr{A}_i$ contribute to the *code* messages. Hence, all messages need to be computed only on the discrete alphabets $\mathscr{A}_i$, $i = 0, 1, \ldots, m-1$.

4. SIMULATIONS AND EXPERIMENTAL RESULTS

To validate the proposed iterative decoding algorithm, the system shown in Fig. 4 has been simulated. In this system,

- y is a realization of a zero-mean, unitary variance, i.i.d. Gaussian process; n is a realization of an L-state HMM ($L = 3$ or 6) with distribution in the j-th state

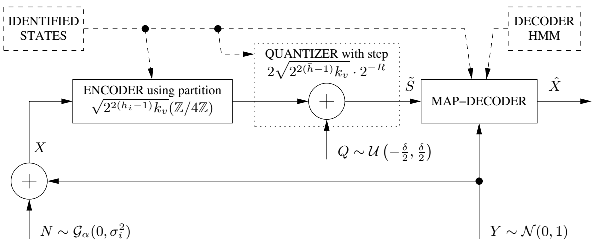


Figure 4: Block diagram of the simulated system for coding with side information.

$\mathcal{G}_\alpha(0,\sigma_j^2)$ ($\alpha = 2$ or 1), $j = 0,1,\ldots,L-1$, and variances chosen in order to model a time-varying *backward correlation signal-to-noise ratio* [10] (i.e. side-to-noise variance ratio) in the range from 5 to 20 dB (see Table 1);

- in order to generate a partial knowledge about the actual noise variances σ_i^2, $i = 0,1,\ldots,m-1$, n is identified within another HMM with possibly different number of states and state variances; this information is used by the encoder and transmitted to the decoder (the required bit-rate is taken into account);
- s is formed using an 8-state $\mathbb{Z}/4\mathbb{Z}$-based trellis-coded quantizer (with $p = 2$) that uses the [13 4] binary code [12]. Since the (two-dimension) normalized volume of the equivalent lattice would be 4, each side is multiplied by $\sqrt{2^{2(h_i-1)}k_v}$, where h_i is the differential entropy of n_i (according to the partial knowledge); then, the normalized volume of the equivalent lattice is k_v times the assumed normalized *typical set* [11] volume (this experimentally optimized factor takes account of both the imperfect typical set-Voronoi region matching and the syndrome corruption due to quantization [10]);
- s is quantized using a uniform, embedded scalar quantizer whose smallest step δ is such that the average bit-rate equals $R = 4$ bit/sample, i.e. $\delta = \sqrt{k_v}2^{\bar{h}-R}$ with $\bar{h} = \sum_{j=0}^{L-1} p(j)h_j$ the average differential entropy of n according to the model used for the identification of the noise; inverse quantization uses the middle point of each quantization interval, and considering only most significant bit-planes permits to address the target bit-rates $R = 1,2,3$ bit/sample as well;
- assuming that the HMM according to which the noise was generated is known, the decoder performs the algorithm described in Section 3. The information regarding the partial knowledge is used to determine the reconstruction alphabets $\mathscr{A}_i$; since they are not finite, only the 21 closest points to y_i are actually considered (decoding is impaired very slightly by this choice due to the fast decay of the state functions $f_{\tilde{N}}(\cdot|\cdot)$).

The results in the following are relative to the traditional ML decoding, the MAP decoding with $j = 0$ iterations, or the so called TURBO decoding (i.e. the MAP decoding with $j = 4$ iterations, that in all examined cases were proven to be sufficient for convergence). They are reported in terms of the error probability P_e and the mean variance of the total reconstruction error, averaged over 1000 sequences of $m = 1000$ samples each. The lower bound

$$D_L(R) = \frac{2^{2\bar{h}}}{2\pi e}2^{-2R}$$

name	L	$p(j)$	$p(l\|j)\cdot 10^3$	σ_j^2 (dB)
HMM(6)	6	0.372	982, 6, 5, 4, 2, 1	−20.0
		0.235	9, 974, 7, 5, 3, 2	−17.0
		0.175	10, 8, 970, 6, 4, 2	−14.0
		0.119	12, 10, 7, 964, 5, 2	−11.0
		0.070	15, 12, 9, 6, 955, 3	−8.0
		0.029	20, 16, 12, 8, 4, 941	−5.0
HMM(3)	3	0.607	986, 10, 4	−18.6
		0.294	20, 974, 7	−12.5
		0.099	29, 16, 954	−6.9
HMM(2)	2	0.782	991, 9	−17.1
		0.218	34, 966	−8.7
HMM(1)	1	1.000	1000	−13.5

Table 1: The considered *stationary* HMMs. $p(j)$ is the stationary distribution (that is used as initial probability), and $p(l|j)$ are the transition probabilities from state j to state l. The HMM(6) model was chosen in order to represent the frequent case where the source-side correlation is normally high (small noise) except for some occasional interval of time (high noise). The HMM(3), HMM(2), and HMM(1) models are obtained from HMM(6) by grouping (in order) the 6 states into 3, 2, and 1 macro-states respectively.

to the reconstruction error variance, with $\bar{h}$ the average differential entropy of n according to the model used for its generation, is reported as well. This bound represents the *Shannon lower bound* [11] for the problem of coding n (i.e. of coding x with y known at the encoder as well), that in the i.i.d. Gaussian case ($L = 1$, $\alpha = 2$) equals the theoretical performance limit of the simulated system, independently from the *pdf* of Y.

Figures 5(a) and 5(b) refer to the case of a Gaussian noise ($\alpha = 2$), generated according to HMM(3) (see Table 1). If the encoder has knowledge about this model (i.e. the same model is artificially used for hidden state identification[3]), the TURBO decoding algorithm does not perform any better than the MAP decoding algorithm, that was shown to perform like the ML decoding. If the encoder only knows the average variance of n (i.e. the HMM(1) model is used for hidden state identification) the MAP decoding performs again like the ML decoding, but TURBO decoding leads to about one order of magnitude less reconstruction errors (see Fig. 5(a)), in particular at bit-rates $R = 3 \div 4$ bit/sample. This translates into an up to 3 dB decrease of the mean error variance (see Fig. 5(b)). Clearly, the performance is still inferior w.r.t. the former case, except for low bit-rates, were a lower P_e is obtained.

Figures 6(a) and 6(b) refer to the case of a Gaussian noise generated according to HMM(6). In all experiments it turned out that ML and MAP decoding lead to the same performances. Again, if the encoder knows (exactly or partially) the generating model, TURBO decoding performs like MAP decoding, but if the encoder only knows the average variance of n, TURBO decoding permits to improve both the error probability and the mean error variance. At high bit-rates, the more the encoder knows about the model the better is the performance; at low bit-rates, however, the performances are about the same independently from the partial knowledge.

The results summarized in Table 2, that refer to the target bit-rate of 4 bit/sample, encompass the Laplacian case ($\alpha = 1$) as well. Differently from the Gaussian case, and w.r.t. ML decoding, MAP decoding is more performing, and TURBO

[3]Note that there may be still uncertainties about the *actual* hidden states.

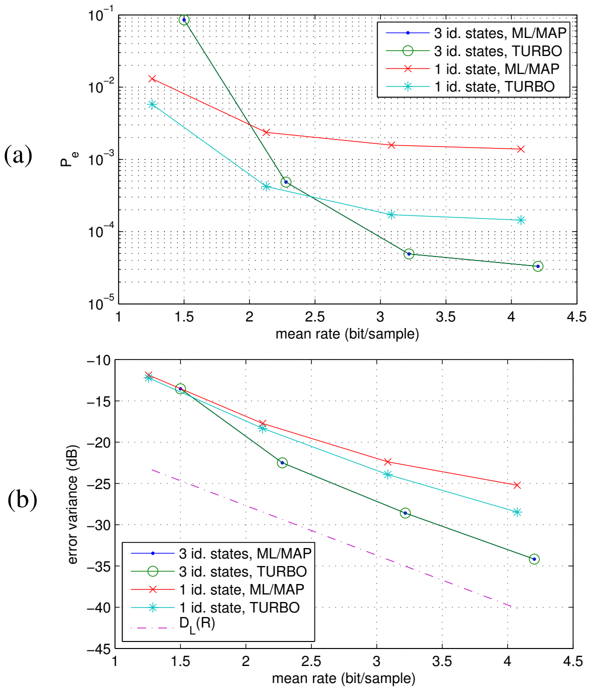


Figure 5: Results using the HMM(3) as generating model.

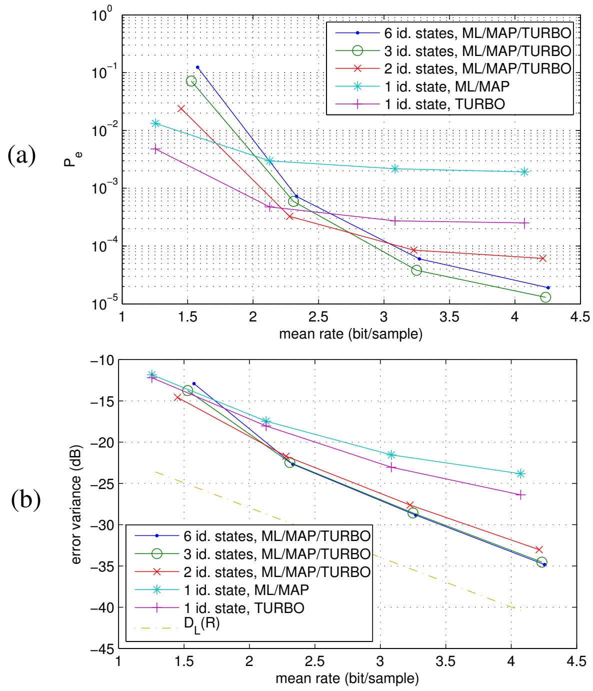


Figure 6: Results using the HMM(6) as generating model.

α	L	Δ_{ML} (dB)	Δ_{MAP} (dB)	Δ_{TURBO} (dB)
2	3	$14.94(-2.9)$	$+0.01(+0.0)$	$-3.25(-1.0)$
	6	$16.60(-2.7)$	$-0.01(-0.0)$	$-2.56(-0.9)$
1	3	$18.67(-2.9)$	$-0.81(-0.4)$	$-0.23(-0.2)$
	6	$19.57(-2.9)$	$-0.58(-0.3)$	$-0.23(-0.2)$

Table 2: Δ_{ML}: performance loss (w.r.t. $D_L(R)$) of ML decoding, with HMM(L) and HMM(1) as the generating model and the identification model respectively (in parentheses, magnitude of P_e); $\Delta_{MAP}(\Delta_{TURBO})$: difference of the performance loss of MAP(TURBO) decoding (in parentheses, relative magnitude of P_e) w.r.t. ML(MAP) decoding.

decoding offers a smaller but sensible decrease of the error probability and variance, that amounts up to 0.6 orders of magnitude and 1 dB respectively.

5. CONCLUSION

An iterative syndrome decoding algorithm for coding with side information over time-varying correlation channels has been proposed that takes advantage of the HMM that models the channel statistics. If the encoder has no knowledge about the channel and cannot rely on feedback from the decoder, this algorithm performs noticeably better than traditional ML decoding. In order to extend its application to real scenarios, future research directions include investigating more specific correlation models and devising procedures for estimation of the actual model parameters at the decoder.

REFERENCES

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

[2] S. S. Pradhan and K. Ramchandran, "Distributed source coding using syndromes (DISCUS): design and construction," *IEEE Trans. Inf. Theory*, vol. 49, no. 3, pp. 626–643, Mar. 2003.

[3] L. Cappellari and G. A. Mian, "A practical algorithm for distributed source coding based on continuous-valued syndromes," in *Proc. of European Signal Process. Conf. (EUSIPCO)*, Florence, Italy, 4-8 Sep. 2006.

[4] B. Girod, A. M. Aaron, S. Rane, and D. Rebollo-Monedero, "Distributed video coding," *Proc. IEEE*, vol. 93, no. 1, pp. 71–83, Jan. 2005.

[5] L. Cappellari, "Wavelet-domain distributed video coding based on continuous-valued syndromes," in *Proc. of European Signal Process. Conf. (EUSIPCO)*, Poznań, Poland, 3-7 Sep. 2007, pp. 1422–1426.

[6] L. R. Rabiner, "A tutorial on hidden Markov models and selected applications in speech recognition," *Proc. IEEE*, vol. 77, no. 2, pp. 257–286, Feb. 1989.

[7] S. Xiao, I. Kozintsev, and K. Ramchandran, "Stochastic wavelet-based image modeling using factor graphs and its application to denoising," in *Proc. of IEEE Intl. Conf. on Acoust., Speech and Signal Process.*, vol. 6, Jun. 2000, pp. 2187–2190.

[8] D. Varodayan, A. Aaron, and B. Girod, "Exploiting spatial correlation in pixel-domain distributed image compression," in *Proc. of Pict. Coding Symp.*, Apr. 2006.

[9] F. R. Kschischang, B. J. Frey, and H.-A. Loeliger, "Factor graphs and the sum-product algorithm," *IEEE Trans. Inf. Theory*, vol. 47, no. 2, pp. 498–519, Feb. 2001.

[10] L. Cappellari, "Distributed source coding using continuous-valued syndromes," *IEEE Trans. Inf. Theory*, 2007, submitted. [Online]. Available: http://arxiv.org/pdf/0710.1949

[11] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. Hoboken, NJ, USA: John Wiley & Sons, Inc., 2006.

[12] M. W. Marcellin and T. R. Fisher, "Trellis coded quantization of memoryless and Gauss-Markov sources," *IEEE Trans. Commun.*, vol. 38, no. 1, pp. 82–93, Jan. 1990.