Joint Source-Channel Coding using Real BCH Codes for Robust Image Transmission

A. Gabay[1], M. Kieffer[2,*], and P. Duhamel[2]

[1] Bloophone S.A., 96 Bd de sébastopol, F-75003 Paris

France Phone: 00 33 1 44 54 05 50, Fax : 00 33 1 44 54 05 15

a.gabay@bloophone.com

[2] LSS – CNRS – Supélec – Université Paris-Sud

Plateau de Moulon, F-91192 Gif-sur-Yvette, France

Phone: 00 33 1 69 85 17 32, Fax : 00 33 1 69 85 17 65

{michel.kieffer,piere.duhamel}@lss.supelec.fr

Abstract

In this paper, a new still image coding scheme is presented. In contrast with standard tandem coding schemes, where the redundancy is introduced after source coding, it is introduced before source coding using real BCH codes. A joint channel model is first presented. The model corresponds to a memoryless mixture of Gaussian and Bernoulli-Gaussian noise. It may represent the source coder, the channel coder, the physical channel, and their corresponding decoder. Decoding algorithms are derived from this channel model and compared to a state-of-art real BCH decoding scheme. A further comparison with two reference tandem coding schemes and the proposed joint coding scheme for the robust transmission of still images has been presented. When the tandem scheme is not accurately tuned, the joint coding scheme outperforms the tandem scheme in all situations. Compared to a tandem scheme well-tuned for a given channel situation, the joint coding scheme shows an increased robustness as the channel conditions worsen.

EDICS Category: COM-SOUR,COM-ERRO

I. Introduction

The classical approach to solve the problem of reliably sending compressed images or video over a noisy channel is to use a *tandem* source-channel coding (TSCC) scheme. Redundancy is inserted

* Corresponding author

Parts of this work have been presented at Globecom 2000 [1], ICASSP 2001 [2], ICC 2002 [3], and ICC 2003 [4]

at the output of the source coder by a channel coder with the intent that the channel decoder will detect and correct all erroneously received data. Thus, protecting against errors results in an increase in bandwidth. Sometimes however, the channel decoder is not able to correct all transmission errors, due to complexity or delay constraints, for example, and this results in highly corrupted reconstructed data with almost all efficient source encoders. On the contrary, if the channel is not too noisy, the excess bandwidth required by channel coding is wasted since the channel decoder does not consider the distortions introduced by source coding. The problem here is that classical communication systems are tuned for a precise target situation, and that a deviation of this situation from the nominal one may result in suboptimality.

Shannon's classical separation result states that the end-to-end performance of a transmission scheme can be optimized by separately optimizing the source encoder-decoder pair and the channel encoder-decoder pair [5]. However, this result holds only in the limit of infinite channel code block length and for stationary channels, two assumptions that are seldom met in practical situations such as mobile or satellite transmission: usually the delay is constrained, which makes the infinite block length assumption unreasonable, and the channels may be quite different at various locations, thus hitting the stationarity assumption. Furthermore, Shannon's theory does not provide a design algorithm for good channel codes with finite block length. In addition, it does not address the design of good source codes when the probability of channel error is non-zero, which is unavoidable for finite-length channel codes. Thus, for practical systems, in which the delay or complexity is constrained, a joint source and channel code design may reduce distortion.

In the last years, Joint Source-Channel Coding (JSCC) received an increasing interest in the research community, while having difficulties to enter practical applications. In [6], an overview of the various techniques available before 1996 may be found. Today, JSCC embeds a large variety of techniques which share the common property of not applying Shannon's separation principle. Albeit the frontiers between techniques are not so clear, one may distinguish four main families of techniques.

- In *concatenated coding* schemes, the Source Coder (SC) and Channel Coder (CC) are two separate units which are jointly optimized in order to get a minimum distortion for a given bitrate. Unequal error protection is the most commonly used technique of this family. Subband image coders or hybrid video coders [7] generate informations which can be grouped in classes according to their sensitivity to noise. Very sensitive data are strongly channel encoded whereas weak channel codes are used for the less sensitive data. See, *e.g.*, [8] for a technique using mainly the modulation and [9], [10] for CC which may easily be adapted to channel conditions or data sensitivity.

- In schemes where the *source coder is optimized for the channel conditions*, the channel is explicitly taken into account in the design of the SC. First results have been obtained by [11], who optimizes the index assignment in order to mitigate the effect of errors introduced by the channel. Multiple description schemes also belong to this family. More precisely, the idea is to optimize the redundancy

between descriptions with the channel conditions, see, *e.g.*, [12]. Entropy codes with error-detecting or correcting capabilities also belongs to this family, see [13] for reversible variable-length codes and [14], [15] for variable-length codes optimized according to the channel conditions.

- In *joint decoding* schemes, information provided by the channel decoder is used to improve the decoding realized by the source decoder [16]. All techniques belonging to this family use the fact that the source coder leaves some redundancy in the bitstream to perform a decoding using channel decoding techniques. In [21], a MAP detection schemes is proposed to mitigate transmission errors, taking the redundancy of images into account via a second-order Markov process. This requires the accurate transmission of the Markov model to the receiver. In [22], a Baum-Welch algorithm is put at work to perform jointly the decoding and the estimation of the parameters of a hidden Markov model for a binary source. See also [17]–[20] for decoding of variable-length encoded data using residual source redundancy.

- In *joint coding* schemes, the location of the CC or even the distinction between SC and CC is not clear anymore. Punctured convolutional codes have been used to realize source coding [23]. The choice of the puncturing matrix makes it possible to smoothly transform a CC in a SC [24], [25]. In a similar flavor, [26] realize the SC and CC with nonlinear transforms with dimensionality change inspired from [27]. A dimension increase corresponds to a channel coding whereas a decrease corresponds to a source coding.

The solution proposed in this paper belongs to the latter family. It consists in using a JSCC scheme which introduces redundancy *before* source coding using real BCH codes. The main advantage of this technique is the ability of the joint source-channel decoder (JSCD) to correct large errors resulting from non perfect transmission *as well as* reducing the level of errors resulting from source coding. Real BCH codes have been introduced in [28], [29], where the analogy between real-valued and finite-field block codes is established. It will be clear from Section III-C that the main difficulty with real-number codes in this situation comes from finite-precision computation and quantization of the encoded data. Some entries of the data block to be decoded may be affected by some high amplitude perturbations (*impulse noise*), but *all* components suffer from quantization or computation noise (*background noise*). In [30], the background noise is neglected and the Berlekamp-Massey (BM) and Forney [31] decoding algorithms for finite-fields BCH codes are directly transposed to real-valued DCT codes, a subclass of which are real BCH codes. However, in actual coding schemes, the background noise is unavoidable. Thus, the error-correcting algorithms developed for finite-field BCH codes cannot be used directly. In [32], a decoding technique derived from the Peterson, Gorenstein, Zierler (PGZ) algorithm is applied for burst of erasures recovery. Erasures locations are assumed to be known. An analysis of the perturbation introduced by the background noise is also provided. The background noise is explicitly taken into account in [33], where an error correcting technique based on Wiener estimation is provided. The unknown impulse error locations are estimated using a multiple hypotheses test.

However, the corresponding results would clearly benefit from an accurate channel model describing the part of the communication scheme lying between the real BCH encoder and decoder. At the same time, [1] and [2] propose a Gaussian-Bernouilli-Gaussian (GBG) channel model derived from [34] to efficiently represent a *joint channel* consisting of a scalar quantizer followed by a binary symmetric channel and the corresponding inverse quantizer. The decoding is performed by a modified version of the PGZ algorithm presented in [35] and taking into account the information provided by the channel model. This technique has been improved using a multiple hypotheses test for estimating the number of impulse errors [4] and by realizing an *a posteriori* control of the correction technique [3], which are computationally less demanding than the techniques presented in [33]. Further results on real BCH codes have been obtained for code vectors corrupted by bursts of erasures (with known locations) in [36] where the background noise is taken into account. In [37], real BCH codes are embedded in the theory of real-valued frames, which allow a further analysis of the influence of the background noise on erasure recovery.

It has to be noted that the proposed approach is complementary to that proposed, *e.g.*, in [16], [21], [22]. Real BCH decoding is performed *after* all other decoding techniques have done their work, since the real BCH decoder is situated at the end of communication chain. The best the intermediate channel or joint-source-channel techniques will perform, the best the real BCH decoder will work too. Moreover, one of the unique specificity of real BCH codes is their ability to mitigate part of the noise introduced by the source coder.

The aim of this paper is to summarize and complement results partly presented by the authors in conference papers [1]–[4] on the use of real BCH codes for robust image transmission over realistic satellite channels. It will be shown that in many practical situations, a GBG model is well suited to represent the part of the communication scheme between the output of the real BCH coder and the input of the real BCH decoder. Thus, data encoded using a real BCH code are affected in the channel by a mixture of Gaussian background noise and impulse noise with Gaussian distributed amplitude. Three problems have then to be solved: (i) determining whether an encoded vector is affected by some impulse error and estimating their number; (ii) determining the amplitude and location of the impulse errors; (iii) realizing the correction and verifying its efficiency. Compared to the conference papers, a detailed presentation of the GBG channel model is provided, its capacity is evaluated and non-trivial illustrative examples are treated. A comparison of the proposed decoding technique to a state-of-art real BCH decoder is performed. Moreover, the usefulness of such techniques in practical satellite image transmission situations is evaluated. Obviously, these results can be extended to more general wireless communication channels.

The joint channel model is the key issue of the presented technique. Thus Section II is devoted to the description and analysis of the joint GBG channel model. In Section III, properties of real BCH codes are recalled, and coding and decoding techniques are presented. The way real BCH codes

may be introduced in actual still image coding schemes is considered in Section IV. Finally, some examples are presented in Section V.

II. JOINT CHANNEL MODEL

In standard channel coding, the channel is the part of the communication scheme that lies between the channel coder and the channel decoder. Similarly, for joint source-channel coding, define the *joint channel* as the part of the communication scheme between the JSCC and the JSCD, see Figure 1. A model of the joint channel is required to be able to evaluate the performance a JSCC/JSCD may attain and to simplify its design. Since it is a joint channel, the model has to account for the distortion introduced both by the source coder-decoder and the transmission.

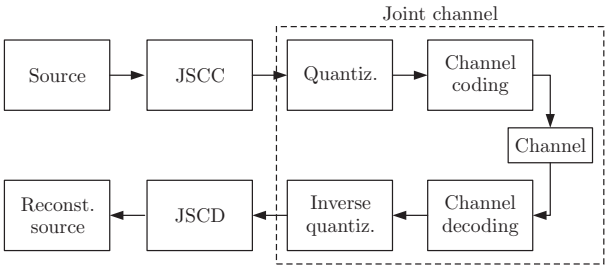

Fig. 1. Definition of the joint channel

A. Introduction

Assume that a source of probability density function (pdf) $p(\mathbf{s})$ generates vectors $\mathbf{s} = (s_1 \dots s_L)^{\mathrm{T}} \in \mathbb{R}^L$ which pass through a quantizer defined by (i) a dictionary $\mathcal{D} = \{\mathbf{d}_1 \dots \mathbf{d}_M\}$, (ii) a partition $\mathcal{P} = \{P_1 \dots P_M\}$ of $\mathbb{R}^L$ in M non-overlapping regions and (iii) a function Q such that

$$Q(\mathbf{s}) = \mathbf{d}_j \text{ if } \mathbf{s} \in P_j,\ j = 1, \dots, M.$$

A binary index $\mathbf{i} = I(\mathbf{d})$ is associated to the code vector $\mathbf{d}$ via the indexing function $I(\cdot)$ and is transmitted on a channel, which is assumed to be memoryless and characterized by its transition probability $p(\mathbf{i}_k|\mathbf{i}_j)$ or $p(\mathbf{d}_k|\mathbf{d}_j)$. At receiver side, if the received index is $\mathbf{i}$, the estimate for the source vector $\mathbf{s}$ is $\widehat{\mathbf{s}} = I^{-1}(\mathbf{i}) \in \mathcal{D}$. The distortion introduced by the quantizer and the channel is then

$$D = \frac{1}{L} \sum_{j=0}^{M} \sum_{k=0}^{M} p(\mathbf{d}_k|\mathbf{d}_j) \int_{P_j} p(\mathbf{s})\, d(\mathbf{s}, \mathbf{d}_k)\, d\mathbf{s}, \tag{1}$$

where $d(\mathbf{s}, \mathbf{d}_k)$ is the distortion measure between $\mathbf{s}$ and $\mathbf{d}_k$, see [11].

When the distortion measure is the quadratic euclidean norm, then (1) may be written as

$$D = E\left(|\mathbf{s} - \widehat{\mathbf{s}}|^2\right) = E\left(|\mathbf{s} - Q(\mathbf{s})|^2\right) + E\left(|Q(\mathbf{s}) - \widehat{\mathbf{s}}|^2\right) + 2.E\left((\mathbf{s} - Q(\mathbf{s}))^{\mathrm{T}} . (Q(\mathbf{s}) - \widehat{\mathbf{s}})\right). \tag{2}$$

The first term in (2) corresponds to the distortion introduced by the quantizer, the second, to the distortion introduced by the channel and the last one to the mutual distortion. For well-tuned quantizers, the last term can be neglected [38] to get

$$D \approx E\left(|\mathbf{s} - Q(\mathbf{s})|^2\right) + E\left(|Q(\mathbf{s}) - \widehat{\mathbf{s}}|^2\right). \tag{3}$$

B. New channel model

The joint channel model has to account for both types of distortion expressed in (3). The proposed *Gaussian-Bernoulli-Gaussian* (GBG) channel model [34] consists of a memoryless channel corrupted by a mixture of Gaussian noise with small variance and of Bernoulli-Gaussian (BG) noise. The Gaussian noise models the quantization noise and corresponds to the first term of (3). The BG noise represents the channel errors that were not corrected by the channel coder/decoder pair and corresponds to the second term of (3).

Assume that the input $c(k)$ and output $r(k)$ of the joint channel are related by

$$r(k) = c(k) + b(k) + \gamma(k) \tag{4}$$

where $b(k)$ is zero-mean Gaussian with variance σ_{g}^2 and $\gamma(k) = \xi(k)\eta_{\mathrm{i}}(k)$ is a BG noise with $\xi(k)$ a Bernoulli process, *i.e.*, an i.i.d. sequence of zeros and ones with Prob$(\xi(k) = 1) = p_{\mathrm{i}}$, and $\eta_{\mathrm{i}}(k)$ is zero-mean Gaussian with variance σ_{i}^2. Obviously, in most situations, $\sigma_{\mathrm{i}}^2 >> \sigma_{\mathrm{g}}^2$ since channel errors, when they occur, impact much more significantly on the distortion than quantization noise.

The pdf associated with $\gamma(k)$ is

$$p(\gamma) = (1 - p_{\mathrm{i}})\,\delta(\gamma) + p_{\mathrm{i}}\left(2\pi\sigma_{\mathrm{i}}^2\right)^{-1/2} \exp\left(-\frac{\gamma^2}{2\sigma_{\mathrm{i}}^2}\right). \tag{5}$$

The pdf of $\eta(k) = b(k) + \gamma(k)$ is thus a convolution of a Gaussian pdf and of (5)

$$p(\eta) = (1 - p_{\mathrm{i}})\left(2\pi\sigma_{\mathrm{g}}^2\right)^{-1/2} \exp\left(-\frac{\gamma^2}{2\sigma_{\mathrm{g}}^2}\right) + p_{\mathrm{i}}\left(2\pi\left(\sigma_{\mathrm{g}}^2 + \sigma_{\mathrm{i}}^2\right)\right)^{-1/2} \exp\left(-\frac{\gamma^2}{2\left(\sigma_{\mathrm{g}}^2 + \sigma_{\mathrm{i}}^2\right)}\right). \tag{6}$$

This model may seem quite rough, since it is well known, for example, that quantization noise is far from being Gaussian, but it is shown in Section II-C that it can describe quite accurately many practical situations.

This GBG joint channel model is thus characterized by three parameters p_{i}, σ_{g}^2 and σ_{i}^2 which have to be estimated for a given channel. They may be obtained in three steps.

1) First evaluate the distortion D_{q} introduced by the source coder assuming that the channel is error-free. This distortion is modeled by the Gaussian part of the GBG model. An estimate of the variance of the Gaussian noise may then be chosen as $\widehat{\sigma}_{\mathrm{g}}^2 = D_{\mathrm{q}}$.
2) Then consider the noise introduced by the channel and evaluate the probability that a sample is in error after source decoding, as a function of the channel, the channel coder-decoder, and

the source coder-decoder parameters. This leads to an estimate $\widehat{p}_{\mathrm{i}}$ of the parameter p_{i} in the BG noise.

3) Finally, evaluate the distortion D_{i} introduced by the channel error on the reconstructed samples after source decoding. Using (5), the distortion introduced by the BG noise would be

$$\widehat{D}_{\mathrm{i}} = (1-p_{\mathrm{i}}) \int_{-\infty}^{+\infty} \gamma^2 \delta(\gamma)\, d\gamma + p_{\mathrm{i}} \int_{-\infty}^{+\infty} \gamma^2 \left(2\pi\sigma_{\mathrm{i}}^2\right)^{-1/2} \exp\left(-\frac{\gamma^2}{2\sigma_{\mathrm{i}}^2}\right) dy = p_{\mathrm{i}}\sigma_{\mathrm{i}}^2. \quad (7)$$

An estimate of the variance of the Gaussian noise in the BG noise may be obtained as

$$\widehat{\sigma}_{\mathrm{i}}^2 = D_{\mathrm{i}}/\widehat{p}_{\mathrm{i}}. \quad (8)$$

C. Examples

The accuracy of the proposed model is verified below through some examples. Another example may be found in Appendix I.

1) Scalar quantizer followed by a BSC: Consider a source s quantized by an $M = 2^b$ levels scalar quantizer before being transmitted over a memoryless BSC with transition probability ε. The scalar quantizer, with dictionary $\mathcal{D} = \{d_1, \ldots, d_M\}$, introduces a distortion $D_{\mathrm{q}} = E\left(|s - Q(s)|^2\right)$, which may be modeled by the zero-mean Gaussian noise b in (4) of variance

$$\widehat{\sigma}_{\mathrm{g}}^2 = D_{\mathrm{q}}. \quad (9)$$

The BSC introduces additional distortion, since bit errors on the channel result in index errors at the decoder. The transition probability $p(\mathbf{i}_k|\mathbf{i}_j) = p(d_k|d_j)$ may be expressed as

$$p(\mathbf{i}_k|\mathbf{i}_j) = \varepsilon^{d_{\mathrm{H}}(\mathbf{i}_j,\mathbf{i}_k)} (1-\varepsilon)^{b-d_{\mathrm{H}}(\mathbf{i}_j,\mathbf{i}_k)}, \quad (10)$$

where $d_{\mathrm{H}}(\cdot,\cdot)$ stands for the Hamming distance. Using (10) and taking into account the source decoder, the distortion between the quantizer output d and its estimate $\widehat{d}$ after source decoding may be written as

$$D_{\mathrm{i}} = E\left(\left|d-\widehat{d}\right|^2\right) = \sum_{j=1}^{M}\sum_{k=1}^{M} \varepsilon^{d_{\mathrm{H}}(\mathbf{i}_j,\mathbf{i}_k)} (1-\varepsilon)^{b-d_{\mathrm{H}}(\mathbf{i}_j,\mathbf{i}_k)} p(d_j) |d_j - d_k|^2, \quad (11)$$

An estimate for the probability p_{i} of the GBG model is the probability that an index is in error

$$\widehat{p}_{\mathrm{i}} = 1 - (1-\varepsilon)^b. \quad (12)$$

By combining (8), (11) and (12), one then gets an estimate for the parameter σ_{i}^2 of the GBG model

$$\widehat{\sigma}_{\mathrm{i}}^2 = \frac{\sum_{j=1}^{M}\sum_{k=1}^{M} p(d_j)\, \varepsilon^{d_{\mathrm{H}}(\mathbf{i}_j,\mathbf{i}_k)} (1-\varepsilon)^{b-d_{\mathrm{H}}(\mathbf{i}_j,\mathbf{i}_k)} p(d_j) |d_j - d_k|^2}{1-(1-\varepsilon)^b}. \quad (13)$$

For a zero-mean unit-variance Gaussian source, Figure 2 illustrates the evolution of $\sigma_{\mathrm{i}}^2/\sigma_{\mathrm{g}}^2$ as a function of the bitrate b for a given transition probability $\varepsilon = 10^{-2}$ of the BSC. This ratio does not significantly depend of the type of quantization used (uniform and non uniform scalar quantizers).

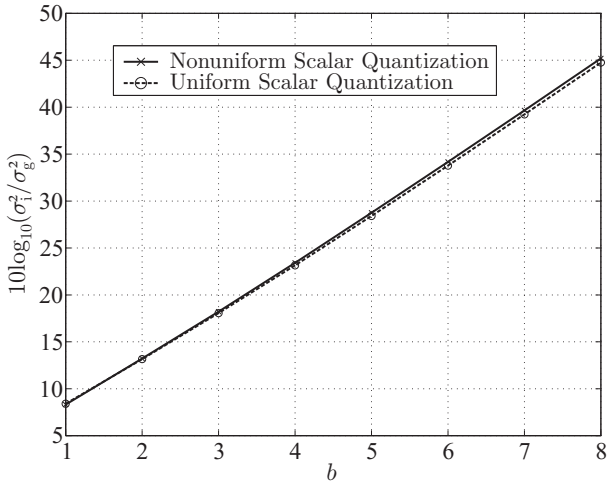

Fig. 2. Relation between quantization and impulse noise variances for uniform and nonuniform scalar quantization followed by a BSC with $\varepsilon = 10^{-2}$

For a scalar quantizer followed by a BSC, considering γ as a high amplitude impulse noise compared to the quantization noise is thus quite reasonable, even at low bitrates.

A comparison of the input-output behavior of a joint channel consisting of a nonuniform scalar quantizer followed by a BSC and an inverse quantizer versus its GBG model is represented on Figure 3. A zero-mean unit-variance Gaussian source has been quantized on M levels, with $M \in \left\{2^1, 2^2, 2^4, 2^6\right\}$. For all quantizers, the difference between the actual channel and its model is less

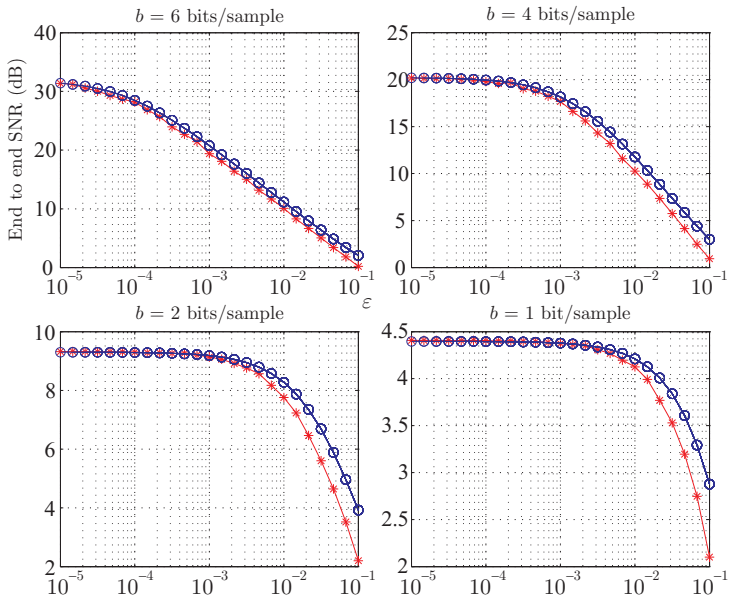

Fig. 3. Comparison of a joint channel (*) and its GBG model (o) as a function of ε (scalar quantization followed by a BSC and an inverse quantization).

than 2 dB. Moreover, below $\varepsilon = 10^{-3}$, it is less than 0.5 dB, except for $M = 2^6$ where it is about 1 dB. This shows that the GBG model is realistic for a large interval of ε and for many scalar

quantizers. The differences come from the fact that, at high values of ε, the quantization noise and the noise introduced by the channel are correlated and the third term in (2) should not be neglected, see [11], [39].

2) Scalar quantization associated with an MPSK followed by an AWGN channel: Consider the same source s that passes through the same scalar quantizer but now followed by an MPSK sending bits over an AWGN channel characterized by its variance $N_0/2$. In the signal space, M points $\{\mathbf{u}_1, \ldots, \mathbf{u}_M\}$ form the constellation of the MPSK.

An estimate of the first parameter σ_{g}^2 of the GBG model (6) is still given by (9).

Assuming that $\mathbf{u}_j$ corresponds to the output level d_j of the scalar quantizer, the probability of transition from $\mathbf{u}_j$ to $\mathbf{u}_k$ may bounded as [40]

$$p\left(\mathbf{u}_k|\mathbf{u}_j\right) \leqslant \frac{\operatorname{erfc}\left(\left|\mathbf{u}_j - \mathbf{u}_k\right| / \left(2\sqrt{N_0}\right)\right)}{2}. \tag{14}$$

An estimate of p_{i} in the GBG model is then obtained by using the union bound [40] and by averaging all transition probabilities (14) to get

$$\widehat{p}_{\mathrm{i}} = \sum_{j=1}^{M} p\left(d_j\right) \sum_{k=1, k\neq j}^{M} \frac{\operatorname{erfc}\left(\left|\mathbf{u}_j - \mathbf{u}_k\right| / \left(2\sqrt{N_0}\right)\right)}{2}. \tag{15}$$

The resulting distortion between the quantizer output d and its estimate $\widehat{d}$ after demodulation and source decoding is

$$D_{\mathrm{i}} = E\left(\left|d - \widehat{d}\right|^2\right) = \sum_{j=1}^{M} \sum_{k=1, k\neq j}^{M} p\left(\mathbf{u}_k|\mathbf{u}_j\right) p\left(d_j\right) \left|d_k - d_j\right|^2. \tag{16}$$

An estimate of the variance of the Gaussian part of the BG process may again be evaluated using (8), (15) and (16)

$$\widehat{\sigma}_{\mathrm{i}}^2 = \frac{\sum_{j=1}^{M} \sum_{k=1, k\neq j}^{M} \operatorname{erfc}\left(\left|\mathbf{u}_j - \mathbf{u}_k\right| / \left(2\sqrt{N_0}\right)\right) p\left(d_j\right) \left|d_k - d_j\right|^2}{2 \sum_{j=1}^{M} p\left(d_j\right) \sum_{k=1, k\neq j}^{M} \frac{\operatorname{erfc}\left(\left|\mathbf{u}_j - \mathbf{u}_k\right| / \left(2\sqrt{N_0}\right)\right)}{2}} \tag{17}$$

To verify the accuracy of this model, consider a zero-mean unit-variance Gaussian source quantized using an M levels non-uniform scalar quantizer. Its output indexes are mapped on an MPSK and sent over an AWGN channel followed by an ML receiver and an inverse quantizer. Figure 4 compares the performance of the joint channel and of its GBG model when $M = 2^3$ and $M = 2^4$.

In both cases, except at low E_{b}/N_0, where the GBG channel is more pessimistic (mostly due to the pessimism of the union bound), the GBG model is again quite realistic to represent the perturbations introduced by the quantization, modulation and channel.

3) Influence of a channel coder: In the previous examples, no channel coder has been considered. The easiest way to account for its influence is to build an equivalent BSC channel model of the group formed by the channel coder, channel and channel decoder. A new transition probability ε' may be evaluated and all previous computations may be performed in a similar way.

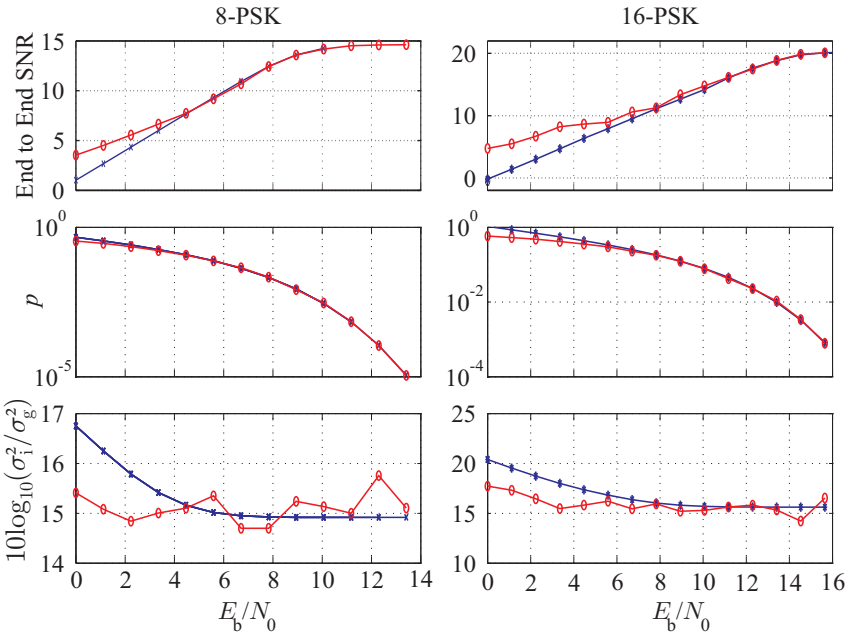

Fig. 4. Comparison of a joint channel (*) and its GBG model (o), ($M = 2^3$ and $M = 2^4$ levels scalar quantization followed by an MPSK, an AWGN channel, an ML receiver and an inverse quantization)

D. Channel capacity

Using the Arimoto-Blahut algorithm [41], it is possible to compute the capacity of the GBG joint channel model for various p_i or $\sigma_\mathrm{i}^2/\sigma_\mathrm{g}^2$. In the considered case, $\sigma_\mathrm{g}^2 = 10^{-1}$ and $\sigma_\mathrm{i}^2 = 1$. For various values of $p_\mathrm{i} \in \left\{10^{-3}, 10^{-2}, 5.10^{-2}, 10^{-1}\right\}$, the capacity is plotted on Figure 5 as a function of the input signal power P. The AWGN channel is the reference corresponding to $p_\mathrm{i} = 0$.

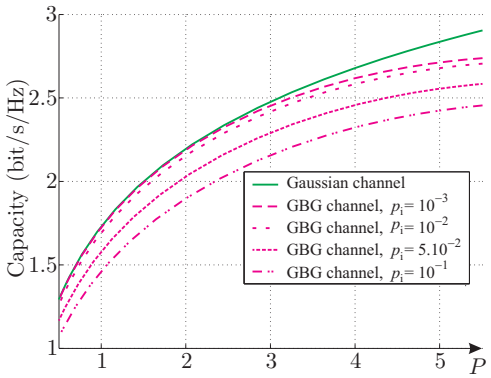

Fig. 5. GBG channel capacity as a function of the input signal power P, for various p ($\sigma_\mathrm{q}^2 = 0.1$, $10\log_{10}\frac{\sigma_\mathrm{i}^2}{\sigma_\mathrm{q}^2} = 10$ dB)

For a moderate input signal power, even with the quite high impulse error probability of 10^{-2}, the channel capacity remains very similar to that of an AWGN channel. This suggests that a well-designed JSCC/JSCD may mitigate the effect of impulse errors in the GBG channel.

E. Comments on the channel model

The purpose of this section was to show that, despite the complexity of the global system, a GBG channel model could describe the impact of the source coding plus that of the residual transmission errors with a very good accuracy in most cases, and could at least give an order of magnitude of the global distortion depending on the system parameters. This is a key point in our approach, and represents one of the main differences with [33], since we heavily rely on this simple model to build efficient decoders.

Note also that our channel model clearly imposes a precise definition of what an *error* is. Because of the presence of a Gaussian noise on all samples and because the emitted signal is continuous, there is no hope that this background noise can be corrected. However, one can hope to get rid of the BG part, which has few contributions (in terms of erroneous samples). Thus, the *error correction capacity* in what follows refers to the ability to correct impulse errors, despite the presence of background noise.

III. Real BCH codes

This section briefly recalls BCH codes on the real field. Encoding and decoding techniques are presented. The way they may be incorporated in JSCC/JSCD schemes to mitigate the effect of a GBG channel is also illustrated.

A. Definition

The definitions of spectral and BCH codes over finite fields given, *e.g.*, in [31] may be translated to define BCH(n,k) codes over the real field (real BCH(n,k) codes) as spectral codes whose elements belong to $\mathbb{R}^n$ and are such that their spectrum vanishes over a set of *consecutive parity frequencies*

$$\mathcal{A} = \{a_0, \ldots, a_0 + n - k - 1\},$$

satisfying for all $f \in \mathcal{A}$, $n - f \in \mathcal{A}$, and where a_0 is the base index, see [29], [33].

Various techniques may be employed to perform a real BCH(n,k) encoding. One could think of polynomial constructions inspired from classical finite fields cyclic codes. This paper focuses on the spectral interpretation and construction of real BCH(n,k) codes with k odd. Our construction also includes some properties that will be useful in our context (interpolation property).

Consider a *real* information vector $\mathbf{x}_k \in \mathbb{R}^k$, an associated BCH$(n,k)$ code vector $\mathbf{c}_n \in \mathbb{R}^n$ may be built as follows

$$\mathbf{c}_n = \mathbf{W}_n \mathbf{P}_{n,k} \mathbf{W}_k^{-1} \mathbf{x}_k, \tag{18}$$

where $\mathbf{W}_n$ is the $n-$dimensional discrete Fourier transform matrix and where

$$\mathbf{P}_{n,k} = \begin{pmatrix} \mathbf{I}_{\lceil n/2 \rceil - t} & \mathbf{0}_{\lceil n/2 \rceil - t, k - \lceil n/2 \rceil + t} \\ \mathbf{0}_{n-k, \lceil n/2 \rceil - t} & \mathbf{0}_{n-k, k - \lceil n/2 \rceil + t} \\ \mathbf{0}_{k - \lceil n/2 \rceil + t, \lceil n/2 \rceil - t} & \mathbf{I}_{k - \lceil n/2 \rceil + t} \end{pmatrix},$$

is an $n \times k$ zero-padding matrix. In $\mathbf{P}_{n,k}$, $t = \lfloor (n-k)/2 \rfloor$ represents the error correcting capacity of the code, $\mathbf{I}_n$ is the $n \times n$ identity matrix and $\mathbf{0}_{n,m}$ is the $n \times m$ null matrix.

Remark 1: In contrast with previous constructions of real BCH codes, the encoding (18) produces a code vector $\mathbf{c}_n$ which is a spectral interpolated version of $\mathbf{x}_k$. The mean value and variance of each component of $\mathbf{c}_n$ are the same as those of the components of $\mathbf{x}_k$. This property is particularly useful in an image coding context: an image after real-BCH encoding will *look like* the original image, allowing image coders to perform well on the BCH-encoded image. Moreover, (18) introduces correlation between samples of $\mathbf{c}_n$, providing a redundant representation of $\mathbf{x}_k$ helpful to correct transmission errors.

A *parity-check* matrix for this code is

$$\mathbf{H}_{n-k,n} = \mathbf{R}_{n-k,n}\mathbf{W}_n^{-1},$$

with $\mathbf{R}_{n-k,n} = \begin{pmatrix} \mathbf{0}_{n-k,\lceil n/2 \rceil - t} & \mathbf{I}_{n-k} & \mathbf{0}_{n-k,k-\lceil n/2 \rceil + t} \end{pmatrix}$. When $\mathbf{c}_n$ is transmitted over a noisy channel and $\mathbf{r}_n = \mathbf{c}_n + \boldsymbol{\eta}_n$ is received, the *syndrome* associated to $\mathbf{r}_n$ is the complex vector

$$\mathbf{s}_{n-k}(\mathbf{r}_n) = \mathbf{H}_{n-k,n}\mathbf{r}_n = \mathbf{H}_{n-k,n}\boldsymbol{\eta}_n \neq \mathbf{0}. \tag{19}$$

When the channel does not introduce any noise, $\mathbf{s}_{n-k}(\mathbf{r}_n) = \mathbf{0}$.

In the next sections, the code vectors are assumed to be transmitted over communication channels described by a GBG model with known characteristics (p_{i}, σ_{g}^2 and σ_{i}^2), see Section II. Two decoding algorithms will be presented.

B. Projection decoding

Consider $\mathbf{x}_k \in \mathbb{R}^k$ and its real BCH(n,k) code vector $\mathbf{c}_n$, which is transmitted over a GBG channel. Assume that $\mathbf{r}_n \in \mathbb{R}^n$ is received. Usually, $\mathbf{r}_n$ does not belong to the code, *i.e.*, $\mathbf{H}_{n-k,n}\mathbf{r}_n \neq \mathbf{0}$. The simplest decoding approach consists in projecting $\mathbf{r}_n$ onto the code subspace to obtain an estimate $\widehat{\mathbf{c}}_n$ of the initial code vector, which is then decoded. The estimate $\widehat{\mathbf{c}}_n$ is simply obtained by canceling all entries of the Fourier transform of $\mathbf{r}_n$ whose indexes belong to $\mathcal{A}$, *i.e.*, $\widehat{\mathbf{c}}_n = \mathbf{W}_n\boldsymbol{\Pi}_n\mathbf{W}_n^{-1}\mathbf{r}_n$, where $\boldsymbol{\Pi}_n = \mathrm{diag}\begin{pmatrix} \mathbf{I}_{1,\lceil n/2 \rceil - t} & \mathbf{0}_{1,n-k} & \mathbf{I}_{1,k-\lceil n/2 \rceil + t} \end{pmatrix}$.

Remark 2: The information vector obtained by projection corresponds to the maximum-likelihood estimate of $\mathbf{x}_k$ from $\mathbf{r}_n$, under the assumption that the GBG channel introduces only background (Gaussian) noise. Obviously, this criterion can be improved by introducing an error correction on the impulses.

C. Decoding with correction of the impulse errors

Assume now that at receiver side $\mathbf{r}_n = \mathbf{c}_n + \mathbf{b}_n + \boldsymbol{\gamma}_n$, where $\mathbf{b}_n$ stands for the Gaussian background noise and $\boldsymbol{\gamma}_n$ is the impulse noise vector containing ν non-zero components at locations $\boldsymbol{\rho}^{(\nu)} =$

$(\rho_1 \dots \rho_\nu)$. The presence of $\boldsymbol{\gamma}_n$ may produce a poor estimate of $\mathbf{x}_k$ when using a projection decoding. This section proposes a variant of the PGZ algorithm for BCH codes on finite fields [31] to remove $\boldsymbol{\gamma}_n$ from $\mathbf{r}_n$. The approach is similar to the two techniques for estimating the characteristics of the impulse noise presented in [33]. The first technique of [33] involves (when an error occurs) many hypotheses tests to perform an exhaustive search of the locations of the impulses. The second, less accurate, uses a variant of the BM algorithm, which is of the same computational complexity as the technique presented here and will thus serve as a reference for comparisons performed in Section III-D. The main difference with [33] is that the joint channel parameters are explicitly taken into account, which allows a more accurate correction in difficult cases.

The syndrome associated with $\mathbf{r}_n$ is

$$\begin{aligned} \mathbf{s}_{n-k}(\mathbf{r}_n) &= \mathbf{H}_{n-k,n}\mathbf{r}_n = \mathbf{R}_{n-k,n}\mathbf{W}_n^{-1}\boldsymbol{\gamma}_n + \mathbf{R}_{n-k,n}\mathbf{W}_n^{-1}\mathbf{b}_n \\ &= (S_{a_0}, \dots, S_{a_0+n-k-1})^{\mathrm{T}}. \end{aligned} \tag{20}$$

First, the number ν and then the locations of the impulse errors have to be determined. Then, a maximum-likelihood (ML) or maximum *a posteriori* (MAP) estimation of the amplitude of the impulse errors is simply derived. At last, *a posteriori* control method of [33] or [3] may be used to verify the efficiency of the correction.

1) Estimation of the number of impulse errors: With the *a priori* probability p_{i} of occurrence of an impulse error in each entry of $\mathbf{r}_n$, the syndrome (20) is the only available information to estimate the number of impulse errors. Several hypotheses may be formulated:

- H_0: there is no impulse error in $\mathbf{r}_n$
- H_1: there is a single impulse error in $\mathbf{r}_n$ ($\nu = 1$)
- $\dots$
- H_t: there are t impulse errors in $\mathbf{r}_n$ ($\nu = t$)
- H_{t+1}: more than t impulse errors have occurred and the correction capacity is overflowed.

Using p_{i}, the *a priori* probabilities π_ν, $\nu = 0 \dots t+1$ are easily evaluated for each hypothesis

$$\pi_\nu = p(H_\nu) = \begin{cases} \binom{\nu}{n} p_{\mathrm{i}}^{\nu} (1-p_{\mathrm{i}})^{n-\nu} & \text{for } 0 \leqslant \nu \leqslant t \\ \sum_{\nu=t+1}^{n} \binom{\nu}{n} p_{\mathrm{i}}^{\nu} (1-p_{\mathrm{i}})^{n-\nu} & \text{for } \nu > t. \end{cases}$$

Bayesian multiple-hypotheses testing (BMHT), see, *e.g.*, [42], is then a natural decision tool to determine the number of impulse errors in $\mathbf{r}_n$. Here, the norm of the syndrome

$$y(\mathbf{r}_n) = \|\mathbf{s}_{n-k}(\mathbf{r}_n)\|^2$$

will be used. Deciding directly on the syndrome, as in [33] is possible, but computationally much more demanding. Moreover, it has been observed in [3] that this choise was not reducing the performance.

The key point of BMHT is the evaluation of the pdf $p(y(\mathbf{r}_n)|\nu)$ for all values of ν. Intuitively, in the absence of errors, $y(\mathbf{r}_n)$ approximatively consists of the sum of n squared Gaussian variables

with the same variance σ_{g}^2, and thus follows a χ^2 distribution with n degrees of freedom. In the presence of a single error, an additional squared gaussian variable with variance σ_{i}^2 has to be taken into account. The resulting distribution is no more χ^2, nevertheless its mean will be larger than that without error. More details may be found in Appendix III.

Once all $p\left(y\left(\mathbf{r}_n\right)|\nu\right)$s have been evaluated, the positive real line $\mathbb{R}^+$ may be partitioned into $t+2$ disjoint decision regions Z_i, one for each hypothesis. Introducing some costs for the wrong decisions, the partitioning is performed by minimizing the a risk function accounting for this costs. Note that all these operations have only to be done once and may be performed off-line.

2) Estimation of the locations: Assume that $\nu \leqslant \lfloor (n-k)/2 \rfloor$ is estimated or known and that $\boldsymbol{\rho}^{(\nu)} = (\rho_1, \ldots, \rho_\nu)$ is the vector of indexes at which $\boldsymbol{\gamma}_n$ is not null. To estimate the locations of all non-zero entries of $\boldsymbol{\gamma}_n$ from the windowed and noisy observations of the inverse Fourier transform of $\boldsymbol{\gamma}_n$ provided by (20), one may use the error-localization polynomial introduced by the PGZ algorithm [31]

$$\Lambda(x) = \prod_{\ell=1}^{\nu} \left(1 - x.\omega^{\rho_\ell}\right) = 1 + \lambda_1 x + \cdots + \lambda_\nu x^\nu, \tag{21}$$

which vanishes at $\omega^{-\rho}$ when ρ belongs to $\boldsymbol{\rho}^{(\nu)}$, with $\omega = e^{-2i\pi/n}$. Using a modified version of the PGZ algorithm for BCH codes on finite fields detailed in Appendix II, an estimate $\widehat{\boldsymbol{\lambda}}_\nu$ of the vector $\boldsymbol{\lambda}_\nu = (\lambda_1, \ldots, \lambda_\nu)^{\mathrm{T}}$ of coefficients of $\Lambda(x)$ can be computed as

$$\widehat{\boldsymbol{\lambda}}_\nu = -\left(\mathbf{S}_{n-k-\nu,\nu}^{\mathrm{T}}.\mathbf{S}_{n-k-\nu,\nu}^{*} - \frac{(n-k-\nu)}{n}\sigma_{\mathrm{g}}^2\mathbf{I}_\nu\right)^{-1}\mathbf{S}_{n-k-\nu,\nu}^{\mathrm{T}}.\mathbf{S}_{n-k-\nu,1}^{*}, \tag{22}$$

with $\mathbf{S}_{n-k-\nu,1} = \left(S_{a_0+\nu} \ldots S_{a_0+n-k-1}\right)^{\mathrm{T}}$ and

$$\mathbf{S}_{n-k-\nu,\nu} = \begin{pmatrix} S_{a_0+\nu-1} & \cdots & S_{a_0} \\ \vdots & \ddots & \vdots \\ S_{a_0+n-k-2} & \cdots & S_{a_0+n-k-\nu+1} \end{pmatrix}.$$

The main difference with the standard PGZ algorithm comes from the $-(n-k-\nu)/n.\sigma_{\mathrm{g}}^2\mathbf{I}_\nu$ term in (22), which accounts for the presence of the Gaussian noise.

Due to this background noise, the indexes where $\boldsymbol{\gamma}_n$ is non-zero are estimated as the ν smallest absolute values of the localization polynomial evaluated at $\omega^{-\rho}$, with $\rho = 0, \ldots, n-1$. The following example illustrates the influence of the background noise on the error localization accuracy.

Example 1: Consider a real BCH$(21, 15)$ code. Figure 6 shows that in absence of Gaussian background noise, the standard PGZ algorithm estimates precisely the locations of the errors. In presence of background noise (here such that $10\log_{10}\sigma_{\mathrm{i}}^2/\sigma_{\mathrm{g}}^2 = 10$ dB), this localization may become more difficult. Using (22) significantly improves the localization performance, as illustrated by Figure 6, where it is clearly seen that if $\sigma_{\mathrm{g}}^2 = 1$ is not taken into account, one misses a minimum of the curve, which is equivalent to missing one error. Further analysis of the influence of round-off errors and of the background noise on the localization accuracy may be found in [43].

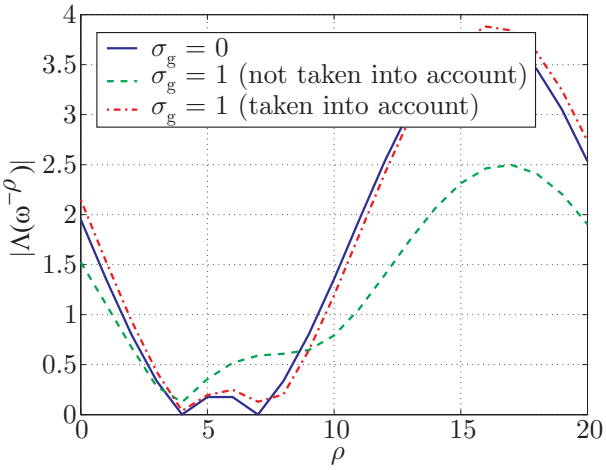

Fig. 6. Influence of the background noise on the error-localization polynomial

3) Estimation of the amplitudes: Assume that an estimate $\widehat{\boldsymbol{\rho}}^{(\nu)} = (\widehat{\rho}_1, \ldots, \widehat{\rho}_\nu)$ of $\boldsymbol{\rho}^{(\nu)}$ has been obtained. Using the vector of ν impulse errors $\boldsymbol{\gamma}_\nu = \left(\gamma_{\widehat{\rho}_1}, \ldots, \gamma_{\widehat{\rho}_\nu}\right)^{\mathrm{T}}$, (20) may be rewritten as

$$\mathbf{s}_{n-k}\left(\mathbf{r}_n\right) = \mathbf{V}_{n-k,\nu}\boldsymbol{\gamma}_\nu + \mathbf{R}_{n-k,n}\mathbf{W}_n^{-1}\mathbf{b}_n, \tag{23}$$

with $\mathbf{V}_{n-k,\nu} = \mathbf{R}_{n-k,n}\left(\begin{array}{ccc} \mathbf{w}_{\widehat{\rho}_1}^{\mathrm{T}} & \ldots & \mathbf{w}_{\widehat{\rho}_\nu}^{\mathrm{T}} \end{array}\right)^{\mathrm{T}}$, and where $\mathbf{w}_\rho$ is the ρ-th line of $\mathbf{W}_n^{-1}$. The ML estimate of $\boldsymbol{\gamma}_\nu$ may then easily be obtained as

$$\widehat{\boldsymbol{\gamma}}_\nu^{\mathrm{ML}} = \left(\mathbf{V}_{n-k,\nu}^{\mathrm{H}}\mathbf{V}_{n-k,\nu}\right)^{-1}\mathbf{V}_{n-k,\nu}.\mathbf{s}_{n-k}\left(\mathbf{r}_n\right),$$

whereas its MAP estimate is

$$\widehat{\boldsymbol{\gamma}}_\nu^{\mathrm{MAP}} = n\sigma_{\mathrm{g}}^{-2}\left(n\sigma_{\mathrm{g}}^{-2}\mathbf{V}_{n-k,\nu}^{\mathrm{H}}\mathbf{V}_{n-k,\nu} + \sigma_{\mathrm{i}}^{-2}\mathbf{I}_\nu\right)^{-1}\mathbf{V}_{n-k,\nu}.\mathbf{s}_{n-k}\left(\mathbf{r}_n\right).$$

4) Complexity issues: The complexity of real BCH encoding is that of two discrete Fourier transforms (DFT), one of dimension k and one of dimension n. Efficient implementations are available, see, *e.g.*, [44]. The syndrome computation may also be performed by DFT or by a matrix-vector product, with complexity $(n-k).n$. The determination of the number of errors requires the computation of the norm of the syndrome, *i.e.*, $(n-k)$ multiplications and additions and at most t comparisons. In the absence of errors, projection decoding is performed; its complexity is similar to that of real BCH encoding. When there are errors, the error localization requires some matrix-matrix and matrix vector products (complexity $(n-k-\nu)^2\nu$) and a $\nu \times \nu$ matrix has to be inverted (complexity ν^3). Note that ν always takes very small values in the context of interest. The error localisation polynomial has then to be evaluated, which may be done by n DFT of dimension n. The ML or MAP estimation are of the same complexity as the error localization. None of these operations requires thus a high computational complexity and implementing a real BCH code is of the same order of complexity as a finite-field BCH code. Nevertheless, for real BCH codes, floating-point operations have to be performed.

D. Comparison with the modified BM algorithm

As seen is Section III-C, the three parameters of the GBG channel model are instrumental for the accurate estimation of the characteristics if the impulse errors. Before applying the real BCH codes in a realistic satellite image transmission scheme, the proposed decoding method is compared to the modified BM algorithm proposed in [33], since both have similar computational complexity.

To perform the comparison, information vectors $\mathbf{x}_k$ are randomly constructed from a uniform source between 0 and 10. These vectors are then BCH encoded to get $\mathbf{c}_n$. A mixture of gaussian noise with variance σ_{g}^2 ranging from 10^{-2} to 1 and BG noise with $\sigma_{\mathrm{i}}^2 = 100$ and $p_{\mathrm{i}} = 10^{-2}$ is added to $\mathbf{c}_n$.

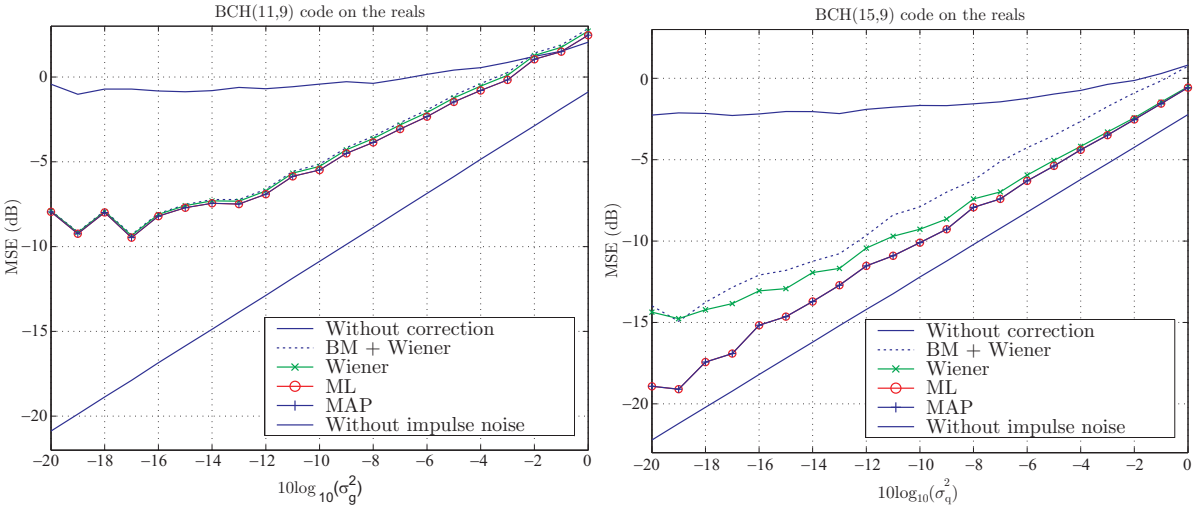

Fig. 7. Comparison of several error-correcting schemes for real BCH(11,9) and real BCH(15,9)

The mean squared error (MSE) between $\mathbf{x}_k$ and its estimate after decoding $\widehat{\mathbf{x}}_k$ are compared using several decoding methods. The two reference schemes are that involving a projection onto the code to perform the correction, see Section III-B, and that without BG noise. The modified BM (to determine the number and locations of impulse errors) combined with the Wiener estimator (to estimate their amplitude) of [33] (BM+Wiener) is compared to the localisation method proposed in this paper, combined with a ML, a MAP, and the Wiener estimator of [33].

Figure 7 shows the results obtained for BCH$(11, 9)$ and BCH$(15, 9)$ codes. For the shorter code, all techniques perform similarly. For the code introducing more redundancy, the proposed techniques outperform the modified BM algorithm. The proposed localisation technique is more efficient, as illustrated by its better performance when it is combined with the Wiener estimator. The ML or MAP estimation technique are also better suited, as shown by the performance increase when compared to the Wiener estimator. In the presented simulations the ML and MAP estimators perform similarly.

For what concerns the block length, using long BCH codes in finite fields poses only the problem of computational complexity. Long codes are more efficient than short codes. To evaluate the influence

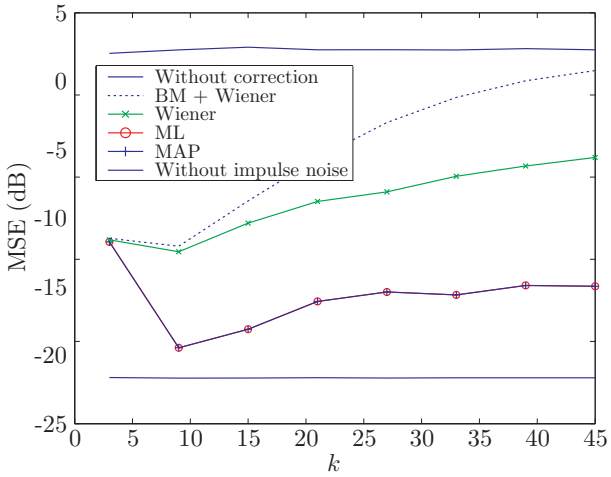

Fig. 8. Influence of the code length on the MSE at decoder output, codes with constant $k/n = 0.6$ have been considered

of the code length, a second set of simulations has been considered, where σ_{g}^2 has been fixed at 10^{-2} and various real BCH codes with the same redundancy are compared, namely BCH$(5, 3)$, BCH$(15, 9)$, BCH$(25, 15)$ . . . BCH$(75, 45)$. Figure 8 provides the MSE as a function of k for the same correction techniques as those used in Figure 7. The same conclusions as before may be taken. Moreover, with real BCH codes, using long codes is not recommended. Without background noise, longer real BCH codes would perform better. However, the influence of the background noise increases with the size of the code. From (23) , one may verify that the impulse noise to background noise ratio is $(\nu\sigma_{\mathrm{i}}^2)/(n\sigma_{\mathrm{g}}^2)$. When n increases, it becomes more difficult to distinguish the impulse noise from the background noise. For real BCH codes, a trade-off has thus to be found between code length and sensitivity to background noise.

IV. Application to robust image transmission

This section shows how real BCH codes may be used in image transmission schemes. For that purpose, we first consider the classical TSCC scheme represented in Figure 9. After a pyramid wavelet decomposition, the subbands are quantized and entropy-coded before adding some redundancy with some channel coder. The inverse operations are realized at receiver side.

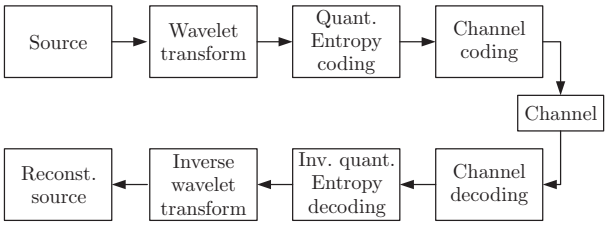

Fig. 9. Reference tandem coding scheme

An extension of this scheme to a possible JSCC one is represented in Figure 10. It is still quite close to a tandem scheme. After a pyramid wavelet decomposition, the subbands are encoded using

real BCH codes. Each encoded subband is quantized and protected with a channel code. At receiver side, a channel decoding is first realized before inverse quantization. A new correction is performed using real BCH decoding. The equivalent GBG channel model that will be used for the real BCH coder/decoder is formed by the quantizer, the channel coder, the physical channel, the channel decoder and the inverse quantizer.

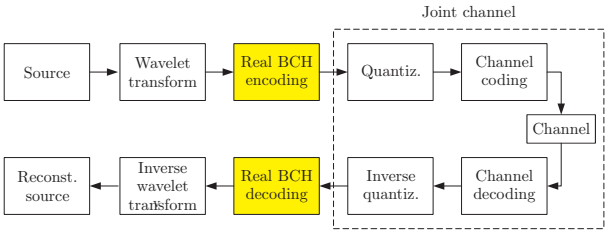

Fig. 10. Joint source-channel coding scheme

Note that in the source coder of the JSCC scheme, no entropy coding is performed. This is a key point to get a valid GBG model: in presence of entropy-coded data, an error resisting to the channel decoding would result in a desynchronisation of any standard entropy decoder that would result in many samples in error. Obtaining an efficient GBG in such cases would be very difficult.

A. Real BCH encoding

Each subband of the pyramid is encoded with a real BCH code. This encoding can be realized on lines or columns of each subband. A more efficient technique is to realize a *product* real BCH encoding by encoding the lines first and then the columns of the resulting encoded subband.

The real BCH error correction technique presented in Section III-C may then be applied alternatively on the lines and on the columns of the received subbands. Errors that may not be corrected by the BCH decoder on the lines may be corrected by the BCH decoder on the columns. For a given amount of redundancy, this technique using a real BCH product code shows a significant improvement when compared to a single real BCH code.

B. Optimization issue

On the proposed JSCC scheme of Figure 10, each subband of the wavelet pyramid can be encoded independently. The number of bits per pixel, the quantization type (SQ, PVQ), the dimension of the PVQ, when used, the length and redundancy of real BCH codes for lines and columns have to be chosen in order to get the best performances for a given channel. All these parameters may be tuned using the Shoham-Gersho algorithm [45].

V. EXAMPLES

Two examples are considered in this section. Both are based on the transmission scheme represented in Figure 9 for the TSCC and in Figure 10 for the JSCC. The first example is quite reminiscent of satellite image transmission, in the 3S context [46]. This context explains the large bit rates: satellite images must be of very good quality, and the target compression ratio is quite small, especially if one compares to the global rate, including the influence of the channel coders that are necessary to secure the transmission.

A. First example

Gray-scale source images, such as that represented in Figure 11, are represented with $\tau_{\mathrm{p}} = 8$ bpp and are processed by a three-level dyadic wavelet transform, resulting in 10 subband images.

Fig. 11. Reference gray-scaled 8bpp picture for the first example

In the TSCC scheme, all these subbands are quantized and entropy coded to get a source coder output bitrate $\tau_{\mathrm{s}} = 2$ bpp. The quantizer efficiency is adjusted using a regulation loop, for more details, see [47]. Two strong channel coders are involved, namely a BCH$(127, 78)$ and a BCH$(63, 39)$ to get a channel bitrate of approximatively $\tau_{\mathrm{c}} = 3.2$ bpp. The compressed data are sent over a BSC channel with transition probability $10^{-5} \leqslant \varepsilon \leqslant 10^{-1}$. The performances of these reference schemes are represented in Figure 12. The global rate $(3.2/8)$ and the transition probabilities were taken as being typical of the 3S context [46].

In the JSCC scheme, all subbands are first encoded using a real BCH$(51, 31)$ code only on the lines or a real BCH$(19, 15)$ product code on both lines and columns. The encoded subbands are quantized using a Lloyd-Max quantizer [48] to get a source coder output rate $\tau'_{\mathrm{s}} = 3.2$ bpp. The bit assignment in the subbands is realized by optimization using the Shoham-Gersho algorithm. No channel code is used in this case, thus $\tau'_{\mathrm{c}} = \tau'_{\mathrm{s}} = 3.2$ bpp. Both TSCC and JSCC introduce thus the same amount of redundancy and work with the same channel conditions. The performances of these JSCC schemes are also represented in Figure 12.

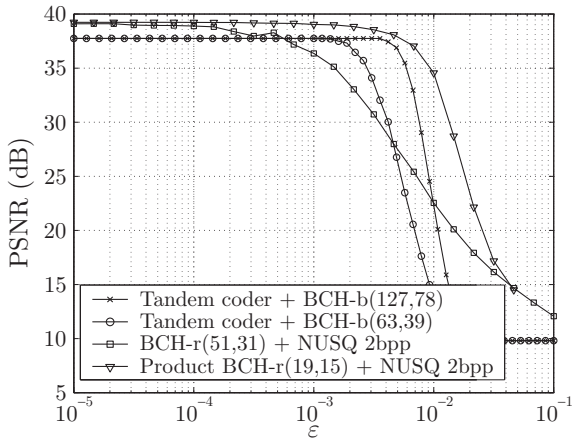

Fig. 12. Performance comparison between the TS-CC and the JS-CC schemes for the first example

The JSCC using the real BCH product code is much more efficient than the JSCC using the real BCH code on the lines, as the code concatenation benefits from the alternating correction on the lines and on the columns. Moreover, this scheme outperforms the TSCC scheme in all channel situations.

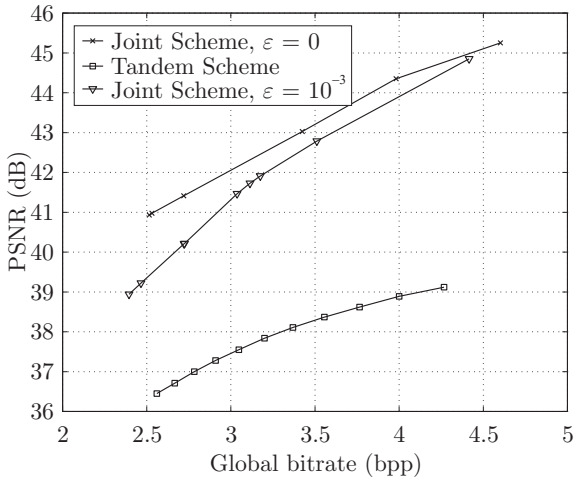

Fig. 13. Performance improvements obtained on the JS-CC scheme using a joint optimisation of the real BCH codes and of the quantizers (first example)

In the previous scheme, the same real BCH$(19, 15)$ code is used in all subbands. However, some subbands are more sensitive to noise than others. For a fixed channel parameter ε, the performance of the JSCC scheme can be further improved by realizing a joint optimization of the real BCH codes *and* of the quantizers used in each subband using a Shoham-Gersho algorithm. This amounts to attributing to each subband the optimal amount of redundancy. The performances that may then be obtained are represented in Figure 13. When compared to the unoptimized case, at 3.2 bpp, an improvement of about 3 dB is obtained on a noise-free channel and of about 2.5 dB on a BSC channel with transition probability $\varepsilon = 10^{-3}$. As a conclusion, in this application, a joint source-channel coder would be more efficient than the tandem one for all considered channel conditions.

B. Second example

In this second example, the numbers are taken from the PLEIADES satellite image transmission scenario [49]. The source and channel codes of the reference scheme have been carefully optimized with respect to given channel conditions, which involves a much more efficient transmission system.

The original gray-scaled images are now initially represented on 10 bpp. They are first processed by the same three-level dyadic wavelet transform to provide again 10 transformed subbands.

In the TSCC scheme, all these subbands are quantized and entropy coded to get a source coder output bitrate $\tau_{\mathrm{s}} = 2$ bpp. The quantizer efficiency is still adjusted using a regulation loop. In this scheme, a Reed-Solomon RS$(255, 238)$ coder is followed by a trellis-coded modulation using an 8PSK and a rate $2.5/3$ convolutional code. The overall channel coding rate is $R_{\mathrm{cc}} = \frac{238}{255}\frac{2.5}{3} = 0.77$. The bitrate on the channel is thus about $\tau_{\mathrm{c}} = 2.57$ bpp. The compressed data are sent over an AWGN channel characterized by its E_{b}/N_0 ratio. Table I summarizes some residual bit error rates (BER) at the output of the RS$(255, 238)$ decoder for various values of E_{b}/N_0 in the region of interest. Errors

Channel coder and modulation	E_{b}/N_0	BER
RS(255,238) - TCM 2.5/3	6.5	2.58×10^{-4}
	6.8	7.65×10^{-7}
	6.95	1.25×10^{-8}

TABLE I

RESIDUAL BER FOR VARIOUS SIGNAL TO NOISE RATIOS FOR THE PLÉÏADES CHANNEL

are irregularly spread in the bitstream. On a noise-free channel, the PSNR is 46.7 dB. For the noisy channel, the performances of the reference scheme are represented in Figure 14.

In the JSCC scheme, all subbands are again first encoded using real BCH product codes along both lines and columns. The encoded subbands are quantized using Lloyd-Max or pyramid vector quantizers to get a source coder output rate $\tau'_{\mathrm{s}} = 2$ bpp. No entropy coding is realized. The bit assignment in the subbands, the real BCH code used for lines and columns and the quantization type are optimized using the Shoham-Gersho algorithm. The same channel code as for the TSCC is used, thus the bitrate on the channel is $\tau'_{\mathrm{c}} = 2.57$ bpp, equal to that for the TSCC scheme. Keeping the channel coder/decoder pair is necessary in this case, as an 8PSK with $E_{\mathrm{b}}/N_0 = 6.8$ dB would suffer at receiver side from a BER of more than 2.10^{-2}, which is not compatible with an efficient real BCH decoding. The performance of the JSCC scheme is represented in Figure 14.

For the JSCC scheme, two optimization situations have been considered. For the first, a noise-free channel is assumed, the optimization process results in a real BCH encoding of the only low frequency subband. In a second, a channel with E_{b}/N_0 around 6.4 dB is considered and all subbands have been encoded using real BCH codes. Table II summarizes the characteristics (source coder efficiency C_{R}, redundancy introduced by the BCH codes R_{BCH}, PSNR obtained without noise) of these schemes.

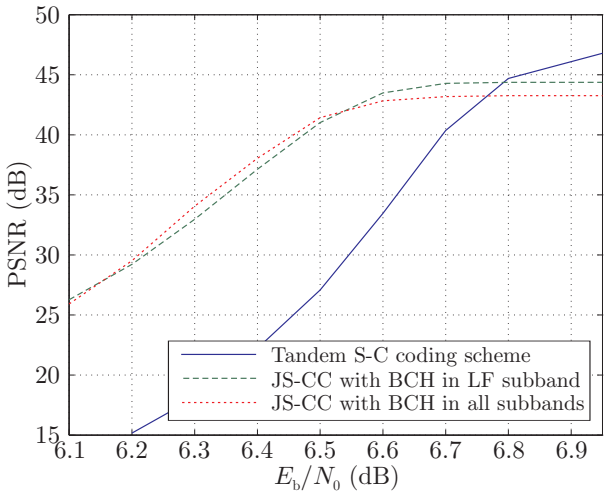

Fig. 14. Compared performances of the tandem and joint source-channel coding schemes (second example)

	C_R	τ_{BCH}	PSNR
BCH code on the LF subband	5.01	1.018	44.3 dB
BCH code of all subbands	5.96	1.2	42.8 dB

TABLE II

CHARACTERISTICS AND PERFORMANCES OF THE JOINT SCHEMES FOR NOISE-FREE CHANNELS

The reduced PSNR in the clear-channel case is due to the reduced number of bits assigned for the quantization in the JSCC in order to allow for the introduction of redundancy with real BCH codes.

In this scenario, where the TSCC has been accurately tuned, the results are less contrasted. The TSCC scheme performs better (about 2.5dB) on a channel with little noise. The JSCC scheme performance decreases slowlier when the channel noise increases. For example, at $E_b/N_0 = 6.5$ dB, the JSCC scheme is more than 10 dB better than the TSCC scheme. The JSCC provides more robust results, at the cost of a slightly reduced nominal performance (the one which is obtained without errors).

Note that in all cases, our JSCC were compared to optimized situations, and that further improvements would still be feasible, introducing, *e.g.*, other joint source-channel decoding methods such as those presented in [21], [22].

Obviously, one could think of more efficient tandem schemes, *e.g.*, based on LDPC. Note, however, that the reference scheme considered here is the actual PLEIADES system, which was also carefully tuned. From our experience, a comparison with a tandem situation involving LDPC codes would certainly result in the same conclusion: more robustness is obtained at the cost of a reduced nominal performance. Again, JSCC intends to cope with highly varying channel conditions in a more robust way.

VI. Conclusions and perspectives

In this paper, a new still image coding scheme is presented. Contrary to standard tandem coding schemes, where the redundancy is introduced after source coding, it is introduced before source coding using real BCH codes. A joint channel model is presented and a comparison with a state-of-art real BCH decoder is given. Decoding algorithms are derived from this channel model. A realistic comparison between a TSCC and a JSCC scheme for the robust transmission of still images has been presented. When the TSCC is tuned for a wide range of channel conditions, the JSCC outperforms the TSCC in all situations. When the TSCC is tuned for precise channel condition, the JSCC shows an increased robustness when compared to TSCC as the channel worsens.

More generally, when the channel conditions are rather precisely known and are not fluctuating, a TSCC scheme designed according to Shannon's separation principle is the solution to be adopted. However, when the transmission condition are fluctuating, JSCC may increase the robustness of the scheme without significant loss at the nominal point of operation of the TSCC.

Further investigations may be conducted in four directions. First, the error-correction efficiency is related to the ratio between the variance of the impulse noise on the variance of the background noise. Standard optimized index assignment results in a reduction of the variance of the impulse noise. It may be interesting to optimize the index assignment in order to increase this variance. This would clearly allow the correction algorithms to be more efficient. Second, in the considered examples, no entropy coder was considered in the JSCC schemes, in order to be able to build an efficient GBG channel model. Using more sophisticated variable-length decoder such as that presented in [18], in combination with interleavers may reduce the number of consecutive errors resulting from wrong variable-length decoding and spread these errors, allowing thus a GBG channel model to be deduced. A third idea is to try to obtain soft outputs of the BCH decoder in order to put it in an iterative scheme. First promising results have been presented in [50], [51]. Finally, on the presented scheme, wavelet transform and real BCH encoding are done separately. Decorrelation and redundancy introduction may be done in a single step using oversampled filterbanks. First results have been obtained in [52].

Ackownledgements

This work was partly supported by CNES grant, by the RNRT COSOCATI project and by the Network of Excellence NEWCOM. The authors would like to thank Dr C. Lambert-Nebout and Dr G. Lesthievent from CNES, and C. Guillaume and R. Triki for their help during this project.

Appendix I

GBG channel model for a pyramid vector quantizer followed by a BSC

Consider a source that is quantized using a Pyramid Vector Quantizer (PVQ) of dimension L at a bitrate b per dimension [53]. A PVQ uses a dictionary formed by vectors $\mathbf{d}_k = (d_{k,1}, \ldots, d_{k,L})^{\mathrm{T}}$

placed on an L-dimensional pyramid of radius K such that

$$d_{k,i} \in \mathbb{Z},\ i = 1, \ldots L \text{ and } \sum_{i=1}^{L} |d_{i,k}| = K. \tag{24}$$

For given values of L and K, there are $N(L,K)$ vectors $\mathbf{d}_k$ satisfying (24). Thus, K has to be chosen as the integer which maximizes the function $N(L,K)$ under the constraint $N(L,K) \leqslant 2^{Lb}$. For the quantization, an L-dimensional source vector $\mathbf{s} = (s_1, \ldots, s_L)^{\mathrm{T}}$ is first multiplied by $K\lambda/L$, where $\lambda = 1/E(|s|)$, to get closer in average to the points of the pyramid. Then, the PVQ assigns the closest point $\mathbf{d}$ on the pyramid to $\mathbf{s}$. An index vector of Lb bits $\mathbf{i} = I(\mathbf{d})$ is finally determined using the index assignment function $I(\cdot)$ described in [53].

The binary index vector is transmitted through a BSC with transition probability ε. The received index is used to build an estimate $\widehat{\mathbf{d}}$ of $\mathbf{d}$**.** The reconstructed vector $\widehat{\mathbf{s}}$ results from the rescaling of $\widehat{\mathbf{d}}$ by $\gamma L/(K\lambda)$, where γ is some coefficient tuned to reduce the reconstruction error.

In order to build a GBG model for this communication scheme, the quantization noise and the noise resulting from channel errors have to be considered. As in the other examples, this model considers a JSCC channel with scalar input and output.

In [53], an approximate expression of the distortion per dimension $D_{\mathrm{PVQ}}(b)$ introduced by the quantization using a PVQ is provided; it is used as an estimate $\widehat{\sigma}_{\mathrm{g}}^2$ for σ_{g}^2

$$\widehat{\sigma}_{\mathrm{g}}^2 = D_{\mathrm{PVQ}}(b) = \frac{e^2}{3\lambda^2} 2^{-2b}. \tag{25}$$

For a given L-dimensional pyramid of radius K, with a BSC of bit transition probability ε, the probability that an index of Lb bits is in error is thus

$$p = 1 - (1-\varepsilon)^{Lb}. \tag{26}$$

The index transition probability becomes

$$p(\mathbf{i}_k|\mathbf{i}_j) = \varepsilon^{d_{\mathrm{H}}(\mathbf{i}_j,\mathbf{i}_k)} (1-\varepsilon)^{Lb - d_{\mathrm{H}}(\mathbf{i}_j,\mathbf{i}_k)}.$$

An index error may affect from one up to L samples in the associated quantized vector. When the index $\mathbf{i}_j$ becomes $\mathbf{i}_k$ at the BSC output, the number of affected samples is $d_{\mathrm{H}}(\mathbf{d}_j, \mathbf{d}_k)$. Thus, the average number of samples affected by index errors is

$$\alpha = \sum_{j=1}^{N(L,K)} p(\mathbf{d}_j) \sum_{k=1, k\neq j}^{N(L,K)} \varepsilon^{d_{\mathrm{H}}(\mathbf{i}_j,\mathbf{i}_k)} (1-\varepsilon)^{Lb - d_{\mathrm{H}}(\mathbf{i}_j,\mathbf{i}_k)} d_{\mathrm{H}}(\mathbf{d}_j, \mathbf{d}_k). \tag{27}$$

By combining (26) with (27), one estimates p_{i} as

$$\widehat{p_{\mathrm{i}}} = \alpha \left(1 - (1-\varepsilon)^{Lb}\right)/L. \tag{28}$$

The evaluation of (27) may be time consuming for pyramids with large $N(L,K)$, as it requires $O\left((N(L,K))^2\right)$ operations. Under the assumption that $P(\mathbf{d}_k) = 1/N(L,K)$ (this is reasonable for

a Laplacian source) and that the channel is not too noisy, which implies $p\left(\mathbf{i}_k|\mathbf{i}_j\right) \approx \varepsilon\left(1-\varepsilon\right)^{Lb-1}$, a good approximation for α is

$$\widetilde{\alpha} = \frac{1}{N\left(L,K\right)} \sum_{j=1}^{N(L,K)} \sum_{k=1,\, d_{\mathrm{H}}(\mathbf{i}_j,\mathbf{i}_k)=1}^{N(L,K)} \varepsilon\left(1-\varepsilon\right)^{Lb-1} d_{\mathrm{H}}\left(\mathbf{d}_j,\mathbf{d}_k\right), \tag{29}$$

which requires only $O\left(LbN\left(L,K\right)\right)$ operations.

To evaluate the distortion resulting from index errors, one may use the mean value of the squared distance between any two points on the pyramid, which is about $2K^2/\left(L+1\right)$ [54]. However, this evaluation does not take into account that errors affecting a single bit in an index are the most likely. When the index $\mathbf{i}_j$ becomes $\mathbf{i}_k$ at the BSC output, the number of affected samples is $d_{\mathrm{H}}\left(\mathbf{d}_j,\mathbf{d}_k\right)$ and the average distortion per sample is $\left|\mathbf{d}_j-\mathbf{d}_k\right|^2/d_{\mathrm{H}}\left(\mathbf{d}_j,\mathbf{d}_k\right)$. Thus, the per-component distortion may be evaluated as follows

$$D_{\mathrm{i}} = \frac{\gamma L}{K\lambda} \sum_{j=1}^{N(L,K)} p\left(\mathbf{d}_j\right) \sum_{k=1,k\neq j}^{N(L,K)} \varepsilon^{d_{\mathrm{H}}(\mathbf{i}_j,\mathbf{i}_k)} \left(1-\varepsilon\right)^{Lb-d_{\mathrm{H}}(\mathbf{i}_j,\mathbf{i}_k)} \frac{\left|\mathbf{d}_j-\mathbf{d}_k\right|^2}{d_{\mathrm{H}}\left(\mathbf{d}_j,\mathbf{d}_k\right)}, \tag{30}$$

where $\frac{\gamma L}{K\lambda}$ accounts for the scaling. From (8), (28) and (30), one may estimate the variance of the Gaussian noise in the BG process as

$$\widehat{\sigma}_{\mathrm{i}}^2 = \frac{LD_{\mathrm{i}}}{\alpha\left(1-\left(1-\varepsilon\right)^{Lb}\right)}.$$

Again evaluating (30) is time consuming. Under the same hypotheses that those used to compute (29), the following approximation of (30) may be obtained

$$\widetilde{D}_{\mathrm{i}} = \frac{1}{N\left(L,K\right)} \frac{\gamma L}{K\lambda} \sum_{j=1}^{N(L,K)} \sum_{k=1,d_{\mathrm{H}}(\mathbf{i}_j,\mathbf{i}_k)=1}^{N(L,K)} \varepsilon\left(1-\varepsilon\right)^{Lb-1} \frac{\left|\mathbf{d}_j-\mathbf{d}_k\right|^2}{d_{\mathrm{H}}\left(\mathbf{d}_j,\mathbf{d}_k\right)}. \tag{31}$$

To evaluate the accuracy of the obtained model, we consider a zero-mean unit variance Gaussian source quantized with PVQ of dimension $L=8$. The binary indexes are sent through a BSC with transition probability $\varepsilon = 0.01$. Figure 15 shows the parameters of an actual channel and those of its GBG model. For the evaluation of $\widehat{\sigma}_{\mathrm{i}}^2$, the two expressions resulting from (30) (full GBG model) and (31) (simplified GBG model) have been used.

A good estimate $\widehat{p}_{\mathrm{i}}$ of the impulse error probability is provided. The estimate of σ_{i}^2 provided by (30) is also satisfying and still reasonable with (31). The most difficult quantity to estimate is σ_{g}^2. Nevertheless, this quantity is a characteristic of the system, and may be estimated more accurately by simulation beforehand.

APPENDIX II

PGZ ALGORITHM ACCOUNTING FOR THE BACKGROUND NOISE

Consider $\mathbf{W}_n^{-1}\boldsymbol{\gamma}_n = \left(\Gamma_0,\Gamma_1,\ldots,\Gamma_{n-1}\right)^{\mathrm{T}}$ and $\mathbf{W}_n^{-1}\mathbf{b}_n = \left(B_0,B_1,\ldots,B_{n-1}\right)^{\mathrm{T}}$, then the syndrome (20) may be rewritten as

$$\mathbf{s}_{n-k}\left(\mathbf{r}_n\right) = \left(S_{a_0},\ldots,S_{a_0+n-k-1}\right)^{\mathrm{T}} = \left(\Gamma_{a_0},\ldots,\Gamma_{a_0+n-k-1}\right)^{\mathrm{T}} + \left(B_{a_0},\ldots,B_{a_0+n-k-1}\right)^{\mathrm{T}}.$$

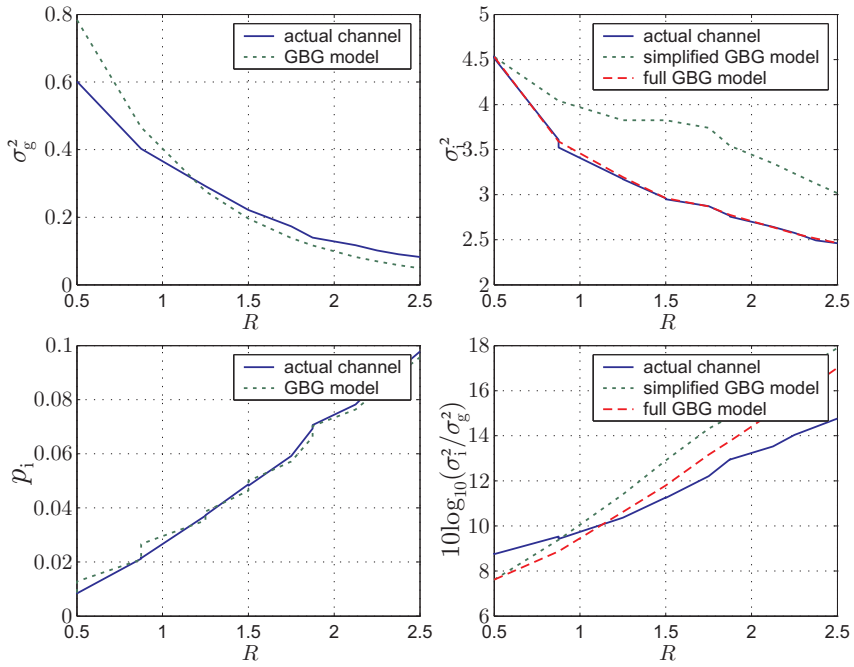

Fig. 15. Comparison of a joint channel and its GBG model ($L = 8$ dimension PVQ followed by a BSC)

Assuming that there is no background gaussian noise, the classical PGZ algorithm would show that $\boldsymbol{\lambda}_\nu$ has to satisfy

$$\boldsymbol{\Gamma}^*_{n-k-\nu,\nu}.\boldsymbol{\lambda}_\nu = -\boldsymbol{\Gamma}^*_{n-k-\nu,1} \tag{32}$$

with $\boldsymbol{\Gamma}_{n-k-\nu,1} = \begin{pmatrix} \Gamma_{a_0+\nu} & \dots & \Gamma_{a_0+n-k-1} \end{pmatrix}^{\mathrm{T}}$ and

$$\boldsymbol{\Gamma}_{n-k-\nu,\nu} = \begin{pmatrix} \Gamma_{a_0+\nu-1} & \cdots & \Gamma_{a_0} \\ \vdots & \ddots & \vdots \\ \Gamma_{a_0+n-k-2} & \cdots & \Gamma_{a_0+n-k-1-\nu} \end{pmatrix}.$$

Assuming that $\boldsymbol{\Gamma}_{n-k-\nu,\nu}$ and $\boldsymbol{\Gamma}_{n-k-\nu,1}$ are available, (32) may be solved in a least-squares sense for $\boldsymbol{\lambda}_\nu$

$$\widehat{\boldsymbol{\lambda}}_\nu = -\left(\boldsymbol{\Gamma}^T_{n-k-\nu,\nu}.\boldsymbol{\Gamma}^*_{n-k-\nu,\nu}\right)^{-1} \boldsymbol{\Gamma}^T_{n-k-\nu,\nu}.\boldsymbol{\Gamma}^*_{n-k-\nu,1}. \tag{33}$$

In what follows, it is shown that $\boldsymbol{\Gamma}^T_{n-k-\nu,\nu}.\boldsymbol{\Gamma}^*_{n-k-\nu,\nu}$, $\boldsymbol{\Gamma}_{n-k-\nu,\nu}$ and $\boldsymbol{\Gamma}_{n-k-\nu,1}$ may be estimated from the syndrome (20).

Let $\mathbf{B}_{n-k-\nu,1} = \begin{pmatrix} B_{a_0+\nu} & \dots & B_{a_0+n-k-1} \end{pmatrix}$ and

$$\mathbf{B}_{n-k-\nu,\nu} = \begin{pmatrix} B_{a_0+\nu-1} & \cdots & B_{a_0} \\ \vdots & \ddots & \vdots \\ B_{a_0+n-k-2} & \cdots & B_{a_0+n-k-1-\nu} \end{pmatrix}$$

using the syndrome (20), $\mathbf{S}_{n-k-\nu,1}$ may be written as

$$\mathbf{S}_{n-k-\nu,1} = \boldsymbol{\Gamma}_{n-k-\nu,1} + \mathbf{B}_{n-k-\nu,1} \tag{34}$$

and $\mathbf{S}_{n-k-\nu,\nu}$ as

$$\mathbf{S}_{n-k-\nu,\nu} = \mathbf{\Gamma}_{n-k-\nu,\nu} + \mathbf{B}_{n-k-\nu,\nu}. \tag{35}$$

Then

$$\begin{aligned}\mathbf{S}^{\mathrm{H}}_{n-k-\nu,\nu}\mathbf{S}_{n-k-\nu,\nu} &= \mathbf{\Gamma}^{\mathrm{H}}_{n-k-\nu,\nu}\mathbf{\Gamma}_{n-k-\nu,\nu} + \mathbf{\Gamma}^{\mathrm{H}}_{n-k-\nu,\nu}\mathbf{B}_{n-k-\nu,\nu} \\ &\quad + \mathbf{B}^{\mathrm{H}}_{n-k-\nu,\nu}\mathbf{\Gamma}_{n-k-\nu,\nu} + \mathbf{B}^{\mathrm{H}}_{n-k-\nu,\nu}\mathbf{B}_{n-k-\nu,\nu}. \end{aligned}\tag{36}$$

Impulse and Gaussian background noise are uncorrelated, thus $E\left(\mathbf{\Gamma}^{\mathrm{H}}_{n-k-\nu,\nu}\mathbf{B}_{n-k-\nu,\nu}\right) = \mathbf{0}_\nu$ and $E\left(\mathbf{B}^{\mathrm{H}}_{n-k-\nu,\nu}\mathbf{\Gamma}_{n-k-\nu,\nu}\right) = \mathbf{0}_\nu$. Moreover, one may easily verify that $E\left(\mathbf{B}_{n-k-\nu,1}\right) = \mathbf{0}_{n-k-\nu,1}$, $E\left(\mathbf{B}_{n-k-\nu,\nu}\right) = \mathbf{0}_{n-k-\nu,\nu}$, and $E\left(\mathbf{B}^{\mathrm{H}}_{n-k-\nu,\nu}\mathbf{B}_{n-k-\nu,\nu}\right) = \frac{n-k-\nu}{n}\sigma^2_{\mathrm{g}}\mathbf{I}_\nu$.

ML estimates for $\mathbf{\Gamma}_{n-k-\nu,\nu}$ and $\mathbf{\Gamma}_{n-k-\nu,1}$ may then be obtained as $\hat{\mathbf{\Gamma}}_{n-k-\nu,\nu} = \mathbf{S}_{n-k-\nu,\nu}$ and $\hat{\mathbf{\Gamma}}_{n-k-\nu,1} = \mathbf{S}_{n-k-\nu,1}$. Using (36) and the two previous estimates, one may deduce an estimate for $\mathbf{\Gamma}^H_{n-k-\nu,\nu}.\mathbf{\Gamma}_{n-k-\nu,\nu}$, as

$$\widehat{\mathbf{\Gamma}^H_{n-k-\nu,\nu}.\mathbf{\Gamma}}_{n-k-\nu,\nu} = \mathbf{S}^H_{n-k-\nu,\nu}\mathbf{S}_{n-k-\nu,\nu} - \frac{(n-k-\nu)}{n}\sigma^2_{\mathrm{g}}\mathbf{I}_\nu.$$

The three previous estimates may then be used in (33) to get (22).

APPENDIX III
PDF OF THE NORM OF THE SYNDROME

The aim of this appendix is to derive a probability distribution of the norm of the syndrome under the assumption that ν errors have corrupted the transmitted codeword.

Considering the vector of impulses $\boldsymbol{\gamma}\left(\boldsymbol{\rho}^{(\nu)}\right) = \left(\gamma_{\rho_1^{(\nu)}}, \ldots, \gamma_{\rho_\nu^{(\nu)}}\right)^T$, (20) may be rewritten as

$$\mathbf{s}_{n-k}\left(\boldsymbol{\rho}^{(\nu)}\right) = \mathbf{V}_{n-k,\nu}\left(\boldsymbol{\rho}^{(\nu)}\right)\boldsymbol{\gamma}\left(\boldsymbol{\rho}^{(\nu)}\right) + \mathbf{R}_{n-k,n}\mathbf{W}_n^{-1}\mathbf{b}_n, \tag{37}$$

with $\mathbf{V}_{n-k,\nu}\left(\boldsymbol{\rho}^{(\nu)}\right) = \mathbf{R}_{n-k,n}\left(\begin{array}{ccc}\mathbf{w}^{\mathrm{T}}_{\widehat{\rho}_1} & \ldots & \mathbf{w}^{\mathrm{T}}_{\widehat{\rho}_\nu}\end{array}\right)^{\mathrm{T}}$, see Section III-C.3. Then,

$$\mathbf{s}_{n-k}\left(\boldsymbol{\rho}^{(\nu)}\right) = \left(\begin{array}{cc}\mathbf{V}_{n-k,\nu}\left(\boldsymbol{\rho}^{(\nu)}\right) & \mathbf{R}_{n-k,n}\mathbf{W}_n^{-1}\end{array}\right)\left(\begin{array}{c}\boldsymbol{\gamma}\left(\boldsymbol{\rho}^{(\nu)}\right)\\ \mathbf{b}_n\end{array}\right)$$

is a linear combination of gaussian variables with differing variances. Consider

$$\mathbf{\Delta}_{\nu+n}(\nu) = \mathrm{diag}\left(\underbrace{1/\sigma_{\mathrm{i}}, \ldots, 1/\sigma_{\mathrm{i}}}_{\nu \text{ times}}, \underbrace{1/\sigma_{\mathrm{g}}, \ldots, 1/\sigma_{\mathrm{g}}}_{n \text{ times}}\right),$$

then $\mathbf{s}_{n-k}\left(\boldsymbol{\rho}^{(\nu)}\right) = \mathbf{Q}\left(\boldsymbol{\rho}^{(\nu)}\right)\mathbf{z}_{\nu+n}$, with

$$\mathbf{Q}\left(\boldsymbol{\rho}^{(\nu)}\right) = \left(\begin{array}{cc}\mathbf{V}_{n-k,\nu}\left(\boldsymbol{\rho}^{(\nu)}\right) & \mathbf{R}_{n-k,n}\mathbf{W}_n^{-1}\end{array}\right)\mathbf{\Delta}^{-1}_{\nu+n}(\nu)$$

and

$$\mathbf{z}_{\nu+n} = \mathbf{\Delta}_{\nu+n}(\nu)\left(\begin{array}{c}\boldsymbol{\gamma}\left(\boldsymbol{\rho}^{(\nu)}\right)\\ \mathbf{b}_n\end{array}\right)$$

is a gaussian vector with iid unit variance entries.

The matrix $\mathbf{Q}^{\mathrm{H}}\left(\boldsymbol{\rho}^{(\nu)}\right)\mathbf{Q}\left(\boldsymbol{\rho}^{(\nu)}\right)$ is hermitian symmetric and thus may be diagonalized as

$$\mathbf{Q}^{\mathrm{H}}\left(\boldsymbol{\rho}^{(\nu)}\right)\mathbf{Q}\left(\boldsymbol{\rho}^{(\nu)}\right)=\mathbf{K}^{\mathrm{H}}\left(\boldsymbol{\rho}^{(\nu)}\right)\mathbf{G}\left(\boldsymbol{\rho}^{(\nu)}\right)\mathbf{K}\left(\boldsymbol{\rho}^{(\nu)}\right),$$

with $\mathbf{K}\left(\boldsymbol{\rho}^{(\nu)}\right)$ unitary and real and $\mathbf{G}\left(\boldsymbol{\rho}^{(\nu)}\right)$ diagonal. The norm of the syndrome may then be written as

$$y\left(\boldsymbol{\rho}^{(\nu)}\right)=\mathbf{s}_{n-k}^{\mathrm{H}}\left(\boldsymbol{\rho}^{(\nu)}\right)\mathbf{s}_{n-k}\left(\boldsymbol{\rho}^{(\nu)}\right)=\left(\mathbf{K}\left(\boldsymbol{\rho}^{(\nu)}\right)\mathbf{z}_{\nu+n}\right)^{\mathrm{H}}\mathbf{G}\left(\boldsymbol{\rho}^{(\nu)}\right)\mathbf{K}\left(\boldsymbol{\rho}^{(\nu)}\right)\mathbf{z}_{\nu+n}.$$

As $\mathbf{z}_{\nu+n}$ is normal with unit variance, $\mathbf{K}\left(\boldsymbol{\rho}^{(\nu)}\right)\mathbf{z}_{\nu+n}$ is also normal with unit variance. Thus,

$$y\left(\boldsymbol{\rho}^{(\nu)}\right)=\sum_{k=1}^{\mathrm{rank}\left(\mathbf{Q}\left(\boldsymbol{\rho}^{(\nu)}\right)\right)}g_k^2\left(\boldsymbol{\rho}^{(\nu)}\right)|z_k|^2,$$

where $g_k\left(\boldsymbol{\rho}^{(\nu)}\right)$ are the diagonal entries of $\mathbf{G}\left(\boldsymbol{\rho}^{(\nu)}\right)$. As $\mathrm{rank}\left(\mathbf{Q}\left(\boldsymbol{\rho}^{(\nu)}\right)\right)=n-k$, the probability distribution of $y\left(\boldsymbol{\rho}^{(\nu)}\right)$ is the convolution of scaled χ^2 distributions with one degree of freedom

$$p\left(y|\boldsymbol{\rho}^{(\nu)},\nu\right)=\bigotimes_{k=1}^{n-k}\chi_1^2\left(g_k\left(\boldsymbol{\rho}^{(\nu)}\right),y\right), \tag{38}$$

with $\chi_1^2\left(g_k,y\right)=\frac{1}{\sqrt{2}g_k^2\Gamma(1/2)}\left(y/g_k^2\right)^{-1/2}\exp\left(-y/\left(2g_k^2\right)\right)$ and may be evaluated using [55].

The pdf of the syndrome under the hypothesis that there are ν impulse errors is obtained by averaging (38) over all possible values of $\boldsymbol{\rho}^{(\nu)}$

$$p\left(y|\nu\right)=\sum_{\boldsymbol{\rho}^{(\nu)}}p\left(y|\boldsymbol{\rho}^{(\nu)},\nu\right)p\left(\boldsymbol{\rho}^{(\nu)}|\nu\right). \tag{39}$$

REFERENCES

[1] A. Gabay, P. Duhamel, and O. Rioul, "Real BCH codes as joint source channel code for satellite images coding," in *Proc. Globecom*, San Francisco, USA, nov 2000, pp. 820–824.

[2] A. Gabay, O. Rioul, and P. Duhamel, "Joint source-channel coding using structured oversampled filters banks applied to image transmission," in *Proc. ICASSP*, 2001, pp. 2581–2584.

[3] F. Abdelkefi, F. Alberge, and P. Duhamel, "A posteriori control of complex Reed Solomon decoding with application to impulse noise cancellation in hiperlan2," in *Proc. ICC*, 2002, pp. 659 – 663.

[4] ——, "Impulse noise correction in hiperlan 2: Improvement of the decoding algorithm and application to PAPR reduction," in *Proceedings of ICC*, 2003, pp. 2094 – 2098.

[5] C. E. Shannon, "A mathematical thoery of communication," *Bell Syst. Tech. J.*, 1948.

[6] S. B. Zahir Azami, P. Duhamel, and O. Rioul, "Joint source channel coding : Panorama of methods," *Proceedings of CNES workshop on Data Compression*, Nov. 1996.

[7] K. Sayood, *Introduction to Data Compression, Second Edition*. San Francisco: Morgan Kaufmann, 2000.

[8] K. Ramchandran, A. Ortega, K. M. Uz, and M. Vetterli, "Multiresolution broadcast for HDTV using joint source/Channel coding," *IEEE Tansactions on SAC*, vol. 11, pp. 3–14, 1993.

[9] J. Hagenauer, "Rate-compatible punctured convolutional codes (RCPC codes) and their applications," *IEEE Trans. Communications*, vol. 36, no. 4, pp. 389–400, 1988.

[10] I. Kozintsev and K. Ramchandran, "Robust image transmission over energy-constrained time-varying channels using multiresolution joint source-channel coding," *IEEE trans. Signal Processing*, vol. 46, no. 4, pp. 1012–1026, 1998.

[11] N. Farvardin and V. Vaishampayan, "Optimal quantizer design for noisy channels: An approach to combined source-channel coding," *IEEE Transactions on Information Theory*, vol. 33, no. 6, pp. 827–837, 1991.

[12] Y. Zhou and W. Y. Chan, "Multiple description quantizer design using a channel optimized quantizer approach," in *Proceedings of the 38th Annual Conference on Information Sciences and Systems (CISS'04)*, 2004.

[13] Y. Takishima, M. Wada, and H. Murakami, "Reversible variable length codes," *IEEE Trans. Commun*, vol. 43, no. 2-4, pp. 158–162, 1995.

[14] K. Lakovic and J. Villasenor, "On design of error-correcting reversible variable length codes," *IEEE Communications Letters*, vol. 6, no. 8, pp. 337– 339, 2002.

[15] J. Wang, L.-L. Yang, and L. Hanzo, "Iterative construction of reversible variable-length codes and variable-length error-correcting codes," *IEEE Communications Letters*, vol. 8, no. 11, pp. 671– 673, 2004.

[16] J. Hagenauer, "Source-controlled channel decoding," *IEEE trans. on Communications*, vol. 43, no. 9, pp. 2449–2457, 1995.

[17] V. Buttigieg and P. Farrell, "A MAP decoding algorithm for variable-length error-correcting codes," in *Codes and Cyphers: Cryptography and Coding IV*. Essex, England: The Inst. of Mathematics and its Appl., 1995, pp. 103–119.

[18] R. Bauer and J. Hagenauer, "Symbol-by-symbol MAP decoding of variable length codes," in *Proc. 3rd ITG Conference Source and Channel Coding*, München, 2000, pp. 111–116.

[19] K. Sayood, H. H. Otu, and N. Demir, "Joint source/channel coding for variable length codes," *IEEE Trans. Commun.*, vol. 48, pp. 787–794, 2000.

[20] R. Thobanen and J. Kliewer, "Robust decoding of variable-length encoded Markov sources using a three-dimensional trellis," *IEEE Communications Letters*, vol. 7, no. 7, pp. 320–322, 2003.

[21] P. Burlina and F. Alajaji, "An error resilient scheme for image transmission over noisy channels with memory," *IEEE trans. on Image Processing*, vol. 7, no. 4, pp. 593–600, 1998.

[22] J. Garcia-Frías and J. D. Villasenor, "Joint turbo decoding and estimation of hidden Markov sources," *IEEE J. Selected Areas in Communications*, vol. 19, no. 9, pp. 1671–1679, 2001.

[23] J. Garcia-Frias and Y. Zhao, "Compression of binary memoryless sources using punctured turbo codes," *IEEE Communications Letters*, vol. 6, no. 9, pp. 394 – 396, 2002.

[24] J. Hagenauer, N. Dütsch, J. Barros, and A. Schaefer, "Incremental and decremental redundancy in turbo source–channel coding," in *Proc. Int. Symposium on Control, Communications and Signal Processing*, March 2004, pp. 595–598.

[25] J. Hagenauer, J. Barros, and A. Schaefer, "Lossless turbo source coding with decremental redundancy," in *Proc. 5th Int. ITG Conference on Source and Channel Coding*, January 2004, pp. 333–339.

[26] F. Hekland, G. E. Øien, and T. A. Ramstad, "Using 2: 1 shannon mapping for joint source-channel coding," in *Proc. DCC*, 2005, pp. 223–232.

[27] C. E. Shannon, "Communication in the presence of noise," *Proc. IRE*, vol. 37, pp. 10–21, 1949.

[28] J. K. Wolf, "Redundancy, the discrete Fourier transform, and impulse noise cancellation," *IEEE Transactions on Communications*, vol. 31, no. 3, pp. 458–461, 1983.

[29] T. G. Marshall, "Coding of real-number sequences for error correction: A digital signal processing problem," *IEEE Journal on Selected Areas in Communications*, vol. 2, no. 2, pp. 381–392, 1984.

[30] J. L. Wu and J. Shiu, "Discrete cosine transform in error control coding," *IEEE trans. on Communications*, vol. 43, no. 5, pp. 1857–1861, 1995.

[31] R. E. Blahut, *Theory and Practice of Error Control Codes*. Reading, MA: Addison-Wesley, 1984.

[32] F. Marvasti, M. Hasan, M. Echhart, and S. Talebi, "Efficient algorithms for burst error recovery using FFT and other transform kernels," *IEEE trans. on Signal Processing*, vol. 47, no. 4, pp. 1065–1075, 1999.

[33] G. R. Redinbo, "Decoding real block codes: Activity detection, wiener estimation," *IEEE trans. on Information Theory*, vol. 46, no. 2, pp. 609–623, 2000.

[34] M. Ghosh, "Analysis of the effect of impulse noise on multicarrier and single carrier QAM systems," *IEEE Transactions on Communication*, vol. 44, no. 2, pp. 145–147, 1996.

[35] O. Rioul, "A spectral algorithm for removing salt and pepper from images," in *Proc. IEEE Digital Signal Processing Workshop*. 3, 1996, pp. 275–278.

[36] G. Rath and C. Guillemot, "Performance analysis and recursive syndrome decoding of DFT codes for bursty erasure recovery," *IEEE trans. on Signal Processing*, vol. 51, no. 5, pp. 1335–1350, 2003.

[37] ——, "Frame-theoretic analysis of DFT codes with erasures," *IEEE trans. on Signal Processing*, vol. 52, no. 2, pp. 447–460, 2004.

[38] R. E. Totty and G. C. Clark, "Reconstruction error in waveform transmission," *IEEE Transactions on Information Theory*, vol. 13, no. 2, pp. 336–338, 1967.

[39] A. J. Kurtenbach and P. A. Wintz, "Quantizing for noisy channels," *IEEE Trans. Communication Technology*, vol. 17, no. 10, pp. 292–302, 1969.

[40] J. B. Anderson, *Digital Transmission Ingineering*, 2nd ed. Wiley, 2005.

[41] R. E. Blahut, "Computation of channel capacity and rate-distorsion functions," *IEEE transactions on Information Theory*, vol. 18, no. 4, pp. 460–473, 1972.

[42] M. Barkat, *Signal, Detection and Estimation*. Norwood, MA: Artech House, 1991.

[43] P. J. S. G. Ferreira, D. M. S. Santos, and V. M. A. Afreixo, "New results and open problems in real-number codes," in *Proceedings of EUSIPCO*, 2004.

[44] P. Duhamel and M. Vetterli, "Fast Fourier transforms: A tutorial review and a state of the art," *Signal Processing*, vol. 19, no. 4, pp. 259–299, 1990.

[45] Y. Shoam and A. Gersho, "Efficient bit allocation for an arbitrary set of quantizers," *IEEE Trans. on Acoustics, Speech, and Signal Processing*, vol. 36, no. 9, pp. 1445–1453, 1988.

[46] "CNES report WGISS-8," 1999, `http://www.space.qinetiq.com/ceos/docs/cnes_report_wgiss.doc`.

[47] C. Lambert-Nebout, C.Latry, G. Moury, C. Parisot, M. Antonini, and M. Barlaud, "On-board optical image compression for future high resolution remote sensing systems," in *Proceedings of SPIE2000*, vol. 4115, San Diego, USA, 2000.

[48] A. K. Jain, *Fundamentals of Digital Image Processing*. Englewood Cliffs, NJ: Prentice-Hall, 1989.

[49] `http://smsc.cnes.fr/PLEIADES/`.

[50] J. Kliewer and A. Mertins, "Soft-input reconstruction of binary transmitted quantized overcomplete expansions," *IEEE Signal Processing Letters*, vol. 11, no. 11, pp. 899–903, 2004.

[51] S. Marinkovic and C. Guillemot, "Joint source-channel coding by means of oversampled filter banks codes," *Eurasip journal on applied signal processing*, no. 4, pp. 510–524, 2005.

[52] F. Labeau, J. Chiang, M. Kieffer, P. Duhamel, L. Vandendorpe, and B. Mack, "Oversampled filter banks as error correcting codes: theory and impulse correction," *IEEE Trans. on Signal Processing*, vol. 53, no. 12, pp. 4619 – 4630, 2005.

[53] T. R. Fischer, "A pyramid vector quantization," *IEEE Trans. Inform. Theory*, vol. 32, pp. 568–583, 1986.

[54] A. C. Hung, E. K. Tsern, and T. H. Meng, "Error-resilient pyramid vector quantization for image compression," *IEEE Trans. Image Processing*, vol. 7, no. 10, pp. 1373–1386, 1998.

[55] R. B. Davies, "The distribution of a linear combination of χ^2 random variables," *Statistical algorithms*, vol. 29, pp. 323–333, 1980.