From Maximum Likelihood to Iterative Decoding

Florence Alberge, Ziad Naja, Pierre Duhamel

FROM MAXIMUM LIKELIHOOD TO ITERATIVE DECODING

F. Alberge, Z. Naja, P. Duhamel

Laboratoire des Signaux et Systèmes
Univ. Paris-Sud, UMR8506 Orsay, F-91405; CNRS, Gif-sur-Yvette, F-91192;
Supelec, Gif-sur-Yvette, F-91192
e-mail:{alberge, naja, pierre.duhamel}@lss.supelec.fr

ABSTRACT

Iterative decoding is considered in this paper from an optimization point of view. Starting from the optimal maximum likelihood decoding, a (tractable) approximate criterion is derived. The global maximum of the approximate criterion is analyzed: the maximum likelihood solution can be retrieved from the approximate criterion in some particular cases. The classical equations of turbo-decoders can be obtained as an instance of an hybrid Jacobi/Gauss-Seidel implementation of the iterative maximization for the tractable criterion. The extrinsics are a natural consequence of this implementation. In the simulation part, we show a practical application of these results.

***Index Terms*—** Maximum likelihood decoding, iterative turbo-decoding, BICM

1. INTRODUCTION

Bit-Interleaved Coded Modulation (BICM) was first suggested by Zehavi in [1] to improve the Trellis Coded Modulation performance over Rayleigh-fading channels. In BICM, the diversity order is increased by using bit-interleavers instead of symbol interleavers. This improvement is achieved at the expense of a reduced minimum Euclidean distance leading to a degradation over non-fading Gaussian channels [1], [2]. This drawback can be overcome by using iterative decoding (BICM-ID) at the receiver. BICM-ID is known to provide excellent performance for both Gaussian and fading channels.

The iterative decoding scheme used in BICM-ID is very similar to serially concatenated turbo-decoders. Indeed, the serial turbo-decoder makes use of an exchange of information between computationally efficient decoders for each of the component codes. In BICM-ID, the inner decoder is replaced by demapping which is less computationally demanding than a decoding step. Even if this paper focus on iterative decoding for BICM, the results can be applied to the large class of iterative decoders including serial or parallel concatenated turbo-decoders.

The turbo-decoder and more generally iterative decoding was not originally introduced as the solution to an optimization problem rendering the analysis of its convergence and stability very difficult. Among the different attempts to provide an analysis of iterative decoding, the EXIT chart analysis and density evolution have permitted to make significant progress [3] but the results developed within this setting apply only in the case of large block length. Another tool of analysis is the connection of iterative decoding to factor graphs [4] and belief propagation [5]. Convergence results for belief propagation exists but are limited to the case where the corresponding graph is a tree which does not include turbo-code. A link between iterative decoding and classical optimization algorithms has been made also in [6] where the turbo-decoding is interpreted as a nonlinear block Gauss-Seidel iteration for solving a constrained optimization problem. In [7], the turbo-decoding is interpreted in a geometric setting as a dynamical system leading to new but incomplete results. The failure to obtain complete results is mainly due to the inability to efficiently describe extrinsic information passing. The relation between the optimal maximum likelihood decoding and iterative decoding is not yet fully understood.

In this paper, we first review the principle of maximum likelihood decoding. An approximate (and tractable) criterion is derived from an equivalent and convenient formulation of the optimal criterion. We prove that, in specific cases, the global maximum of the approximate criterion yields the maximum likelihood optimum. We then consider the iterative maximization and prove that the choice of a particular scheduling leads to the classical updates used in the iterative (turbo) decoding. In the simulation part, these results are applied to the detection of suspicious solutions *ie* with a possible large number of errors.

2. MAXIMUM-LIKELIHOOD DECODING

A conventional BICM system [2] is built from a serial concatenation of a convolutional encoder, a bit interleaver and an M-ary bits-to-symbol mapping (where $M = 2^m$) as shown in Figure 1. The sequence of information bits $\mathbf{b}$ of length n_b is first encoded by a convolutional encoder to produce the output encoded bit sequence $\mathbf{c}$ of length n which is then scrambled by a bit interleaver (as opposed to the channel symbols in the symbol interleaved coded sequence) operating on bit indexes. Let $\mathbf{d} = \pi(\mathbf{c})$ denote the interleaved sequence. Then, m consecutive bits of $\mathbf{d}$ are grouped as a symbol. The complex transmitted signal s_k, $1 \leq k \leq n/m$, is then chosen from an M-ary constellation ψ where ψ denotes the mapping scheme. For simplicity, we consider transmission over the AWGN channel. The received signals can be written as:

$$y_k = s_k + n_k \quad 1 \leq k \leq \frac{n}{m} \tag{1}$$

where n_k is a complex white Gaussian noise with independent in-phase and quadrature components having two-sided power spectral density σ_c^2. The maximum likelihood sequence detection takes the

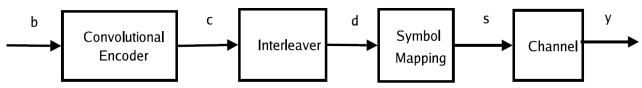

Fig. 1. BICM transmission scheme

form:

$$\hat{\mathbf{b}}_{MLD} = \arg \max_{\mathbf{b}\in\{0,1\}^{n_b}} p(\mathbf{y} \mid \mathbf{b}) \tag{2}$$

where $p(\mathbf{y} \mid \mathbf{b})$ is the likelihood function which results from concatenating the encoder with the channel. Since there is a one-to-one correspondence between the binary message $\mathbf{b}$ and the interleaved sequence $\mathbf{d}$, eq. (2) is equivalent to searching $\hat{\mathbf{d}}_{MLD}$ as:

$$\hat{\mathbf{d}}_{MLD} = \arg \max_{\mathbf{d}\in\{0,1\}^{n}} p_{ch}(\mathbf{y} \mid \mathbf{d}) I_{co}(\mathbf{d}) \tag{3}$$

where $p_{ch}(\mathbf{y} \mid \mathbf{d})$ is the probability of receiving $\mathbf{y}$ when the sequence transmitted through the channel is the mapping of $\mathbf{d}$ and where $I_{co}(\mathbf{d})$ is the indicator function of the code meaning that $I_{co}(\mathbf{d}) = 1$ if $\mathbf{c} = \pi^{-1}(\mathbf{d})$ is a codeword and 0 elsewhere. Another way to tackle this problem consists in finding the prior PMF on $\mathbf{d}$ which maximizes the *a posteriori* probability of having received $\mathbf{y}$

$$\hat{\mathbf{p}}_{MLD}(\mathbf{d}) = \arg \max_{\mathbf{p}\in\mathcal{E}_s} \sum_{\mathbf{d}} \mathbf{I}_{co}(\mathbf{d}) \mathbf{p}_{ch}(\mathbf{y} \mid \mathbf{d}) \mathbf{p}(\mathbf{d}) \tag{4}$$

where $\mathcal{E}_S$ stands for the set of all possible **separable** PMFs on $\mathbf{d}$. A PMF $\mathbf{p}(\mathbf{d})$ is separable if $\mathbf{p}(\mathbf{d}) = \prod_i p_i(d_i)$ with $p_i(d_i)$ the probability for bit i to be equal to d_i. The optimal solution $\hat{\mathbf{p}}_{MLD}(\mathbf{d})$ takes the form

$$\hat{\mathbf{p}}_{MLD}(\mathbf{d}) = \begin{cases} 1, & \mathbf{d} = \hat{\mathbf{d}}_{MLD} \\ 0, & otherwise \end{cases} \tag{5}$$

Any other weighting (with the constraint $\sum_{\mathbf{d}} p(\mathbf{d}) = 1$) produces a lower likelihood. The formulation in (4) is equivalent to the original problem in (3). The practical implementation of this optimal optimization problem is dismissed due to the presence of a random bit interleaver and to the (large) numerical value of n. In the next section, we present a sub-optimal criterion derived from (4) in a constructive way and establish some properties.

3. A SUB-OPTIMAL MAXIMUM LIKELIHOOD DECODING

3.1. Assumptions and approximations

We observed that the optimal maximum likelihood decoding is infeasible due to the interleaver and to the computational complexity involved by the computation and storage of the 2^n taps of the PMF. A solution regarding the interleaver is to consider separately the two-blocks (mapping and coding) in a particular sense to be defined later. The problem of the computational complexity can be handled by working on the bit-marginals rather than on the PMF of the whole sequence. For that purpose we split the variable $\mathbf{p}(\mathbf{d})$ into the product of the two separable PMFs ($\mathbf{l}(\mathbf{d})$ and $\mathbf{q}(\mathbf{d})$) and introduce the computation of the bit-marginals into the optimal criterion as

$$\begin{aligned} &\left(\hat{\mathbf{l}}_{MLD}(\mathbf{d}), \hat{\mathbf{q}}_{MLD}(\mathbf{d})\right) = \\ &\arg \max_{\mathbf{l},\mathbf{q}\in\mathcal{E}_s} \sum_{d_k} \sum_{\mathbf{d}:d_k} \mathbf{I}_{co}(\mathbf{d}) \mathbf{p}_{ch}(\mathbf{y} \mid \mathbf{d}) \mathbf{l}(\mathbf{d}) \mathbf{q}(\mathbf{d}) \end{aligned} \tag{6}$$

The double sum above is exactly the same as the some over all the words $\mathbf{d}$. The global maximum is again obtained for the optimal choice of the weights $\mathbf{l}(\mathbf{d})\mathbf{q}(\mathbf{d})$. Since $\mathbf{l}(\mathbf{d})$ and $\mathbf{q}(\mathbf{d})$ are PMF, $\sum_{\mathbf{d}} \mathbf{l}(\mathbf{d})\mathbf{q}(\mathbf{d}) \leq 1$, and the optimal weighting strategy is again

$$\hat{\mathbf{l}}_{MLD}(\mathbf{d}) = \hat{\mathbf{q}}_{MLD}(\mathbf{d}) \begin{cases} 1, & \mathbf{d} = \hat{\mathbf{d}}_{MLD} \\ 0, & otherwise \end{cases} \tag{7}$$

The formulation in (6) is then an equivalent form on the original problem since the two solutions $\hat{\mathbf{l}}_{MLD}(\mathbf{d})$ and $\hat{\mathbf{q}}_{MLD}(\mathbf{d})$ both select the optimal sequence $\hat{\mathbf{d}}_{MLD}$ of the maximum likelihood decoding problem. Let $\mathcal{C}_{MLD}$ denote the criterion in (6). The direct maximization of $\mathcal{C}_{MLD}$ is still untractable. We need to separate the coding part from the mapping and channel part, this can done by replacing the bit-marginals of the product of two PMFs by the product of the bit-marginals of the two PMFs taken separately. This is of course an approximation leading to the new criterion $\tilde{\mathcal{C}}_k$ defined as

$$\tilde{\mathcal{C}}_k = \left(\sum_{\mathbf{d_k}} \left(\sum_{\mathbf{d}:d_k} \mathbf{I}_{co}(\mathbf{d})\mathbf{q}(\mathbf{d})\right) \left(\sum_{\mathbf{d}:\mathbf{d_k}} \mathbf{p_{ch}}(\mathbf{y} \mid \mathbf{d})\mathbf{l}(\mathbf{d})\right)\right) \tag{8}$$

This approximation deserves some comments. First, the bit-marginals in $\tilde{\mathcal{C}}_k$ are now computable in practice. For example $\sum_{\mathbf{d}:d_k} \mathbf{I}_{co}(\mathbf{d})\mathbf{q}(\mathbf{d})$, $1 \leq k \leq n$, $d_k \in \{0,1\}$ is exactly the output given by a BCJR [8]. Next, the criterion $\tilde{\mathcal{C}}_k$ is dependant of k: the quantities involved in the criterion are not the same for two different values of k (whereas $\mathcal{C}_{MLD}$ is independent of k). This suggests that criterion $\tilde{\mathcal{C}}_k$ should be used for the **maximization over the kth bit-marginal**. The maximization of the unique criterion $\mathcal{C}_{MLD}$ has been turned into a distributed optimization of the n criteria $\tilde{\mathcal{C}}_k$. Last, the criteria $\mathcal{C}_{MLD}$ and $\tilde{\mathcal{C}}_k$, $1 \leq k \leq n$, are the same (meaning there is no approximation) if the two PMFs involved $\mathbf{I}_{co}(\mathbf{d})\mathbf{q}(\mathbf{d})$ and $\mathbf{p}_{ch}(\mathbf{y} \mid \mathbf{d})\mathbf{l}(\mathbf{d})$) are separable. We can notice that for $\mathbf{l}(\mathbf{d}) = \hat{\mathbf{l}}_{MLD}(\mathbf{d})$ and $\mathbf{q}(\mathbf{d}) = \hat{\mathbf{q}}_{MLD}(\mathbf{d})$ (defined in (7)) the two PMFs are indeed separable. This is also true for all the class of "Kronecker" PMFs in which the global optimum is always lying.
In the next subsection, we focus on the maximization of the sub-optimal criteria $\tilde{\mathcal{C}}_k$ and derive some interesting properties.

3.2. Sub-optimal criterion and global maximum

We prove in the two propositions below that in some special cases, the criteria $\tilde{\mathcal{C}}_k$ yields the same global maxima as the optimal criterion $\mathcal{C}_{MLD}$.

Proposition 1 *The maximum of any criterion $\tilde{\mathcal{C}}_k$, $1 \leq k \leq n$ is obtained for $\mathbf{q} = \hat{\mathbf{q}}$ and $\mathbf{l} = \hat{\mathbf{l}}$ such that*

$$\hat{\mathbf{l}}(\mathbf{d}')\hat{\mathbf{q}}(\mathbf{d}) = \begin{cases} 1, & (\mathbf{d}, \mathbf{d}') = (\hat{\mathbf{d}}, \hat{\mathbf{d}}') \\ 0, & otherwise \end{cases} \tag{9}$$

where $(\hat{\mathbf{d}}, \hat{\mathbf{d}}') = \arg\max_{\mathbf{d},\mathbf{d}'\in\{0,1\}^n} \mathbf{p}_{ch}(\mathbf{y} \mid \mathbf{d}')\mathbf{I}_{co}(\mathbf{d})$.

Proof: Since $\mathbf{q}$ and $\mathbf{l}$ are in $\mathcal{E}_s$, $\tilde{\mathcal{C}}_k$ reads

$$\tilde{\mathcal{C}}_k = \sum_{\mathbf{d_k}} \left(\sum_{\mathbf{d}:d_k} \sum_{\mathbf{d}':d_k} \mathbf{I}_{co}(\mathbf{d})\mathbf{p}(\mathbf{y} \mid \mathbf{d}') \prod_j q_j(d_j) l_j(d'_j)\right) \tag{10}$$

and $\sum_{d_k} \sum_{\mathbf{d}:d_k} \sum_{\mathbf{d}':d_k} \prod_j q_j(d_j) l_j(d'_j) = \sum_{d_k} l_k(d'_k) q_k(d_k)$. In other words,

$$\frac{1}{\sum_{d_k} l_k(d'_k) q_k(d_k)} \sum_{d_k} \sum_{\mathbf{d}:d_k} \sum_{\mathbf{d}':d_k} \prod_j q_j(d_j) l_j(d'_j) = 1 \tag{11}$$

Thus $\frac{1}{\sum_{d_k} l_k(d'_k) q_k(d_k)} \mathcal{C}_k$ is maximum for $\hat{\mathbf{q}}(\mathbf{d}')\hat{\mathbf{l}}(\mathbf{d}) = 1$ for $(\mathbf{d}, \mathbf{d}') = (\hat{\mathbf{d}}, \hat{\mathbf{d}}')$ and $\hat{\mathbf{q}}(\mathbf{d})\hat{\mathbf{l}}(\mathbf{d}') = 0$ for the other pairs involved in (10). All the pairs $(\mathbf{d}, \mathbf{d}')$ have the same value for the bit k, this is true in particular for $(\hat{\mathbf{d}}, \hat{\mathbf{d}}')$. Then $\sum_{d_k} l_k(d'_k) q_k(d_k)$ and consequently $\tilde{\mathcal{C}}_k$ are also maximized by $\hat{\mathbf{q}}(\mathbf{d})\hat{\mathbf{l}}(\mathbf{d}')$.

□

Let $\hat{\mathbf{d}}_{ch} = \arg\max_{\mathbf{d}\in\{0,1\}^n} p(\mathbf{y} \mid \mathbf{d})$. Let suppose that $\hat{\mathbf{d}}_{ch}$ is a **codeword**. This is likely to be so at high SNR. Then $(\hat{\mathbf{d}}_{ch}, \hat{\mathbf{d}}_{ch}) = \arg\max_{\mathbf{d},\mathbf{d}'\in\{0,1\}^n} p_{ch}(\mathbf{y} \mid \mathbf{d}')I_{co}(\mathbf{d})$ and also $\hat{\mathbf{d}}_{ch} = \hat{\mathbf{d}}_{MLD} = \arg\max_{\mathbf{d}\in\{0,1\}^n} p_{ch}(\mathbf{y} \mid \mathbf{d})I_{co}(\mathbf{d})$. Then each criterion $\tilde{\mathcal{C}}_k$ has a global maximum at $(\hat{\mathbf{l}}_{MLD}(\mathbf{d}),\ \hat{\mathbf{q}}_{MLD}(\mathbf{d}))$. We can also remark that the others global maximizers of the individual criterion $\tilde{\mathcal{C}}_k$ are not global maximizers of all the others criteria $\tilde{\mathcal{C}}_i$, with $1 \leq i \leq n$ and $i \neq k$. As a conclusion, if the channel probability $p_{ch}(\mathbf{y} \mid \mathbf{d})$ reaches its maximum for a particular value of $\mathbf{d}$ corresponding to a codeword then the joint global maximization of criteria $\mathcal{C}_k$ for $1 \leq k \leq n$ yields the same solution (given in (7)) than the maximum likelihood decoding. We can now define a new criterion $\tilde{\mathcal{C}} = \sum_{k=1}^{n} \tilde{\mathcal{C}}_k$ which appears to be a relevant approximation of the optimal criterion $\mathcal{C}$.

Proposition 2 *Let suppose that $\tilde{\mathcal{C}}$ has a global maximum at $(\hat{\mathbf{l}}_{\tilde{\mathcal{C}}}, \hat{\mathbf{q}}_{\tilde{\mathcal{C}}})$. If $(\hat{\mathbf{l}}_{\tilde{\mathcal{C}}}, \hat{\mathbf{q}}_{\tilde{\mathcal{C}}})$ is such that $\hat{\mathbf{l}}_{\tilde{\mathcal{C}}}\hat{\mathbf{q}}_{\tilde{\mathcal{C}}}(\mathbf{d}) = 1$ at $\mathbf{d} = \mathbf{d}_0$ and 0 otherwise then $\mathbf{d}_0 = \hat{\mathbf{d}}_{MLD} = \arg\max_{\mathbf{d}\in\{0,1\}^n} p_{ch}(\mathbf{y} \mid \mathbf{d})I_{co}(\mathbf{d})$.*

Proof: $\tilde{\mathcal{C}}$ has a global maximum at $(\hat{\mathbf{l}}_{\tilde{\mathcal{C}}}, \hat{\mathbf{q}}_{\tilde{\mathcal{C}}})$ then $\tilde{\mathcal{C}}(\hat{\mathbf{l}}_{\tilde{\mathcal{C}}}, \hat{\mathbf{q}}_{\tilde{\mathcal{C}}}) \geq \tilde{\mathcal{C}}(\mathbf{l}, \mathbf{q})$ for any PMF $\mathbf{l}$, $\mathbf{q}$. From the definition of $\tilde{\mathcal{C}}$ we have $\tilde{\mathcal{C}}(\mathbf{l}, \mathbf{q}) \geq n\mathcal{C}_{MLD}(\mathbf{l}, \mathbf{q})$. Moreover $\tilde{\mathcal{C}}(\hat{\mathbf{l}}_{\tilde{\mathcal{C}}}, \hat{\mathbf{q}}_{\tilde{\mathcal{C}}}) = n\mathcal{C}_{MLD}(\hat{\mathbf{l}}_{\tilde{\mathcal{C}}}, \hat{\mathbf{q}}_{\tilde{\mathcal{C}}})$. Thus $\mathcal{C}_{MLD}(\hat{\mathbf{l}}_{\tilde{\mathcal{C}}}, \hat{\mathbf{q}}_{\tilde{\mathcal{C}}}) \geq \mathcal{C}_{MLD}(\mathbf{l}, \mathbf{q})$ for any PMF $\mathbf{l}$, $\mathbf{q}$.

□

If we manage to find the global maximum of the sub-optimal criterion $\tilde{\mathcal{C}}$ and if the corresponding argument turns out to be a Kronecker PMF then this is also the argument of a global maximum of the optimal criterion $\mathcal{C}_{MLD}$ associated with the maximum likelihood decoding. The practical usefulness of criterion $\tilde{\mathcal{C}}$ will be emphasized in section 4. In the next subsection, we build an iterative strategy of maximization.

3.3. Iterative maximization

We observed in section 3.2 that the sub-optimal criterion $\tilde{\mathcal{C}}_k$ was derived from $\mathcal{C}_{MLD}$ when dealing with the kth bit-marginal. We propose here to consider a distributed maximization strategy where $l_k(d_k)$ and $q_k(d_k)$ are chosen in order to maximize $\tilde{\mathcal{C}}_k$ as

$$\left(\hat{l}_k, \hat{q}_k\right) = \arg \max_{l_k, q_k \in \mathcal{F}} \tilde{\mathcal{C}}_k \tag{12}$$

where $\mathcal{F}$ is the set of all possible PMFs on d_k. The solution of (12) is given by

$$\begin{array}{c} \hat{l}_k(d_k)\hat{q}_k(d_k) = 1 \quad if \\ f_{d_k}(\mathbf{q}, I_{co}) f_{d_k}(\mathbf{l}, p_{ch}(\mathbf{y} \mid \mathbf{d})) > f_{\overline{d}_k}(\mathbf{q}, I_{co}) f_{\overline{d}_k}(\mathbf{l}, p_{ch}(\mathbf{y} \mid \mathbf{d})) \\ \hat{l}_k(d_k)\hat{q}_k(d_k) = 0 \quad otherwise \end{array} \tag{13}$$

where $f_{d_k}(\mathbf{q}, I_{co}) = \sum_{\mathbf{d}:d_k} \mathbf{I}_{co}(\mathbf{d}) \prod_{j\neq k} q_j(d_j)$, $f_{d_k}(\mathbf{l}, p_{ch}(\mathbf{y} \mid \mathbf{d})) = \sum_{\mathbf{d}:d_k} \mathbf{p}_{ch}(\mathbf{y} \mid \mathbf{d}) \prod_{j\neq k} l_j(d_j)$ and $\overline{d}_k = 1 - d_k$. An iterative process propagating hard estimates (0 or 1) is likely to get stuck in a local minima. A classical solution is to propagate instead soft-estimates (in $[0; 1]$) and take hard decisions at the end of the iterative process. For the maximization problem in (12), possible soft estimates are :

$$\begin{array}{l} \hat{l}_k(d_k)\hat{q}_k(d_k) \propto f_{d_k}(\mathbf{q}, I_{co}) f_{d_k}(\mathbf{l}, p_{ch}(\mathbf{y} \mid \mathbf{d})) \\ \hat{l}_k(\overline{d}_k)\hat{q}_k(\overline{d}_k) \propto f_{\overline{d}_k}(\mathbf{q}, I_{co}) f_{\overline{d}_k}(\mathbf{l}, p_{ch}(\mathbf{y} \mid \mathbf{d})) \end{array} \tag{14}$$

Equation (14) characterizes the product $\hat{l}_k\hat{q}_k$. The individual values of $\hat{l}_k$ and $\hat{q}_k$ depend on the scheduling of the successive updates. Since we have no prior information, a natural choice for the initialization is $l_k^{(0)}(d_k) = q_k^{(0)}(d_k) = \frac{1}{2}$ for $1 \leq k \leq n$ and $d_k \in \{0; 1\}$. Let consider first the update of variables q_k. Following a Jacobi implementation, $q_k^{(it)}$ is obtained from (14) by setting $l_k(d_k) = l_k^{(it-1)}(d_k)$ for $1 \leq k \leq n$ and $q_i(d_i) = q_i^{(it-1)}(d_i)$ for $i \neq k$. In particular, the update at the first iteration is:

$$q_k^{(1)}(d_k) \propto \sum_{\mathbf{d}:d_k} I_{co}(\mathbf{d}) \sum_{\mathbf{d}:d_k} p_{ch}(\mathbf{y} \mid \mathbf{d}) \tag{15}$$

for $d_k \in \{0; 1\}$ and for all $k \in \{1, ..., n\}$. The update of $\mathbf{l}_k$ also comes from (14) by setting $l_i(d_i) = l_i^{(it-1)}(d_i)$ for $i \neq k$ and $q_k(d_k) = q_k^{(it)}(d_k)$ for $1 \leq k \leq n$ since $q_k^{(it)}(d_k)$ has just been computed and is available. This is an hybrid Jacobi/Gauss-Seidel implementation. At the first iteration, the update for l_k is:

$$l_k^{(1)}(d_k) \propto \frac{\sum_{\mathbf{d}:d_k} I_{co}(\mathbf{d}) \prod_{j\neq k} q_j^{(1)}(d_j)}{\sum_{\mathbf{d}:d_k} I_{co}(\mathbf{d})} \tag{16}$$

The generalization to iteration (it) reads:

$$\begin{array}{l} q_k^{(it)}(d_k) \propto \sum_{\mathbf{d}:d_k} I_{co}(\mathbf{d}) \sum_{\mathbf{d}:d_k} p_{ch}(\mathbf{y} \mid \mathbf{d}) \prod_{j\neq k} l_j^{(it-1)}(d_j) \\ l_k^{(it)}(d_k) \propto \frac{\sum_{\mathbf{d}:d_k} I_{co}(\mathbf{d}) \prod_{j\neq k} q_j^{(it)}(d_j)}{\sum_{\mathbf{d}:d_k} I_{co}(\mathbf{d})} \end{array} \tag{17}$$

for $d_k \in \{0; 1\}$ and for all $k \in \{1, ..., n\}$. In general, for convolutive codes, $\sum_{\mathbf{d}:d_k} I_{co} = \sum_{\mathbf{d}:\overline{d}_k} I_{co}$. The iterative updates are then obtained through:

$$\begin{array}{l} q_k^{(it)}(d_k) \propto \sum_{\mathbf{d}:d_k} p_{ch}(\mathbf{y} \mid \mathbf{d}) \prod_{j\neq k} l_j^{(it-1)}(d_j) \\ l_k^{(it)}(d_k) \propto \sum_{\mathbf{d}:d_k} I_{co}(\mathbf{d}) \prod_{j\neq k} q_j^{(it)}(d_j) \end{array} \tag{18}$$

This is exactly the equations that are used in the iterative decoding of BICM where $l_k(d_k)$, $q_k(d_k)$ are usually called *extrinsics* and $l_k(d_k)q_k(d_k)$ (after normalization) is the *APP* (*A Posteriori Probability*) [9]. The extrinsics are a direct consequence of the maximization of a well-defined criterion and of the choice of a particular scheduling in the computation of the successive updates. The algorithm stops in general when an agreement is reached between the APP computed at the demapper and the APP computed at the decoder. This corresponds with our notations to:

$$q_k^{(it)}(d_k)l_k^{(it-1)}(d_k) \propto q_k^{(it)}(d_k)l_k^{(it)}(d_k) \tag{19}$$

for $d_k \in \{0; 1\}$ and for all $k \in \{1, ..., n\}$. This means that the tentative maximization (via soft estimates) of $\tilde{\mathcal{C}}_k$ for $1 \leq k \leq n$ with respect either to $l_k(d_k)$ or to $q_k(d_k)$ produces exactly the same result regarding the product (APP) $l_k(d_k)q_k(d_k)$. The derivation in this paper are mainly based on the block structure (code, mapping) of the BICM and not on the specificity of each block. The line of arguments followed in this paper and the conclusions apply to a wide range of problems including iterative turbo-decoding. For instance, the serial turbo-decoder is obtained from BICM by replacing the PMF $p_{ch}(\mathbf{y} \mid \mathbf{d})$ with $p_{ch}(\mathbf{y} \mid \mathbf{d})I_{co1}(\mathbf{d})$ where $I_{co1}(\mathbf{d})$ is the indicator function of the inner coder involved at the transmitter. Since, no assumptions has been made on $p_{ch}(\mathbf{y} \mid \mathbf{d})$ within this paper, the whole conclusions apply to serial turbo-decoders as well as to any decoder with a structure similar to BICM.

4. SIMULATIONS

The criterion $\tilde{\mathcal{C}} = \sum_{k=1}^{n} \tilde{\mathcal{C}}_k$ is an approximation of the MLD criterion $\mathcal{C}$. The solution for the approximate criterion $\tilde{\mathcal{C}}$ is expected to be close to the MLD solution and hence to achieve low bit error rate (BER). Iterative turbo-decoding is known to provide excellent performance in terms of BER. We can however observe at low SNR a degradation in the performance mainly due to inaccurate solutions of the iterative decoder for some isolated frames. In the following, we use $\tilde{\mathcal{C}}$ has a detector to separate the acceptable solutions (few errors expected) from the suspicious ones (many more errors expected).
We used a classical transmitter BICM scheme with a $(5, 7)$ convolutional code of rate $1/2$. The number of information bits is $n_b = 400$ (a frame). The code bits were passed through a random interleaver and modulated to 16-QAM symbols. The signal to noise ratio is defined as $\frac{E_b}{N_0}$, where E_b denotes the energy per information bit and N_0 is the noise variance. We consider an environment with varying $\frac{E_b}{N_0}$ in the range $\{4dB, 5dB, ..., 11dB, 12dB\}$. The values are chosen randomly from a uniform distribution. The iterative decoding is performed using eq. (18). For each frame, the iterative process is run until an agreement is reached for all the bits (eq. 19) or the maximum number of iterations is reached. The solutions are qualified of acceptable if $\sum_{k=1}^{n} \log(\tilde{\mathcal{C}}_k)$ is greater (at the end of the iterative process) than a certain threshold (to be chosen) and are qualified of suspicious if $\sum_{k=1}^{n} \log(\tilde{\mathcal{C}}_k)$ is under the threshold. We compute BER_a resp. BER_s the bit error rates of the acceptable solutions resp. the BER of suspicious solutions. We expect BER_s to be many more large than BER_a. We also evaluate the proportion of suspicious solutions identified (p_s) and the false alarm proportion ($p_{s,false}$) which counts the solutions wrongly qualified of suspicious. We consider that the solutions with less than 6 errors (among the 400 bits) should not have been qualified of suspicious. The algorithm stops when the total number of errors among the acceptable sequences is greater than 200. The results are reported in table 1.

Threshold	-20	-10	-5
BER_a	$8,78.10^{-4}$	$4,68.10^{-4}$	$2,08.10^{-4}$
BER_s	$0,205$	$0,13$	$9,28.10^{-2}$
p_s %	$6,4\%$	$10,8\%$	$14,8\%$
$p_{false,s}$ %	$2,5\%$	$36,4\%$	$53,83\%$

Table 1. Evaluation of Acceptable/Suspicious solutions

The validity of $\tilde{\mathcal{C}}$ as a relevant approximation of $\mathcal{C}$ is confirmed by this simulation. The BER is strongly correlated with the value of $\tilde{\mathcal{C}}$ reached at the end of the iterative process. If the value is close to the global maximum, the solution is also close (or equal) to the optimal MLD solution. At the opposite, a low value of $\tilde{\mathcal{C}}$ is often associated with a large number of errors. This is also observable in fig. (2) where the distribution of the errors (histograms) are plotted for the acceptable solutions and compared to the histogram obtained for the suspicious solutions in the case where the threshold is -20. We can conclude from this experiment that the suspicious solutions are mainly due to local minima ($\tilde{\mathcal{C}}$ is often very far from the threshold) rather than to an inaccuracy of $\tilde{\mathcal{C}}$ as an approximate criterion for $\mathcal{C}$. For a practical point of view, the simulation above illustrates how we can guarantee a given performance (in terms of BER) independently of the SNR (at the cost of a rejection/re-emission of the suspicious frames).

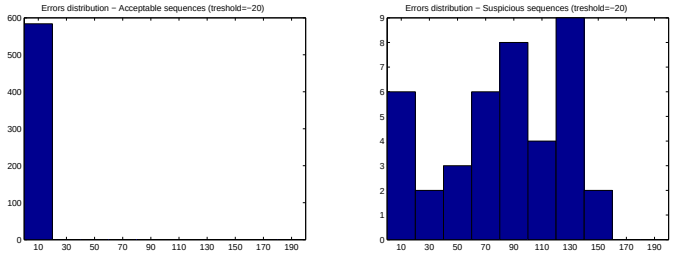

Fig. 2. Histogram representation of the error distribution for : (left) acceptable solutions - (right) suspicious solutions

5. CONCLUSION

In this paper, we derived the (turbo) iterative decoding from maximum likelihood decoding. The approximation needed to obtain a tractable solution are clearly stated and some proofs have been established concerning the global maximum of the approximate criterion. It has also be shown that the turbo-decoding follows from an hybrid Jacobian/Gauss-Seidel implementation of the maximization process. The propagation of extrinsics is naturally introduced and is a direct consequence of the scheduling. In the simulation part, a possible application of these results has been presented. In this paper, we proved that iterative turbo-decoding can be interpreted as a distributed optimization strategy reminiscent of mixed cooperative/non cooperative games. The theoretical tools developed within the game theory framework constitute a new open perspective for completing the analysis of turbo-like decoding.

6. REFERENCES

[1] E. Zehavi, "8-PSK trellis codes for a Rayleigh fading channel," *IEEE Trans. on Commun.*, vol. 40, pp. 873–883, May 1992.

[2] G. Caire, G.Taricco, and E. Biglieri, "Bit-interleaved coded modulation," *IEEE Trans. on Inform. Theory*, vol. 4, pp. 927–946, May 1998.

[3] S. ten Brink, "Convergence behavior of iteratively decoded parallel concatenated codes," *IEEE Trans. on Commun.*, vol. 49, pp. 1727–1737, Oct 2001.

[4] F.R. Kschischang, B.J. Frey, and H.A. Loeliger, "Factor graphs and the sum-product algorithm," *IEEE Trans. on Inform. Theory*, vol. 47, pp. 498–519, Feb. 2001.

[5] J. Pearl, *Probabilistic Reasoning in Intelligent Systems: Network of Plausible Inference*, San Francisco, CA: Morgan Kaufmann, 1988.

[6] J. M. Walsh, P.A. Regalia, and C. R. Johnson, "Turbo decoding as Iterative Constrained Maximum-Likelihood Sequence Detection," *IEEE Trans. on Inform. Theory*, vol. 52, pp. 5426–5437, Dec. 2006.

[7] T. Richardson, "The geometry of turbo-decoding dynamics," *IEEE Trans. on Inform. Theory*, vol. 46, no. 1, pp. 9–23, 2000.

[8] L.R. Bahl, J. Cocke, F. Jelinek, and J.Raviv, "Optimal Decoding of Linear Codes for Minimizing Symbol Error Rate," *IEEE Trans. on Inform. Theory*, pp. 284–287, March 1974.

[9] F. Alberge, "Iterative decoding as Dykstra's algorithm with alternate I-projection and reverse I-projection," in *EUSIPCO Proc.*, Lausanne, Switzerland, August 2008.