

GENERALIZED MAP: SEQUENCE DETECTION FOR NON-IDEAL FREQUENCY SELECTIVE CHANNEL KNOWLEDGE

Noura Sellami¹, Mohamed Siala², Aline Roumy³ and Ines Kammoun¹

¹ ISECS, Route Menzel Chaker km 0.5, B.P 868, 3018 Sfax, Tunisia

² SUP'COM, Cité Technologique des Communications, 2083 El Ghazala Ariana, Tunisia

³ IRISA-INRIA, Campus de Beaulieu, 35042 Rennes Cedex, France

ABSTRACT

In this paper, we consider the problem of Maximum *a Posteriori* (MAP) equalization of the received signal over a frequency selective channel when the channel is not perfectly known at the receiver. The derivation of the MAP criterion in this case leads to an expression for which no exact implementation exists in the literature. In this paper, we propose to solve the problem by using the Expectation-Maximization (EM) algorithm. The algorithm we propose has linear-time complexity per iteration. Simulations show that few iterations are required to reach the performance of the MAP equalizer with perfect channel knowledge.

Index Terms— MAP equalization, channel estimation, EM algorithm, frequency selective channel

1. INTRODUCTION

An important source of degradation in high data rate communication systems is the presence of intersymbol interference (ISI) between consecutive data symbols originating from the frequency selectivity of mobile radio channels. To combat the effects of intersymbol interference, an equalizer has to be used. In practice, the receiver does not know perfectly the channel and has to estimate it. In this paper, we consider the case where the equalizer has an imperfect channel estimate provided by another module in the receiver, namely the channel estimator. In most previous works, the equalizer assumes perfect channel knowledge and uses the channel estimate as if it was the true channel. This approach is obviously sub-optimal and the equalization algorithm has to be rederived in order to take into account the channel estimation errors. This problem of equalization for non-ideal channel knowledge has been tackled in [1, 2]. In [1], Tüchler *et al.* considered a Maximum *a Posteriori* (MAP) equalizer in the case of transmission over a frequency selective channel. They derived the metric to be minimized for this problem and suggested to use sequential decoding or sphere decoding to solve it, since it

was no feasible for a direct practical implementation. In [2], an optimum MAP discrete-time rake receiver has been proposed for CDMA systems when the channel is not perfectly known. In this case, the algorithm can be implemented since the probability density function (pdf) involved in the derivation of the MAP criterion can be factored unlike the pdf in [1]. In this paper, we consider, as in [1], a data transmission system over a frequency selective channel. Our aim is to find a practical implementation to solve the exact problem of MAP equalization with imperfect channel knowledge. To do this, we propose to use an iterative algorithm following the Expectation Maximization (EM) approach [3]. The Maximization step is performed by a Viterbi algorithm. Thus, the algorithm we propose has linear-time complexity per iteration. Moreover, simulation results show that few iterations are required to reach the optimal performance achieved when the channel is perfectly known at the receiver.

Throughout this paper scalars are lower case, and vectors and matrices are bold lower and upper case, respectively. The operators $(\cdot)^T$, $(\cdot)^*$ and $(\cdot)^\dagger$ denote respectively transposition, conjugation and transconjugation. The $L \times L$ identity matrix is denoted by \mathbf{I}_L .

2. SYSTEM MODEL

We consider a data transmission system over a frequency selective channel. The input information bit sequence is mapped to the symbol alphabet \mathcal{A} . We assume that transmissions are organized into bursts of K symbols. The channel is supposed to be invariant during one burst. The baseband complex signal sampled at the symbol rate and received at time k is

$$r_k = \sum_{l=0}^{L-1} h_l s_{k-l} + w_k = \mathbf{s}_k^T \mathbf{h} + w_k, \quad \forall 0 \leq k \leq K + L - 2 \quad (1)$$

where:

- 1) $s_k \in \mathbb{C}$, for $0 \leq k \leq K-1$, are the transmitted symbols and take their values from the alphabet \mathcal{A} . The virtual symbols $s_k \in \mathbb{C}$, for $1-L \leq k \leq -1$, are assumed to be known at the receiver and can be set to 0. We introduce the vector $\mathbf{s}_k = [s_k \dots s_{k-L+1}]^T \in \mathbb{C}^L$.

A. Roumy is supported in part by the Network of Excellence in Wireless Communications (NEWCOM), E. C. Contract no. 507325.

- 2) $w_k \in \mathbb{C}$, for $0 \leq k \leq K + L - 2$, are AWGN complex samples with pdf $\mathcal{N}_{\mathbb{C}}(0, \sigma^2)$, where $\mathcal{N}_{\mathbb{C}}(\alpha, \sigma^2)$ denotes a Gaussian distribution with mean α and variance σ^2 .
- 3) $\mathbf{h} = [h_0 \dots h_{L-1}]^T \in \mathbb{C}^L$ is the channel realization vector and is circularly symmetric complex Gaussian distributed with pdf $\mathcal{N}_{\mathbb{C}}(\mathbf{0}, \mathbf{R}_h)$, where \mathbf{R}_h is the channel covariance matrix. The pdf of \mathbf{h} is given by

$$p(\mathbf{h}) = \frac{1}{\det(\pi \mathbf{R}_h)} \exp(-\mathbf{h}^\dagger \mathbf{R}_h^{-1} \mathbf{h}) \quad (2)$$

- 4) K and L denote the burstlength and the channel memory, respectively.

Let $\mathbf{s} = [s_{K-1}, \dots, s_{1-L}]^T$ be the $(L + K - 1)$ -long vector of symbols and \mathbf{S} be the associated $(K + L - 1) \times L$ Hankel matrix having the last column $[s_{K-1}, \dots, s_{1-L}]^T$ and the last row $[s_0, \dots, s_{1-L}]$. The received signal model in (1) can be rewritten in a matrix form as

$$\mathbf{r} = \mathbf{S}\mathbf{h} + \mathbf{w} \quad (3)$$

with $\mathbf{r} = [r_{K+L-2}, \dots, r_0]^T$ and $\mathbf{w} = [w_{K+L-2}, \dots, w_0]^T$. It follows that:

$$\begin{aligned} p(\mathbf{r}|\mathbf{h}, \mathbf{s}) &= \frac{1}{(\pi\sigma^2)^{K+L-1}} \exp\left(-\sum_{k=0}^{K+L-2} \frac{|r_k - \mathbf{s}_k^T \mathbf{h}|^2}{\sigma^2}\right) \\ &= \frac{1}{(\pi\sigma^2)^{K+L-1}} \exp\left(-\frac{(\mathbf{r} - \mathbf{S}\mathbf{h})^\dagger (\mathbf{r} - \mathbf{S}\mathbf{h})}{\sigma^2}\right) \end{aligned} \quad (4)$$

The channel estimator computes estimates $\hat{\mathbf{h}}$ of the channel that can be modeled as a noisy version of \mathbf{h} :

$$\hat{\mathbf{h}} = \mathbf{h} + \mathbf{b}$$

with \mathbf{b} being a complex AWGN, independent of \mathbf{h} . This is a general model that encompasses the least mean square error estimator of the channel that uses a training sequence [4]. It follows that

$$p(\hat{\mathbf{h}}|\mathbf{h}) = \frac{1}{\det(\pi \mathbf{R}_b)} \exp\left(-(\hat{\mathbf{h}} - \mathbf{h})^\dagger \mathbf{R}_b^{-1} (\hat{\mathbf{h}} - \mathbf{h})\right) \quad (5)$$

$$p(\hat{\mathbf{h}}) = \frac{1}{\det(\pi(\mathbf{R}_h + \mathbf{R}_b))} \exp\left(-\hat{\mathbf{h}}^\dagger (\mathbf{R}_h + \mathbf{R}_b)^{-1} \hat{\mathbf{h}}\right) \quad (6)$$

3. PROBLEM STATEMENT

When the channel is known, the data estimate according to the MAP sequence criterion is given by

$$\hat{\mathbf{s}} = \arg \max_{\mathbf{s}} p(\mathbf{s}|\mathbf{r}, \mathbf{h}). \quad (7)$$

which can be solved efficiently by the Viterbi algorithm [5]. Here, however, an estimate $\hat{\mathbf{h}}$ but not the true channel \mathbf{h} is

available at the receiver. Replacing \mathbf{h} by $\hat{\mathbf{h}}$ in the MAP algorithm leads to an error for which a closed form expression is derived in [6]. Hence, in this paper, we consider the exact MAP problem that takes into account the channel estimate error. This MAP problem reads:

$$\hat{\mathbf{s}} = \arg \max_{\mathbf{s}} p(\mathbf{s}|\mathbf{r}, \hat{\mathbf{h}}). \quad (8)$$

This criterion was first derived in [1] but no efficient implementation was yet proposed. In this paper, we propose to use the EM-algorithm which has a linear-time complexity per iteration.

4. PROPOSED SOLUTION: EM-ALGORITHM

We consider the EM algorithm [3, 7] to solve the problem (8) with a reasonable complexity. The algorithm consists in two steps:

$$E: \quad Q(\mathbf{s}, \mathbf{s}^i) = \int \log p(\mathbf{s}|\mathbf{r}, \mathbf{h}, \hat{\mathbf{h}}) p(\mathbf{h}|\mathbf{r}, \hat{\mathbf{h}}, \mathbf{s}^i) d\mathbf{h} \quad (9)$$

$$M: \quad \mathbf{s}^{i+1} = \arg \max_{\mathbf{s}} Q(\mathbf{s}, \mathbf{s}^i) \quad (10)$$

where E stands for the expectation step and M for the maximization step of the EM algorithm. It can be shown that under some mild conditions and for close enough initial estimate \mathbf{s}^0 to the global maximum [7], the algorithm converges to the MAP estimated sequence $\hat{\mathbf{s}}$ defined in (8):

$$\mathbf{s}^i \xrightarrow{i \rightarrow \infty} \hat{\mathbf{s}}$$

In order to justify the previous result, we show that the APP (*a posteriori* probability) $p(\mathbf{s}|\mathbf{r}, \hat{\mathbf{h}})$ in (8) increases with any increase in the function $Q(\mathbf{s}, \mathbf{s}^i)$ in (9), with respect to \mathbf{s} . By using the inequality $\log x \leq x - 1, \forall x > 0$, coming from the concavity of the log function, we obtain:

$$\begin{aligned} &Q(\mathbf{s}, \mathbf{s}^i) - Q(\mathbf{s}^i, \mathbf{s}^i) \\ &= \frac{1}{p(\mathbf{s}^i|\mathbf{r}, \hat{\mathbf{h}})} \int \log \frac{p(\mathbf{s}, \mathbf{h}|\mathbf{r}, \hat{\mathbf{h}})}{p(\mathbf{s}^i, \mathbf{h}|\mathbf{r}, \hat{\mathbf{h}})} p(\mathbf{s}^i, \mathbf{h}|\mathbf{r}, \hat{\mathbf{h}}) d\mathbf{h} \\ &\leq \frac{1}{p(\mathbf{s}^i|\mathbf{r}, \hat{\mathbf{h}})} \int \left(\frac{p(\mathbf{s}, \mathbf{h}|\mathbf{r}, \hat{\mathbf{h}})}{p(\mathbf{s}^i, \mathbf{h}|\mathbf{r}, \hat{\mathbf{h}})} - 1 \right) p(\mathbf{s}^i, \mathbf{h}|\mathbf{r}, \hat{\mathbf{h}}) d\mathbf{h} \\ &= \frac{1}{p(\mathbf{s}^i|\mathbf{r}, \hat{\mathbf{h}})} \left(p(\mathbf{s}|\mathbf{r}, \hat{\mathbf{h}}) - p(\mathbf{s}^i|\mathbf{r}, \hat{\mathbf{h}}) \right) \end{aligned}$$

Hence, whenever $Q(\mathbf{s}^{i+1}, \mathbf{s}^i) \geq Q(\mathbf{s}^i, \mathbf{s}^i)$ (which holds thanks to (10)), $p(\mathbf{s}^{i+1}|\mathbf{r}, \hat{\mathbf{h}}) \geq p(\mathbf{s}^i|\mathbf{r}, \hat{\mathbf{h}})$.

5. EFFICIENT IMPLEMENTATION

This section details a possible implementation of this algorithm showing its linear-time complexity per iteration.

5.1. E-step

First, by noticing that $p(\mathbf{s}|\mathbf{r}, \mathbf{h}, \hat{\mathbf{h}}) = \frac{p(\mathbf{r}|\mathbf{s}, \mathbf{h})p(\mathbf{s})}{p(\mathbf{r}|\hat{\mathbf{h}})}$, the function Q in (9) can further be simplified into:

$$Q(\mathbf{s}, \mathbf{s}^i) = \int \log p(\mathbf{r}|\mathbf{s}, \mathbf{h}) p(\mathbf{h}|\mathbf{r}, \hat{\mathbf{h}}, \mathbf{s}^i) d\mathbf{h} + \log p(\mathbf{s}) - \underbrace{\int \log p(\mathbf{r}|\hat{\mathbf{h}}) p(\mathbf{h}|\mathbf{r}, \hat{\mathbf{h}}, \mathbf{s}^i) d\mathbf{h}}_{\alpha}$$

Since α , the third term of the right-hand side, is independent of \mathbf{s} , it does not influence the maximization in (10). Thus, we focus on the computation of $\int \log p(\mathbf{r}|\mathbf{s}, \mathbf{h}) p(\mathbf{h}|\mathbf{r}, \hat{\mathbf{h}}, \mathbf{s}^i) d\mathbf{h}$.

5.1.1. The pdf $p(\mathbf{h}|\mathbf{r}, \hat{\mathbf{h}}, \mathbf{s}^i)$

We now compute the pdf $p(\mathbf{h}|\mathbf{r}, \hat{\mathbf{h}}, \mathbf{s}^i)$ and skip the exponent i for notation simplicity. Using the independence between $(\mathbf{h}, \hat{\mathbf{h}})$ and \mathbf{s} , and between \mathbf{r} and $\hat{\mathbf{h}}$ given (\mathbf{h}, \mathbf{s}) , the Bayes' rule reads:

$$p(\mathbf{h}|\mathbf{r}, \hat{\mathbf{h}}, \mathbf{s}) = \frac{p(\mathbf{r}|\mathbf{h}, \mathbf{s})p(\mathbf{h}|\hat{\mathbf{h}})}{p(\mathbf{r}|\hat{\mathbf{h}}, \mathbf{s})}. \quad (11)$$

We proceed by computing $p(\mathbf{h}|\hat{\mathbf{h}})$. Using the Bayes' rule and the expressions of the pdfs (2), (5) and (6), we obtain

$$\begin{aligned} p(\mathbf{h}|\hat{\mathbf{h}}) &= \frac{p(\hat{\mathbf{h}}|\mathbf{h})p(\mathbf{h})}{p(\hat{\mathbf{h}})} \\ &= \frac{1}{\det(\pi\mathbf{\Sigma})} \exp(-(\mathbf{h} - \mathbf{u})^\dagger \mathbf{\Sigma}^{-1}(\mathbf{h} - \mathbf{u})) \end{aligned} \quad (12)$$

with $\mathbf{\Sigma}^{-1} = \mathbf{R}_b^{-1} + \mathbf{R}_h^{-1}$ and $\mathbf{u} = \mathbf{\Sigma} \mathbf{R}_b^{-1} \hat{\mathbf{h}}$. The derivation of $p(\mathbf{r}|\hat{\mathbf{h}}, \mathbf{s})$ needs an additional expectation since

$$p(\mathbf{r}|\hat{\mathbf{h}}, \mathbf{s}) = \int p(\mathbf{r}|\mathbf{h}, \mathbf{s}) p(\mathbf{h}|\hat{\mathbf{h}}) d\mathbf{h}$$

Using (4) and (12), we obtain

$$p(\mathbf{r}|\hat{\mathbf{h}}, \mathbf{s}) = \frac{1}{\det(\pi\mathbf{R}_c)} \exp(-(\mathbf{r} - \mathbf{c})^\dagger \mathbf{R}_c^{-1}(\mathbf{r} - \mathbf{c})) \quad (13)$$

with

$$\mathbf{R}_c = \sigma^2 \mathbf{I}_{K+L-1} + \mathbf{S} \mathbf{\Sigma} \mathbf{S}^\dagger \quad (14)$$

$$\mathbf{c} = \mathbf{R}_c \frac{\mathbf{S}}{\sigma^2} \left(\frac{1}{\sigma^2} \mathbf{S}^\dagger \mathbf{S} + \mathbf{\Sigma}^{-1} \right)^{-1} \mathbf{\Sigma}^{-1} \mathbf{u} \quad (15)$$

Finally, by using the Bayes' rule in (11) and the pdfs in (4), (12) and (13), the desired pdf is given by:

$$p(\mathbf{h}|\mathbf{r}, \hat{\mathbf{h}}, \mathbf{s}) = \frac{1}{\det(\pi\mathbf{R}_d)} \exp(-(\mathbf{h} - \mathbf{d})^\dagger \mathbf{R}_d^{-1}(\mathbf{h} - \mathbf{d})) \quad (16)$$

with

$$\mathbf{R}_d^{-1} = \frac{1}{\sigma^2} \mathbf{S}^\dagger \mathbf{S} + \mathbf{\Sigma}^{-1} = \frac{1}{\sigma^2} \mathbf{S}^\dagger \mathbf{S} + \mathbf{R}_b^{-1} + \mathbf{R}_h^{-1} \quad (17)$$

$$\mathbf{d} = \mathbf{R}_d \left(\frac{1}{\sigma^2} \mathbf{S}^\dagger \mathbf{r} + \mathbf{\Sigma}^{-1} \mathbf{u} \right) = \mathbf{R}_d \left(\frac{1}{\sigma^2} \mathbf{S}^\dagger \mathbf{r} + \mathbf{R}_b^{-1} \hat{\mathbf{h}} \right) \quad (18)$$

5.1.2. The expectation of the metric

Using the above derived pdf in (16), we are now able to perform the expectation:

$$\mathbb{E}_{\mathbf{h}|\mathbf{r}, \hat{\mathbf{h}}, \mathbf{s}^i} [|r_k - \mathbf{s}_k^T \mathbf{h}|^2] = |r_k - \mathbf{s}_k^T \mathbf{d}^i|^2 + \mathbf{s}_k^T \mathbf{R}_d^i \mathbf{s}_k^*$$

where \mathbf{d}^i and \mathbf{R}_d^i correspond to (18) and (17), respectively, with the variable \mathbf{s} set to the value \mathbf{s}^i . Finally, the function is given by:

$$Q(\mathbf{s}, \mathbf{s}^i) = -\frac{1}{\sigma^2} \sum_{k=0}^{K+L-2} (|r_k - \mathbf{s}_k^T \mathbf{d}^i|^2 + \mathbf{s}_k^T \mathbf{R}_d^i \mathbf{s}_k^*) - (K+L-1) \log(\pi\sigma^2) + \log p(\mathbf{s}) - \alpha \quad (19)$$

Hence, by noticing that the function $Q(\mathbf{s}, \mathbf{s}^i)$ is quadratic in \mathbf{s} (it consists of an Euclidean distance metric with an extra quadratic term), the maximization of $Q(\mathbf{s}, \mathbf{s}^i)$ over all possible \mathbf{s} can be performed with the Viterbi algorithm.

5.2. M-step

From (19), the maximization step (10) can be performed recursively by using the Viterbi algorithm with branch metric $(|r_k - \mathbf{s}_k^T \mathbf{d}^i|^2 + \mathbf{s}_k^T \mathbf{R}_d^i \mathbf{s}_k^*)$, states $\sigma_k = (s_k, \dots, s_{k-L+2})$, and *a priori* probability (for the symbols) $p(\mathbf{s})$. This *a priori* corresponds to the knowledge available about the sequence \mathbf{s} . There can be either no knowledge (equiprobable sequences), or knowledge about the training sequences (midamble) or partial knowledge about the symbols provided by a decoder in turbo-equalization.

5.3. Summary of the algorithm: an efficient implementation of the algorithm

In this paper, we have proposed to use the EM algorithm in order to solve the exact MAP problem given in (8) and have shown that this algorithm admits a linear-time complexity per iteration implementation:

- **Initialization** with \mathbf{s}^0
- **For each iteration index i**
 1. Compute \mathbf{R}_d^i and \mathbf{d}^i defined in (17) and (18), respectively, with \mathbf{s}^i used instead of \mathbf{s} .
 2. Perform a Viterbi equalizer with branch metric

$$|r_k - \mathbf{s}_k^T \mathbf{d}^i|^2 + \mathbf{s}_k^T \mathbf{R}_d^i \mathbf{s}_k^*$$

and *a priori* probability (for the symbols) $p(\mathbf{s})$. The metric minimizing sequence at the output of the Viterbi equalizer is \mathbf{s}^{i+1} .

Notice that the complexity of the algorithm can be reduced, when K is sufficiently high, by approximating $\mathbf{S}^\dagger \mathbf{S}$ in (17) by $(K + L - 1)\mathbf{I}_L$.

6. SIMULATION RESULTS

We consider the transmission of BPSK symbols over a frequency selective channel. The symbols are assumed to be equiprobable. The channel length is set to $L = 5$. The different channel taps are modeled as independent zero mean complex Gaussian random variables with variance $1/L$. The channel is quasi stationary, i.e. it is time invariant during the transmission of a burst of $K = 512$ information bits and changes independently from burst to burst. The channel estimate $\hat{\mathbf{h}}$ is obtained by a least mean square error estimator using a training sequence of length K_0 [4]. Figures 1 and 2 show the Bit Error Rate (BER) obtained using the EM algorithm from one to four iterations, with respect to the signal to noise ratio (SNR), when the training sequence length is equal respectively to $K_0 = 14$ and $K_0 = 20$. Here, $SNR = E_b/N_0$ where E_b is the energy per transmitted bit and $N_0 = \sigma^2$. Simulations show that in both cases, the performance achieved at the first iteration is roughly equal to the performance obtained when the MAP equalizer assumes that $\mathbf{h} = \hat{\mathbf{h}}$ (curve labelled: 'simple use of channel estimate'). We also notice that the performance of the EM algorithm at the fourth iteration approaches the performance obtained when the equalizer has perfect channel knowledge (dotted curve). Notice that in the case of quasi stationary channels, the criterion derived in [1] cannot be implemented in a practical way since the pdf involved in the derivation of the MAP criterion cannot be factored. The approximate rule given in ([1, eq(7)]) leads to the same performance as the one obtained at the first iteration of our iterative receiver.

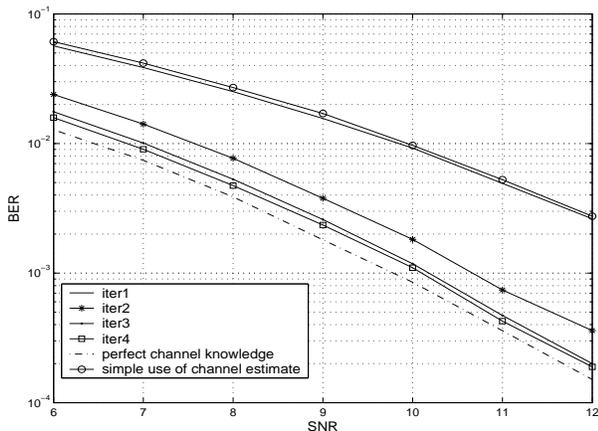


Fig. 1. BER performance for one to four iterations of the EM algorithm when $K_0 = 14$

7. CONCLUSION

In this paper, we considered the problem of MAP equalization when the channel is not perfectly known at the receiver.

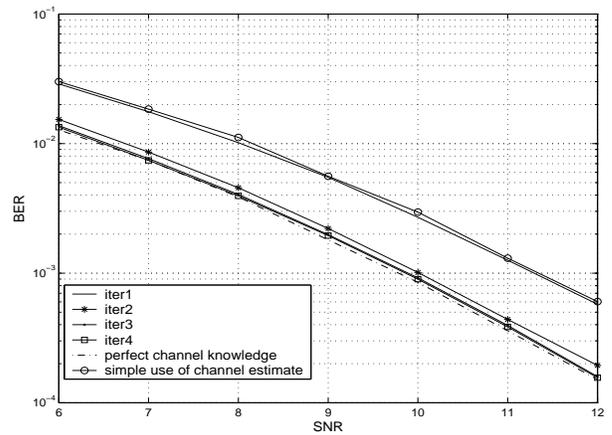


Fig. 2. BER performance for one to four iterations of the EM algorithm when $K_0 = 20$

The MAP criterion was first derived in this case in [1] but no exact practical implementation was proposed. In this paper, we proposed a linear-time complexity per iteration implementation using the EM-algorithm when the channel is quasi stationary. Each M step consisted in a Viterbi algorithm. Simulation results show that after few iterations, our receiver performance attains the performance achieved when the channel is perfectly known. The case of transmissions over time varying channels will be considered in a future work.

8. REFERENCES

- [1] M.Tüchler and M.Mecking, "Equalization for non-ideal channel knowledge," *Conf. on Inf. Sciences and Systems*, The Johns Hopkins University, March 2003.
- [2] M.Siala, "Maximum a Posteriori Decorrelating Discrete-Time Rake Receiver," *Annals of Telecommunications*, vol. 59, n° 3-4, March-April 2004.
- [3] A.P. Dempster, N.M. Laird, and D.B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *Journ. of the Royal Stat. Soc.*, vol. 39, pp. 1-38, 1977.
- [4] S.Crozier, D.Falconer, and S.Mahmoud, "Least sum of squared errors (LSSE) channel estimation," *IEE Proceedings*, vol. 138, pp.371-378, August 1991.
- [5] G.D.Forney, Jr., "Maximum-likelihood sequence estimation for digital sequences in the presence of intersymbol interference," *IEEE Trans. Inf. Theory*, vol. 18, pp. 363-378, May 1972.
- [6] N. Sellami, A. Roumy, and I. Fijalkow, "The impact of both a priori information and channel estimation errors on the MAP equalizer performance," *IEEE Trans. on Signal Process.*, vol. 54, no. 7, pp. 2716-2724, July 2006.
- [7] C. Wu, "On the convergence properties of the EM algorithm," *Annals of Stat.*, vol. 11, no. 1, pp. 95-103, 1983.