

MMSE-based and EM Iterative Channel Estimation Methods

Xavier Wautelet, Cédric Herzet, Antoine Dejonghe, Luc Vandendorpe
Communications and Remote Sensing Laboratory, Université catholique de Louvain
2, place du Levant, B-1348 Louvain-la-Neuve, Belgium
{wautelet,herzet,dejonghe,vandendorpe}@tele.ucl.ac.be

Abstract—This paper compares several iterative channel-and-noise estimation algorithms for a turbo equalizer. Bit-interleaved coded multilevel modulation is considered. The fractionally-spaced equalizer is based on the minimum mean square error (MMSE) criterion.

The effectiveness of a new iterative MMSE-based channel estimation algorithm is shown. Various new or known estimators derived from the expectation-maximization (EM) algorithms are also investigated. All those channel estimators can use either soft or hard symbol estimates fed back by the outer decoder. Simulation results show the gain brought by the refinement of the channel estimates thanks to the iterative process. These results also enable to compare the performance of the various estimators.

I. INTRODUCTION

In a coded system, turbo or iterative equalization has shown its powerfulness to jointly equalize the channel and decode the bits over frequency-selective channels [1]. In order to work properly, the equalizer needs to be fed with estimates of the channel taps and of the noise variance. The customary methods to obtain those values were either data-aided or decision-directed techniques [2]. However, the performance of turbo equalizers is very sensitive to channel-estimation errors. In order to improve the channel state information (CSI) quality, several papers have already been devoted to an iterative refinement of the channel-parameter estimates, e.g. [3]-[12]. Optimally, channel estimation, equalization and decoding should be jointly performed but this usually leads to a receiver with an intractable complexity. Performing those tasks iteratively can be seen as a suboptimal way of jointly achieving those operations.

Several algorithms can be considered to refine the CSI through the iterations. There are the classical hard-decision-directed methods which benefit in this case from the better decision reliability through the iterations. There are also new soft-decision-directed variants of known techniques such as the least-square (LS) or the Kalman estimators [10]. An attractive choice is the expectation-maximization (EM) algorithm [13] which provides a theoretical framework to make sure

of the system convergence. The EM algorithm has already been considered in several papers [4]-[8], [12] to estimate the channel parameters. In these papers, the unknown transmitted symbols are the nuisance parameters which prevent an easy channel estimation. The EM algorithm enables to iteratively refine the channel estimates thanks to symbol a posteriori probabilities. Note that another application of the EM algorithm can be considered to deal with the joint equalization and channel estimation [3], [9]. Since the final goal of the receiver is a low BER, this algorithm can be used to directly estimate the symbols. In this latter case, the channel parameters are the nuisance parameters and the symbol decisions are iteratively refined, computing a posteriori probabilities on the channel taps. Unfortunately, it usually leads to intractable solutions when the channel-impulse-response length is large and/or when the modulation constellation size is high.

In this paper, we investigate a communications system based on bit-interleaved coded modulation (BICM) [14]. The channel is frequency-selective and quasi-static, i.e. the channel taps remain constant over one burst length. The turbo receiver is made up of three main parts. First, the channel estimator: several algorithms are considered for this block. Secondly, the equalizer/demodulator which is based on the minimum mean square error (MMSE) criterion [15]. Thirdly, the classical soft-in soft-out (SISO) BCJR [16] decoder. With appropriate interleaving/deinterleaving, the channel estimator, the equalizer/demodulator and the decoder exchange soft information about the coded bits and hard information about the channel taps and the noise variance.

There are several contributions in this paper. The first new channel estimator is based on the MMSE criterion. The second one is based on a simplification of the EM algorithm called the expectation-conditional-maximization (ECM) algorithm. We also justify the unbiased EM-based channel estimator, which has been proposed in [12]. Furthermore, the bias of the noise variance estimate is calculated. Note that most of the considered channel estimators are fed with soft symbol information but they can also be fed with hard decisions taken on the bits at the decoder output. Eventually, the performance of the new channel-estimation algorithms are compared with those of

Antoine Dejonghe thanks the Belgian NSF for its financial support. This research is partly funded by the Federal Office for Scientific, Technical and Cultural Affairs (OSTC, Belgium) through the IAP contract No P5/11.

known estimators.

The rest of this paper is organized as follows. In section II, the transmitter, the channel model and the receiver are described. In section III, the channel estimators are derived. The simulation results are presented in section IV and the conclusion is given in section V.

II. SYSTEM MODEL

The transmitter scheme is depicted in Fig. 1. Bit-interleaved coded modulation (BICM) is used. The information bits u_a are organized in frames. A frame is encoded by a convolutional encoder, then interleaved by a random permutation. The interleaved bits x_c are grouped and mapped to one of the M possible complex symbols in the considered multilevel/phase constellation. The symbols are zero mean and have variance σ_s^2 . The resulting length- L_s frame of complex symbols s_k ($k = 1, \dots, L_s$) is transmitted over the channel (the symbol s_k is sent at time kT , where T is the symbol period.)

The channel is assumed to be quasi-static, i.e. the channel remains static over one burst length and may vary from one burst to the other. The lowpass equivalent channel impulse response (including the pulse-shaping filter) is denoted by $h(t)$. We assume $h(t)$ can be truncated without loss of accuracy. We only keep its values from time $-L_1T$ to L_2T and $L \triangleq L_1 + L_2 + 1$ is called the channel length. $r(t)$ denotes the received signal and $n(t)$ is the complex envelope of the additive white gaussian noise with two-sided power spectral density (psd) $N_0/2$.

We use fractional sampling at the receiver. Samples are taken at a rate M_s/T after low-pass filtering. For $l = 1 - L_1, \dots, L_s + L_2$ and $m = 0, \dots, M_s - 1$, the received samples $r_{l,m} \triangleq r(lT + mT/M_s)$ are sufficient statistics for symbol detection and channel estimation. We also define the channel taps $h_{l',m} \triangleq h(l'T + mT/M_s)$ for $l' = -L_1, \dots, L_2$ and the noise samples $n_{l,m} \triangleq n(lT + mT/M_s)$. The channel equation is thus:

$$r_{l,m} = \sum_{l'=-L_1}^{L_2} h_{l',m} s_{l-l'} + n_{l,m}. \quad (1)$$

A compact representation may be obtained by stacking up the samples. For the channel-estimation purpose, we only use the received samples for $l = 1 + L_2, \dots, L_s - L_1$. The vectors of the m^{th} polyphase component of the received samples and of the noise samples are respectively \underline{r}_m :

$$\underline{r}_m \triangleq [r_{1+L_2,m} \ r_{2+L_2,m} \ \dots \ r_{L_s-L_1,m}]_{(L_r \times 1)}^T, \quad (2)$$

where $L_r = L_s + 1 - L$ and \underline{n}_m :

$$\underline{n}_m \triangleq [n_{1+L_2,m} \ n_{2+L_2,m} \ \dots \ n_{L_s-L_1,m}]_{(L_r \times 1)}^T. \quad (3)$$

The same way, the vector of the m^{th} polyphase component of the channel impulse response is denoted by

$$\underline{h}_m \triangleq [h_{-L_1,m} \ \dots \ h_{L_2,m}]_{(L \times 1)}^T. \quad (4)$$

The transmitted symbols are placed in a matrix \underline{S} :

$$\underline{S} = \begin{bmatrix} s_L & s_{L-1} & \dots & s_1 \\ s_{L+1} & s_L & \dots & s_2 \\ \vdots & \vdots & \ddots & \vdots \\ s_{k-L_2} & s_{k+1-L_2} & \dots & s_{k+L_1} \\ \vdots & \vdots & \ddots & \vdots \\ s_{L_s} & s_{L_s-1} & \dots & s_{L_s+1-L} \end{bmatrix}_{(L_r \times L)} \quad (5)$$

The observation model then becomes:

$$\underline{r}_m = \underline{S} \underline{h}_m + \underline{n}_m. \quad (6)$$

The presence of an interleaver between the convolutional encoder and the channel input makes it possible to use a turbo or iterative receiver. Without taking into account the channel estimator, the iterative receiver is made up of two main parts. As represented in Fig. 2, the equalizer/demodulator and the decoder are separated by bit (de)interleavers and they exchange soft extrinsic information under the form of bit log-likelihood ratio's (LLR's).

The equalizer/demodulator, the inner SISO stage, has to mitigate intersymbol interference (ISI) and to properly demodulate the symbols. Because of its low complexity, we use the MMSE-based equalizer described in [17]. This equalizer is fractionally spaced, which avoids the cascade of global matched filter and noise whitening filter. Its design relies on an approximation [15] of the optimal MMSE receiver [18]. On the basis of the received samples, of the channel parameter estimates and of the symbol a priori probabilities $P_a(s_k)$ ($k = 1, \dots, L_s$), the equalizer outputs symbol extrinsic probabilities. Then the demodulator computes the bit extrinsic LLR's. According to the turbo principle, the inner SISO module exchanges soft extrinsic information with the outer SISO decoder which is implemented by means of the BCJR algorithm [16].

III. CHANNEL ESTIMATION

The MMSE-based equalizer needs to be fed with estimates of the channel-tap values and of the noise variance. We do not focus here on an adaptive channel estimation algorithm. The channel remains static over the burst length but it can change a lot from one burst to the other; for instance, the following burst is sent over a different frequency. Hence, there is a need for channel acquisition at each burst.

A. ML-Based Estimators

A classical way of doing that is to use a training sequence. Since the channel taps are assumed constant over the burst duration and the pilot symbols are known, it is very easy to perform the channel estimation according to the maximum-likelihood criterion. If

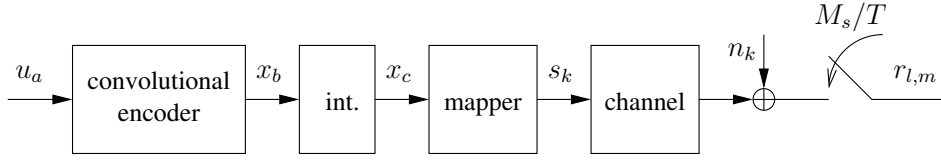


Fig. 1. Transmitter structure

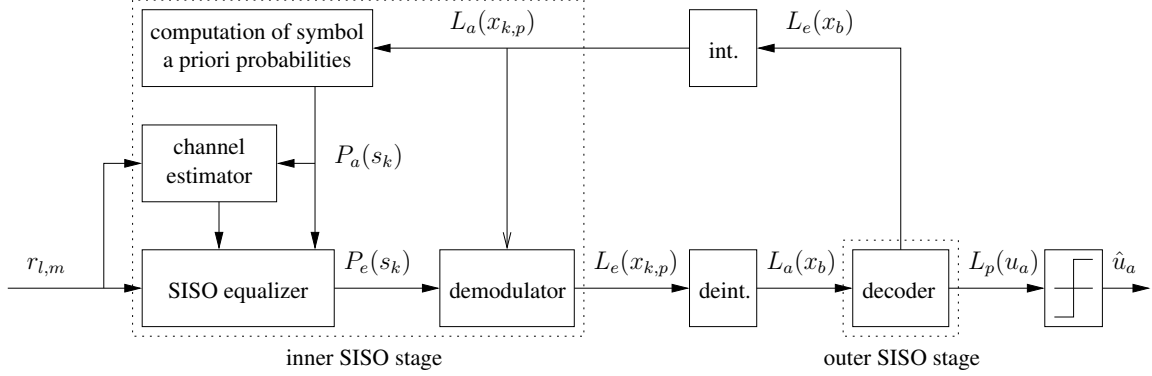


Fig. 2. Receiver scheme

\underline{S}_p contains the transmitted training sequence and $\hat{\underline{h}}_m$ is the estimate vector of the m^{th} polyphase component of the channel impulse response, we get:

$$\hat{\underline{h}}_m = \left(\underline{S}_p^H \underline{S}_p \right)^{-1} \underline{S}_p^H r_m. \quad (7)$$

Note that (7) also corresponds to the least-square estimate of the channel. Finally, it is worth noticing that (7) is the efficient estimator of the channel taps, i.e. this estimator is unbiased and its variance is minimal.

The pilot symbols can be chosen so that the matrix $\underline{S}_p^H \underline{S}_p$ is diagonal [8]. The inversion of this matrix is then very easy. Furthermore, the optimal training sequence consists of such pilots symbols. For more details about the training sequence generation, the readers can refer to [8].

The noise variance can simply be evaluated by measuring the variance of the difference between the received signal r_m and the estimated useful part of it $\underline{S}_p \hat{\underline{h}}_m$:

$$\hat{\sigma}_n^2 = \sum_{m=0}^{M_s-1} \frac{1}{M_s L_r} (r_m - \underline{S}_p \hat{\underline{h}}_m)^H (r_m - \underline{S}_p \hat{\underline{h}}_m). \quad (8)$$

This equation is the ML estimate of the noise variance when the noise variance and the channel taps are jointly estimated. This joint estimation leads to a bias of the noise-variance estimate. This is due to the bias which appears when one try to jointly estimate the mean and the variance of a random variable. More details about this can be found in the appendix.

The performance of turbo equalizers is very sensitive to channel-estimation errors. The data-aided (DA) ML estimation, i.e. the estimation from the training sequence, provides an initial estimate of the channel parameters but it is not reliable enough for the equalizer.

Of course, many pilot symbols can be sent to improve the estimation quality but it implies a huge waste of bandwidth and of power to send those symbols. Rather than that expensive solution, the estimation process can iteratively exploit the information embedded in the whole frame. The simplest idea is a decision-directed (DD) estimator, as in [2]. The equations for that DD estimator are identical to (7):

$$\hat{\underline{h}}_m = \left(\tilde{\underline{S}}^H \tilde{\underline{S}} \right)^{-1} \tilde{\underline{S}}^H r_m \quad (9)$$

and (8):

$$\hat{\sigma}_n^2 = \sum_{m=0}^{M_s-1} \frac{1}{M_s L_r} (r_m - \tilde{\underline{S}} \hat{\underline{h}}_m)^H (r_m - \tilde{\underline{S}} \hat{\underline{h}}_m) \quad (10)$$

but the pilot-symbol matrix \underline{S}_p is replaced by $\tilde{\underline{S}}$, which is filled with the hard decisions on the symbols. Those decisions are computed on the basis of the LLR's at the output of the outer decoder. As shown in Fig. 2, the channel and noise-variance estimator supplies an initial CSI to the equalizer. With appropriate interleaving/deinterleaving, the equalizer/demodulator and the decoder exchange soft information about the coded bits. The channel estimator is fed with the symbol hard decisions built from the decoder output and it can deliver a novel refined CSI to the equalizer.

B. EM-Based Estimators

The iterative DD method is one way to provide a better channel estimate but its convergence is not assured and it is not the theoretical optimal algorithm. As shown in (7) and (8), it is very easy to compute the ML channel estimate if the transmitted symbols are known. However, when the symbols are unknown, the ML function exploiting the code is much more complicated to compute.

The expectation-maximization (EM) algorithm [13] enables to iteratively solve this problem. The EM is usually used to estimate the channel taps only. Here we derive it for the joint estimation of the channel taps and of the noise variance. Let us define the vector of parameters to be estimated \mathcal{B} , which contains the noise variance σ_n^2 and the channel taps \underline{h}_m . Using the EM formalism, the received samples r_m ($m = 0, \dots, M_s - 1$) form the incomplete data set \mathcal{R} , a set from which it is difficult to calculate the ML estimate $\hat{\mathcal{B}}$. Both the received samples and the transmitted symbols \underline{S} form the complete data set \mathcal{Z} , a set from which it is easy to calculate the ML channel-parameter estimates. The symbols \underline{S} are obviously unavailable at the receiver but the EM algorithm provides a framework to iteratively estimate the channel parameters without any a priori information about the symbols. At iteration n , the first step of this method, the expectation step, is the computation of the following function:

$$Q(\tilde{\mathcal{B}}, \hat{\mathcal{B}}^{(n-1)}) = \int_{\mathcal{Z}} p(\mathcal{Z}|\mathcal{R}, \hat{\mathcal{B}}^{(n-1)}) \ln p(\mathcal{Z}|\tilde{\mathcal{B}}) d\mathcal{Z}. \quad (11)$$

This function $Q(\tilde{\mathcal{B}}, \hat{\mathcal{B}}^{(n-1)})$ can be evaluated for any $\tilde{\mathcal{B}}$. The maximization of $Q(\tilde{\mathcal{B}}, \hat{\mathcal{B}}^{(n-1)})$ is the second step of the algorithm:

$$\hat{\mathcal{B}}^{(n)} = \arg \max_{\tilde{\mathcal{B}}} \left\{ Q(\tilde{\mathcal{B}}, \hat{\mathcal{B}}^{(n-1)}) \right\}. \quad (12)$$

It can be shown [19] that the sequence $\mathcal{B}^{(n)}$ converges under fairly general conditions to the ML estimate of \mathcal{B} .

Since the received samples are known, (11) can be rewritten as:

$$Q(\tilde{\mathcal{B}}, \hat{\mathcal{B}}^{(n-1)}) = \int_{\underline{S}} p(\underline{S}|\mathcal{R}, \hat{\mathcal{B}}^{(n-1)}) \ln (p(\mathcal{R}|\underline{S}, \tilde{\mathcal{B}})p(\underline{S}|\tilde{\mathcal{B}})) d\underline{S}. \quad (13)$$

Moreover, \underline{S} is independent of $\tilde{\mathcal{B}}$. Only keeping the terms dependent of $\tilde{\mathcal{B}}$ in $Q(\tilde{\mathcal{B}}, \hat{\mathcal{B}}^{(n-1)})$ yields

$$Q'(\tilde{\mathcal{B}}, \hat{\mathcal{B}}^{(n-1)}) = \int_{\underline{S}} p(\underline{S}|\mathcal{R}, \hat{\mathcal{B}}^{(n-1)}) \ln p(\mathcal{R}|\underline{S}, \tilde{\mathcal{B}}) d\underline{S}, \quad (14)$$

with:

$$\begin{aligned} \ln p(\mathcal{R}|\underline{S}, \tilde{\mathcal{B}}) &= -M_s L_r \ln(\pi \tilde{\sigma}_n^2) \\ &- \frac{1}{\tilde{\sigma}_n^2} \sum_{m=0}^{M_s-1} (r_m - \underline{S} \tilde{h}_m)^H (r_m - \underline{S} \tilde{h}_m). \end{aligned} \quad (15)$$

The maximization step gives the following equation for the channel-tap estimation:

$$\hat{h}_m = E[\underline{S}^H \underline{S}]^{-1} E[\underline{S}]^H r_m, \quad (16)$$

and this expression for the noise-variance estimation:

$$\begin{aligned} \hat{\sigma}_n^2 &= \frac{1}{M_s L_r} \sum_m r_m^H r_m + \hat{h}_m^H E[\underline{S}^H \underline{S}] \hat{h}_m \\ &- 2\text{Re} \left\{ r_m^H E[\underline{S}] \hat{h}_m \right\}. \end{aligned} \quad (17)$$

Fortunately, the joint estimation of the channel parameters does not lead to a coupled (non linear) system between the noise variance and the taps. Indeed, the taps can be estimated before the noise variance. This latter quantity can then be evaluated on the basis of the tap estimates.

The probabilities $p(\underline{S}|\mathcal{R}, \hat{\mathcal{B}}^{(n-1)})$ in (14) are given by the decoder. It has not yet been proved but it is widely believed that after convergence of a turbo receiver, the probabilities at the decoder output are equal to $p(\underline{S}|\mathcal{R}, \hat{\mathcal{B}}^{(n-1)})$. For complexity reasons, it is not possible to reach convergence of the turbo equalizer/decoder at each EM iteration. So, $p(\underline{S}|\mathcal{R}, \hat{\mathcal{B}}^{(n-1)})$ and thus $E[\underline{S}]$ and $E[\underline{S}^H \underline{S}]$, are approximated by taking the decoder output after only one or a few iterations in the equalizer/decoder system.

The iterative refinement of the channel-parameter estimates progresses as follows. At the first iteration, the initial estimation is achieved with the DA ML method, i.e. only with the training sequence. For the next EM iterations, the EM algorithm is applied, using the APP on the symbols delivered by the outer decoder.

It has previously been shown [6] that the EM channel-tap estimator is biased. An EM-based estimator has been proposed in [12] to reduce this bias. Here, we prove that this estimator is actually the unbiased version of the EM algorithm. Indeed, the expectation of the channel tap estimate is:

$$E[\hat{h}_m] = E[\underline{S}^H \underline{S}]^{-1} E[\underline{S}]^H E[r_m]. \quad (18)$$

Since the noise mean is zero and using (6), we get

$$E[\hat{h}_m] = E[\underline{S}^H \underline{S}]^{-1} E[\underline{S}]^H E[\underline{S}] h_m. \quad (19)$$

Unfortunately, the channel-estimate mean is not the true channel-tap value. The bias factor is

$$E[\underline{S}^H \underline{S}]^{-1} E[\underline{S}]^H E[\underline{S}]. \quad (20)$$

For the first iterations, the information about the symbols is bad. During these iterations, $E[\underline{S}]$ is then close to zero, whereas $E[\underline{S}^H \underline{S}]$ keeps the same magnitude order. As a result, the channel-tap estimates tend to zero. To solve this problem, it is possible to unbiased the channel estimates by multiplying the EM estimate by the bias inverse. We then get the unbiased EM solution:

$$\hat{h}_m = (E[\underline{S}^H \underline{S}] E[\underline{S}])^{-1} E[\underline{S}]^H r_m. \quad (21)$$

This is exactly the reduction to the single-input single-output case of the estimator proposed in [12]. It will be referred to as the unbiased EM (UEM) algorithm.

The complexity of the EM-based estimators is mainly due to the inversion of the matrix $E[\underline{S}^H \underline{S}]$ which has to be done once per iteration. An approximated solution with a lower complexity has been proposed in [6]: only the diagonal elements of the matrix $E[\underline{S}^H \underline{S}]$ are kept. The matrix $E[\underline{S}^H \underline{S}]$ is actually nearly diagonal if the frame length is long and if the transmitted symbols are independent. This ‘‘diagonal

approximation” leads to a linear complexity (L) instead of a cubic one (L^3) for the matrix inversion.

Another low-complexity estimator can be designed, using the expectation-conditional-maximization (ECM) algorithm [20]. With the EM algorithm, all the taps corresponding to one polyphase component of the channel impulse response are jointly estimated (16). That leads to M_s linear systems, each with L equations and L unknowns. The basic idea of the ECM algorithm is to compute one tap estimate, assuming the other taps are known. The chosen values for the other taps are the last available estimates of those ones. All those already-computed values can be grouped in a vector $\tilde{\mathbf{h}}_{l,m}^{(n)}$:

$$\tilde{\mathbf{h}}_{l,m}^{(n)} = [\hat{h}_{-L_1,m}^{(n)} \dots \hat{h}_{l-1,m}^{(n)} 0 \hat{h}_{l+1,m}^{(n-1)} \dots \hat{h}_{L_2,m}^{(n-1)}]^T. \quad (22)$$

During one ECM iteration, each tap is estimated turn by turn:

$$\hat{h}_{l,m}^{(n)} = \frac{\{E[\underline{S}]^H \mathbf{r}_m\}_l - \{E[\underline{S}^H \underline{S}] \tilde{\mathbf{h}}_{l,m}^{(n)}\}_l}{\{E[\underline{S}^H \underline{S}]\}_{l,l}}, \quad (23)$$

where $\{A\}_{b,c}$ denotes the element located at the row b and the column c of the matrix A and $\{a\}_b$ denotes the element located at the row b of the vector a . There are $M_s L$ linear equations such as (23), each one with only one unknown.

As the EM algorithm, the ECM algorithm also suffers from a bias. With the same reasoning as the one used for the UEM, the unbiased ECM estimator can be obtained:

$$\hat{h}_{l,m} = \frac{\{E[\underline{S}]^H \mathbf{r}_m\}_l - \{E[\underline{S}^H \underline{S}] \tilde{\mathbf{h}}_{l,m}\}_l}{\{E[\underline{S}]^H E[\underline{S}]\}_{l,l}}. \quad (24)$$

The EM, UEM, ECM and unbiased-ECM algorithms make use of soft information on the symbols. Those methods can also be applied in a DD fashion, i.e. on the basis of hard decisions on the symbols. In this case, the EM and UEM equations are the same as the ones of the DD ML estimator: (9) and (10). The DD version of the unbiased ECM is also equal to the DD ECM. Finally, the “diagonal approximation” can also be applied with the DD ML estimator.

C. MMSE-Based Estimators

The observation-model equation (6) is linear. It is therefore straightforward to derive a channel-tap estimator based on the minimum mean square error (MMSE) criterion. According to the linear-estimation theory [21], this estimate is given by:

$$\hat{h}_{l,m} = E[h_{l,m}] + \underline{w}_{l,m}^H (\mathbf{r}_m - E[\mathbf{r}_m]), \quad (25)$$

where

$$\underline{w}_{l,m} = \text{cov}\{\mathbf{r}_m, \mathbf{r}_m\}^{-1} \text{cov}\{\mathbf{r}_m, h_{l,m}\}. \quad (26)$$

Usually, it is not possible to compute (25) because the receiver does not have any knowledge about $E[h_{l,m}]$. Moreover, the computation of (26) requires

the knowledge of unknown quantities such as the noise variance σ_n^2 , the covariance matrix of the channel taps $\text{cov}\{\underline{h}_m, \underline{h}_m\}$ and the mathematical expectation of the product of two taps $E[\underline{h}_m \underline{h}_m^H]$. In [5], an MMSE channel-tap estimator is used but those quantities are assumed to be known and it is only applied with known pilot symbols. We propose here a method to do MMSE channel estimation without a priori knowledge about the channel statistics. Moreover, this method makes use of the whole information frame and not only of the training sequence.

First, the channel parameters $\hat{\underline{h}}_m^{(1)}$ and $\hat{\sigma}_n^{2(1)}$ are estimated from the pilot symbols using the ML criterion, as in the first iteration of the EM-based algorithms. The DA ML estimator is efficient, so its variance is given by the Cramer Rao bound. If the training sequence is chosen according to [8], it can be shown that the tap-estimate variance $\sigma_{\hat{h}}^2$ corresponding to the Cramer Rao bound is:

$$\sigma_{\hat{h}}^2 = \frac{\sigma_n^2}{(L_p + 1 - L)\sigma_s^2}, \quad (27)$$

where L_p is the training-sequence length. To be accurate, this is true only if the ML estimator makes use of the received samples which are independent on the unknown symbols. There are only $M_s(L_p + 1 - L)$ such received samples, which are only dependent on the pilot symbols.

The channel statistics are unknown at the receiver. Ideally, the conditional a posteriori probabilities $P(h_{l,m} | \hat{\underline{h}}_m^{(1)}, \hat{\sigma}_n^{2(1)})$ should be calculated to use all the a priori and the DA information. Since the a priori information about the channel is unavailable, we propose to approximate the a posteriori means by:

$$E[\underline{h}_m] = \hat{\underline{h}}_m^{(1)}, \quad (28)$$

and the a posteriori covariances by:

$$\text{cov}\{\underline{h}_m, \underline{h}_m\} = \sigma_{\hat{h}}^2 \underline{I}, \quad (29)$$

where \underline{I} is the identity matrix and thus,

$$E[\underline{h}_m \underline{h}_m^H] = E[\underline{h}_m] E[\underline{h}_m]^H + \sigma_{\hat{h}}^2 \underline{I}. \quad (30)$$

Using (25)-(30), the MMSE estimate can be calculated. After some manipulations, we get:

$$\begin{aligned} \text{cov}\{\mathbf{r}_m, \mathbf{r}_m\} &= E[\underline{S} \underline{h}_m \underline{h}_m^H \underline{S}^H] \\ &\quad - E[\underline{S}] E[\underline{h}_m] E[\underline{h}_m]^H E[\underline{S}]^H \\ &\quad + \hat{\sigma}_n^{2(1)} \underline{I} \end{aligned} \quad (31)$$

and

$$\text{cov}\{\mathbf{r}_m, h_{l,m}\} = E[\underline{S}] \underline{e}_l \sigma_{\hat{h}}^2, \quad (32)$$

where the vector \underline{e}_l is filled with zeros except its l^{th} element whose value is 1.

During the iterations, $E[\underline{h}_m]$ and $\sigma_{\hat{h}}^2$ are no longer modified. However, new APP's on the symbols are available at the decoder output and the channel estimation can be refined at each iteration. Regarding the

noise variance, it can be estimated either with the EM algorithm or with the DD ML algorithm. The initial noise-variance estimate $\hat{\sigma}_n^2(1)$ should not be kept for all the iterations. Indeed, the imperfectness of the channel estimation leads to an additional noise at the equalizer output. $\hat{\sigma}_n^2(1)$ does not account for this supplementary noise while it would have to do so [11]. Eventually, a DD MMSE channel estimator can also be considered. In this case, \underline{S} or $E[\underline{S}]$ are replaced by the hard-decision matrix $\underline{\tilde{S}}$ in the equations (25) up to (32).

IV. SIMULATION RESULTS

In this section, we compare the channel-estimation algorithms presented in section III through simulation results. The channel to be estimated is the Porat channel, a length-5 channel with symbol-spaced taps: $[2 - 0.4j, 1.5 + 1.8j, 1, 1.2 - 1.3j, 0.8 + 1.6j]$. Note that the simulated system is not totally realistic. Indeed, the length-5 channel is considered as the equivalent impulse response of the pulse shaping filter, followed by the physical channel and by the receive filter. Moreover, the pulse-shaping-filter bandwidth is assumed to be limited to $1/2T$. Eventually, there is no oversampling at the receiver but the symbol timing is assumed to be perfectly estimated. In a practical system, a fractional receiver could be used to avoid timing estimation. Despite those few unrealistic assumptions, the results are interesting because the goal of this section is only to compare the channel-estimator performance.

All the simulations have been run with the following common parameters. The frame size is 2000 information bits, except for the last figure for which it is equal to 200 information bits. The convolutional encoder always begins and ends a frame at the all-zero state (i.e. trellis termination is performed). The rate-1/2 convolutional encoder has constraint length 5 and octal generator polynomials $[23_8, 35_8]$. The interleaver is totally random and varies from frame to frame to average the effects of good and poor interleavers. The modulation is 8-PSK with Gray mapping. The equalizer filter length is equal to the channel length. Although the Porat channel is used for all the frames, the channel estimation does not use the estimates of the previous frames. A new channel acquisition based on 20 pilot symbols is performed at each frame.

The bit error rate (BER) is measured through Monte Carlo simulations. The results are reported versus the E_b/N_0 ratio. The simulations are stopped after at least 100 frame errors for each E_b/N_0 ratio, for the results to be statistically relevant. All the curves correspond to the BER after 6 EM iterations. It should be mentioned that when pilots are used (all curves but ‘perf’ in Fig. 3), the E_b has not been corrected to account for the fact that the power is spread over pilot and information bits. There is thus a small penalty of 0.13dB for the corresponding curves of Fig. 3 and 4.

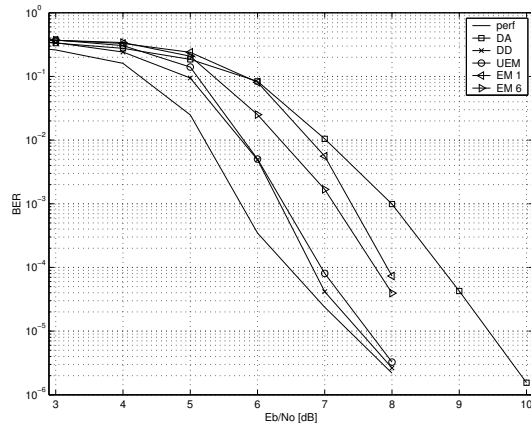


Fig. 3. Comparison of the iterative-channel-estimator performance. perf: with perfect channel knowledge; DA: data aided; DD: decision directed; UEM: unbiased EM; EM 1: one equalizer/decoder iteration per EM iteration; EM 6: 6 equalizer/decoder iterations per EM iteration. 2000 information bits.

Fig. 3 shows the BER for different receivers. Each receiver uses a different channel-estimation algorithm. The curve labelled ‘perf’ corresponds to the performance of the turbo receiver fed with perfect channel knowledge. This curve is a lower bound on the performance of the proposed estimators. The curve labelled ‘DA’ corresponds to the DA estimator. It is only based on the training sequence and therefore is not iterative. At a low BER, the DA-receiver performance is more than 2 dB poorer than the same receiver with perfect parameter knowledge.

Assuming that the probabilities $p(\underline{S}|\mathcal{R}, \hat{\mathcal{B}}^{(n-1)})$ in (14) are given by the decoder after convergence, we should wait until convergence of the equalizer/decoder before doing one more EM iteration. This solution is too complex. For the ‘EM 6’ receiver, we assume the convergence is reached after 6 iterations, so there are 6 equalizer/decoder iterations inside one EM iteration. This ‘EM 6’ receiver performs 1 dB better than the DA ML one at a $\text{BER} \approx 10^{-4}$. All the other iterative algorithms (‘EM 1’, ‘UEM’, etc.) even use a stronger approximation: they only use one equalizer/decoder iteration per EM iteration. However, the ‘EM 1’ performance is only 0.2 dB poorer than the one of ‘EM 6’. Both ‘EM 1’ and ‘EM 6’ algorithms suffer from the bias (20). That explains their weak performance w.r.t. the UEM and the DD algorithms. The UEM receiver performs 1 dB better than the ‘EM 1’. Although the UEM and the DD curves are close to each other, the DD receiver performance is the best one among the receiver with an embedded channel estimator. At a BER below $5 \cdot 10^{-5}$, it performs 0.1 dB better than the UEM and is only 0.1 dB away from the ‘perf’ curve.

The DD ML algorithm is the best one in Fig. 3. Its complexity per iteration evolves as $\mathcal{O}(LM_s L_s + L^3)$ for the channel-tap estimation. In Fig. 4, the DD estimator is compared to its two lower-complexity variants: the DD ML with the “diagonal approxima-

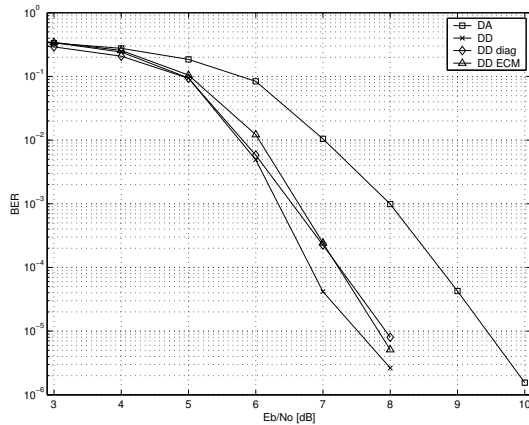


Fig. 4. Comparison of the low-complexity estimators w.r.t the DD and the DA estimators. DD diag: DD with the “diagonal approximation”; DD ECM: hard-decision-directed variant of the expectation-conditional-maximization algorithm. 2000 inform. bits.

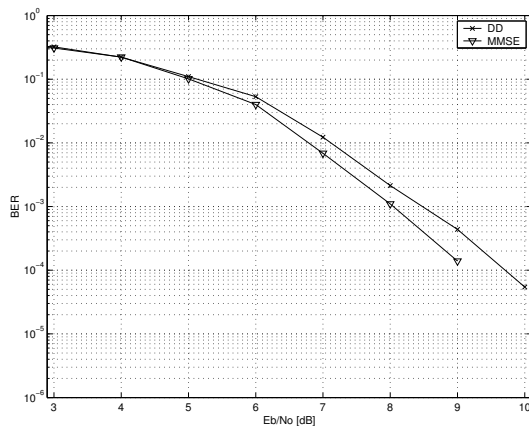


Fig. 5. Comparison of the MMSE and the DD estimators. 200 information bits.

tion” and the DD variant of the ECM. The complexity of both algorithms evolves as $\mathcal{O}(LM_sL_s)$. At a BER $\approx 10^{-5}$, the gaps between the DD ML curve and the ‘DD diag’ and ‘DD ECM’ ones are 0.4 and 0.3 dB respectively. There is thus a small loss due to the approximations brought to lower the complexity. However, at a very small cost, both of the approximated DD solutions perform nearly 1.5 dB better than the non-iterative DA ML method.

Fig. 5 compares the DD ML receiver to the MMSE-based receiver. The disadvantage of the MMSE channel estimator is its high complexity due to the inversions of the M_s matrices $\text{cov}\{\underline{r}_m, \underline{r}_m\}$. The size of the matrix $\text{cov}\{\underline{r}_m, \underline{r}_m\}$ is $L_s \times L_s$. For a simulation-time reason, the frame size in Fig. 5 is 200 information bits instead of 2000. Due to this size reduction, the receiver is less efficient. Indeed, the interleaver size is smaller and the number of data symbols to estimate the channel is smaller too. That explains the 3 dB gap between the DD curves in Fig. 3 and 5. At a BER of 2×10^{-4} , the MMSE curve is 0.5 dB better than the DD one. It seems that the MMSE channel estimator is the best one from the performance point of view but its

complexity is by far higher than the one of the other considered iterative estimators.

V. CONCLUSION

Two new iterative channel-estimation algorithms have been proposed in this paper: the ECM and the MMSE ones. Simulation results show that the turbo-equalization scheme benefits from the iterative refinement of the channel-parameter estimates. All the iterative estimators considered here outperform the classical DA ML method. The performance of several channel-parameter estimators have been compared. It appears that the MMSE one gives very good performance but its complexity is very high. Both UEM and DD receivers perform very closely to the receiver having perfect knowledge of the channel parameters. Both of the low-complexity methods only suffer from a small penalty w.r.t the UEM and the DD receivers. Finally, the EM algorithm is the worst iterative method over the Porat channel, because of the bias of the channel-tap estimate.

Future research could be devoted to the robustness of those algorithms to a varying channel. The estimator performance could also be checked in a realistic system with oversampling.

REFERENCES

- [1] A. Glavieux, C. Laot, J. Labat, “Turbo equalization over a frequency selective channel,” *Proc. Int. Symp. on Turbo codes and related topics*, pp. 96-102, Brest, France, Sep. 1997.
- [2] H. Meyr, M. Moeneclaey, S. A. Fechtel, “Digital communication receivers: synchronization, channel estimation and signal processing,” *Wiley Series in Telecommunications and Signal Processing*, USA, 827 p., 1998.
- [3] C. N. Georghiades, J. C. Han, “Sequence estimation in the presence of random parameters via the EM algorithm,” *IEEE Trans. Commun.*, vol. 45, pp. 300-308, Mar. 1997.
- [4] H. Zamiri-Jafarian, S. Pasupathy, “EM-based recursive estimation of channel parameters,” *IEEE Trans. Commun.*, vol. 47, pp. 1297-1302, Sept. 1999.
- [5] Y. Li, C. N. Georghiades, G. Huang, “Iterative maximum-likelihood sequence estimation for space-time coded systems,” *IEEE Trans. Commun.*, vol. 49, pp. 948-951, June 2001.
- [6] M. Kobayashi, J. Boutros, G. Caire, “Successive interference cancellation with SISO decoding and EM channel estimation,” *IEEE Journ. on Sel. Areas in Comm.*, vol. 19, pp. 1450-1460, Aug. 2001.
- [7] A. O. Berthet, B. S. Ünäl, R. Visoz, “Iterative decoding of convolutionally encoded signals over multipath Rayleigh fading channels,” *IEEE Journ. on Sel. Areas in Comm.*, vol. 19, pp. 1729-1743, Sept. 2001.
- [8] G. Caire, U. Mitra, “Structured multiuser channel estimation for block-synchronous DS/CDMA,” *IEEE Trans. Commun.*, vol. 49, pp. 1605-1617, Sept. 2001.
- [9] B. Lu, X. Wang, Y. G. Li, “Iterative receivers for space-time block-coded OFDM systems in dispersive fading channels,” *IEEE Trans. on Wireless Commun.*, vol. 1, pp. 213-225, Apr. 2002.
- [10] R. Otnes, M. Tüchler, “Soft iterative channel estimation for turbo equalization: comparison of channel estimation algorithms,” in *Proc. Intern. Conf. on Communications Systems*, Singapore, Nov. 2002.
- [11] R. Otnes, M. Tüchler, “On iterative equalization, estimation, and decoding,” in *Proc. Intern. Conf. on Communications*, Anchorage, USA, May 2003.
- [12] X. Wautelet, A. Dejonghe, L. Vandendorpe, “MMSE-based Fractional Turbo Receiver with EM Channel Estimation for Space-Time BICM over MIMO Fading Channels,” in *Proc. 3rd International Symposium on Turbo Codes & Related Topics*, Brest, France, September 1-5, 2003, pp. 519-522.

- [13] A. P. Dempster, N. M. Laird, D. B. Rubin, "Maximum-likelihood from incomplete data via the EM algorithm," *J. Roy. Stat. Soc., Ser. B*, vol. 39, no. 1, pp. 1-38, Jan 1977.
- [14] G. Caire, G. Taricco, E. Biglieri, "Bit-interleaved coded modulation," *IEEE Trans. Inform. Theory*, vol. 44, pp. 927-946, May 1998.
- [15] M. Tüchler, A. Singer, R. Koetter, "Minimum mean squared error equalization using a-priori information," *IEEE Trans. on Signal Processing*, vol. 50, pp. 673-683, Mar. 2002.
- [16] L.R. Bahl, J. Cocke, F. Jelinek, J. Raviv, "Optimal decoding of linear codes for minimizing symbol error rate," *IEEE Trans. Inform. Theory*, vol. 20, pp. 284-287, Mar. 1974.
- [17] X. Wautelet, A. Dejonghe, L. Vandendorpe, "MMSE-based Fractional Turbo Receiver for Space-Time BICM over Frequency-Selective MIMO Fading Channels," *IEEE Trans. on Signal Processing*, to be published.
- [18] X. Wang, H. V. Poor, "Iterative (Turbo) soft interference cancellation and decoding for coded CDMA," *IEEE Trans. Commun.*, vol. 47, pp. 1046-1061, July 1999.
- [19] C. F. J. Wu, "On the convergence properties of the EM algorithm," *Ann. Statistics*, vol. 11, no. 1, pp. 95-103, 1983.
- [20] G. J. McLachlan, T. Krishnan, "The EM algorithm and extensions," *Wiley Series in Probability and Statistics*, USA, 274 p., 1997.
- [21] J. M. Mendel, "Lessons in estimation theory for signal processing, communications, and control," *Prentice Hall Signal Processing Series*, New Jersey, USA, 561 p., 1995.

APPENDIX

The DA ML noise-variance estimate (8) is not biased if the channel taps are known. However, by analogy with the joint estimation of the mean and the variance of a random variable, the ML criterion leads to a biased estimate of σ_n^2 when the noise variance and the channel taps are jointly estimated.

Using (7) and the fact that $(\underline{S}^H \underline{S})^H = \underline{S}^H \underline{S}$, it is straightforward to show that:

$$\begin{aligned} (\underline{r}_m - \underline{S} \underline{h}_m)^H (\underline{r}_m - \underline{S} \underline{h}_m) &= \\ (\underline{r}_m - \underline{S} \hat{\underline{h}}_m)^H (\underline{r}_m - \underline{S} \hat{\underline{h}}_m) &+ \\ + (\underline{S} \hat{\underline{h}}_m - \underline{S} \underline{h}_m)^H (\underline{S} \hat{\underline{h}}_m - \underline{S} \underline{h}_m). \end{aligned} \quad (33)$$

With this relation, the expectation of (8) can be ex-

pressed as:

$$\begin{aligned} E[\hat{\sigma}_n^2] &= \sum_{m=0}^{M_s-1} \left(\frac{E[(\underline{r}_m - \underline{S} \underline{h}_m)^H (\underline{r}_m - \underline{S} \underline{h}_m)]}{M_s L_r} \right. \\ &- \left. \frac{E[(\underline{S} \hat{\underline{h}}_m - \underline{S} \underline{h}_m)^H (\underline{S} \hat{\underline{h}}_m - \underline{S} \underline{h}_m)]}{M_s L_r} \right) \\ &= \frac{M_s L_r \sigma_n^2}{M_s L_r} \\ &- \sum_{m=0}^{M_s-1} \frac{E[(\hat{\underline{h}}_m - \underline{h}_m)^H \underline{S}^H \underline{S} (\hat{\underline{h}}_m - \underline{h}_m)]}{M_s L_r}. \end{aligned} \quad (34)$$

If the pilot symbols are chosen according to [8], we get:

$$\underline{S}^H \underline{S} = L_r \sigma_s^2 \underline{I}. \quad (35)$$

Since the DA ML channel-tap estimator is efficient, the tap-estimate variance is:

$$E[(\hat{h}_{l,m} - h_{l,m})^H (\hat{h}_{l,m} - h_{l,m})] = \frac{\sigma_n^2}{L_r \sigma_s^2}. \quad (36)$$

Using the last two relations, (34) becomes:

$$\begin{aligned} E[\hat{\sigma}_n^2] &= \sigma_n^2 - \frac{M_s L}{M_s L_r} L_r \sigma_s^2 \frac{\sigma_n^2}{L_r \sigma_s^2} \\ &= \frac{L_r - L}{L_r} \sigma_n^2. \end{aligned} \quad (37)$$

The bias shown by (37) becomes negligible when $L_r \gg L$. Usually, this is not the case during the acquisition stage because there are only a few pilot symbols. The bias effect is the underestimation of the noise variance. That can lead to convergence problems in the turbo receiver. To avoid the bias, we can use the following unbiased estimate:

$$\hat{\sigma}_n^2 = \sum_{m=0}^{M_s-1} \frac{1}{M_s (L_r - L)} (\underline{r}_m - \underline{S} \hat{\underline{h}}_m)^H (\underline{r}_m - \underline{S} \hat{\underline{h}}_m). \quad (38)$$

This is the one that is used for the simulations. It has not yet been shown that the EM estimate (17) is biased by the same factor. However, for the simulations, we also multiply by $\frac{L_r}{L_r - L}$ the estimates given by (17) or by (10).