Semi-Blind Channel Estimation Using the EM Algorithm in Iterative MIMO APP Detectors

Mohammad-Ali Khalighi and Joseph Jean Boutros

Abstract— We consider channel estimation in multiple-input multiple-output (MIMO) systems using iterative detection at the receiver. Space-time bit-interleaved coded modulation (BICM) and soft-input soft-output maximum a posteriori (MAP) symbol detection and decoding are considered. Channel coefficients are updated at each iteration of the detector using a semi-blind estimation approach based on the expectation maximization (EM) algorithm. We first show that a "classical" and non-optimized EM implementation, as already proposed in some previous works, gives a biased estimate of the channel coefficients. We then try to optimize the EM implementation and propose a modification to it that provides an unbiased channel estimate and leads to a better convergence of the iterative detector. We show that considerable improvement in the receiver performance can be obtained by using our proposed modified unbiased (MU) EM algorithm, especially for large number of transmit antennas and short training sequences. We also show that when MIMO signal detection is strongly asymmetric in the sense of too few receive antennas, the EM-based channel estimation may be of little interest. Moreover, we consider a simple semi-blind estimation scheme, based on hard decisions on reliable decoded data bits, and compare its performance with the EM based estimation methods.

Index Terms— MIMO systems, channel estimation, semi-blind estimation, training sequences, expectation maximization, bit-interleaved coded modulation, iterative detection.

I. INTRODUCTION

THE potential of multiple-input multiple-output (MIMO) systems in providing huge spectral efficiencies has been revealed recently [2], [3], [4]. In a rich scattering propagation medium, by using multiple antennas at both sides of the communication link, very high data transmission rates can be achieved, provided that appropriate transmit/receive schemes are employed. Hence, MIMO systems are among the most promising solutions to the ever-growing need for very high data rates. When coherent signal detection is to be performed at the receiver, channel state information (CSI) is required, for which a channel estimation step is necessary. Channel estimation plays a critical role in the performance of the receiver. Most current systems use a training-based channel estimation scheme in the form of time-multiplexed pilot symbols. This scheme is usually referred to as pilot symbol assisted modulation (PSAM) [5].

For finite length blocks, if channel estimation is to be done on each block of symbols, the insertion of pilot symbols can result in a considerable reduction of the achievable data rate. This loss in data rate becomes important specially at low signal-to-noise ratios (SNR) and when the channel undergoes rapid variations [6].[1] Thus, we are interested in minimizing the number of pilot symbols in a frame. However, by reducing the number of pilot symbols, the channel may be learned improperly and channel estimation errors may become important. This can result in a considerable performance degradation and the need to retransmit data. The performance degradation can be compensated by smart signal processing at the receiver. In fact, instead of traditional channel estimation methods based on pilot symbols only, we can use *semi-blind* channel estimation approaches that, in addition to pilot symbols, make use of data symbols in channel estimation. In this way, a considerable performance improvement can be achieved at the price of increased receiver complexity [7], [8].

In this paper, we consider a semi-blind channel estimation scheme based on the expectation maximization (EM) algorithm [9], [10]. The MIMO system we consider uses iterative detection at the receiver and the channel estimation part is inserted in the turbo-detector, a scheme already considered in [11], [12]. We use the bit-interleaved coded modulation (BICM) scheme for space-time coding [13], [14]. The advantage of the BICM is its flexibility regarding the choice of the code and the bit-symbol mapping, as well as its conformity to iterative detection [15]. The EM algorithm based on the maximum-likelihood (ML) criterion is used to update the channel coefficients at each iteration of the turbo-detector. At the first iteration, a primary channel estimate is obtained based on the pilot sequences only, that allows the EM algorithm, used in the succeeding iterations, to bootstrap.

First, we consider the "classical" EM implementation, as considered in [11], [12], that will be called mixing EM (Mix-EM). Next, we show that Mix-EM provides a biased estimate of channel coefficients. We show that the origin of this bias is the residual co-antenna interference (CAI) after MIMO detection. Inspired by the idea of [16] for the case of code-division multiple-access (CDMA) multi-user detection, we then propose a modification to Mix-EM that allows to obtain an unbiased estimate while minimizing the estimation errors' variance. We call this approach modified-unbiased EM (MU-

[1]Notice that, given the finite number of unknown channel parameters, the channel capacity is the same for known and unknown CSI at reception since capacity is defined for asymptotically large frame lengths.
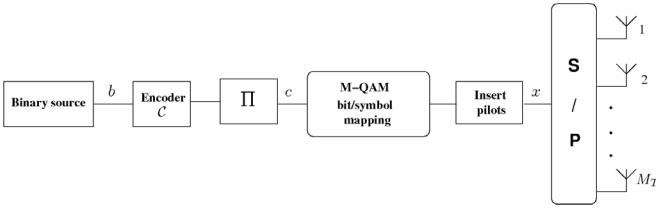
Fig. 1.   Block diagram of the BICM transmission scheme.

Fig. 2.   Block diagram of the receiver.

EM). We show that the improvement achieved by MU-EM, as compared to Mix-EM, depends on the channel diversity order and the number of pilot symbols in a frame. The interesting result is that when we have a small amount of receive diversity available in the underlying MIMO channel, i.e., when signal detection is strongly asymmetric, EM-based channel estimation becomes practically of no interest, even with the proposed modification of MU-EM.

For the sake of comparison, we will also consider a simple semi-blind estimation scheme for which, at each iteration, only the hard decisions of those symbols detected with high enough reliability are taken into account in channel estimation. We call this approach, already proposed in some works such as [17], *thresholded hard-decision* (Th-HD). We show that the performance of the Th-HD method depends highly on the choice of the threshold P_{TH} that determines whether or not the decoder soft-outputs are reliable enough. The practical limitation is that the optimum threshold value depends on the MIMO configuration, i.e., the number of transmit and receive antennas, as well as on the actual SNR.

The paper is organized as follows. System model and assumptions are presented in Section II. Iterative signal detection is described in Section III. Next, we consider the channel estimation problem in Section IV. We first present Mix-EM and show that it provides a biased estimate, and then present our modification of MU-EM. We also present and describe the Th-HD approach. Simulation results are presented in Section V to demonstrate and compare the performances achieved via different estimation schemes. Finally, we draw some conclusions in Section VI.

II. ASSUMPTIONS AND SYSTEM MODEL

Consider a MIMO structure with M_T transmit and M_R receive antennas. The communication channel is assumed to be frequency non-selective. The channel matrix H of dimension $(M_R \times M_T)$ is assumed to be constant over the duration of a symbol frame (i.e., quasi-static MIMO channel). The BICM coding scheme used at the transmitter is described in Fig.1. The binary data b are encoded by a non-recursive non-systematic convolutional (NRNSC) code $\mathcal{C}$, before being interleaved (the block Π). Quasi-random interleaving is used and no optimization [18] is done on the interleaver design. The output bits c are then transformed to symbols according to a given constellation set. We add some pilot symbols to each frame of data symbols for channel estimation. The symbols x of a frame are then multiplexed before being transmitted through M_T antennas. Since a quasi-static and flat channel is assumed, the placement of pilot sequences in a frame has
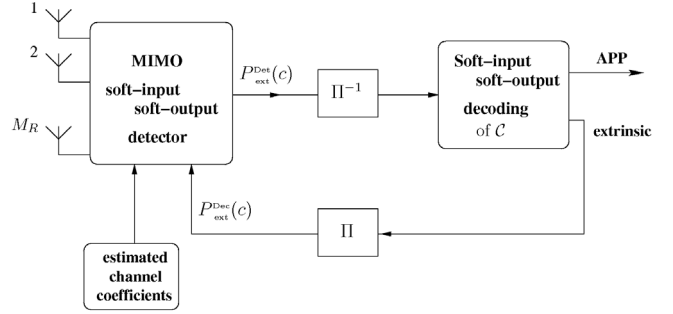
no effect on channel estimation [19], thus no optimization is required for their placement and they are put in a preamble.

QPSK modulation is considered in this paper, and transmit symbol power is normalized, i.e., $|x|^2 = 1$. The following notations are used throughout the paper: bold face upper case letters are used for matrices; bold face lower case letters are used for vectors; $.^*$, $.^H$, $|.|$, $\|.\|$, and E$\{.\}$ denote respectively complex conjugate, conjugate transpose, absolute value, vector Frobenius norm, and expected value. I is the Identity matrix and $diag(.)$ denotes a diagonal matrix with the given diagonal terms.

Considering a frame of transmitted symbols associated with the channel matrix H, the received vector y of dimension $(M_R \times 1)$ at a given time index is given by:

$$y = Hx + n \qquad (1)$$

where n is the vector of additive complex white Gaussian noise of zero mean and variance N_0. We call the vector x of M_T symbols a *compound* symbol. We denote by N_{ps} the number of compound pilot symbols in a frame, corresponding to $N_p = N_{ps}BM_T$ pilot bits, B being the number of bits per symbol. We also denote by N_{ds} the number of compound data symbols in a frame, corresponding to $N_d = N_{ds}BM_T$ data bits. Finally, $N_s = N_{ds} + N_{ps}$ denotes the total number of compound symbols in a frame.

III. ITERATIVE MIMO DETECTION AND DECODING

At the receiver, the detection of transmitted symbols and the decoding task are performed in an iterative manner. Soft-input soft-output (SISO) signal detection and channel decoding are performed, based on maximizing the *a posteriori* probabilities (APP). The block diagram of the receiver is shown in Fig.2. SISO decoding is performed using the well known forward-backward algorithm [20]. We just recall here the formulation of the APP MIMO detector.

Our formulations are based on probabilities instead of likelihood ratios; this is more appropriate when using the EM algorithm. Suppose first that the channel matrix H and the complex noise variance N_0 are perfectly known at the receiver. The MIMO detector provides at its output extrinsic probabilities on coded bits c. Let $\mathcal{Q}$ be the cardinality of x of size $q \triangleq |\mathcal{Q}| = 2^{BM_T}$. Let also c_i, $i = 1, \cdots, BM_T$ be the bits corresponding to a compound symbol $x \in \mathcal{Q}$. The

extrinsic probability on the bit c_j at the MIMO detector output, $P_{\text{ext}}^{\text{Det}}(c_j)$, is calculated as follows [18]:

$$P_{\text{ext}}^{\text{Det}}(c_j = 1) = K \sum_{\substack{\boldsymbol{x} \in \mathcal{Q} \\ c_j = 1}} \exp\left(-\frac{\|\boldsymbol{y} - \boldsymbol{H}\boldsymbol{x}\|^2}{N_0}\right) \prod_{\substack{i=1 \\ i \neq j}}^{BM_T} P_{\text{ext}}^{\text{Dec}}(c_i) \tag{2}$$

where K is the normalization factor satisfying $P_{\text{ext}}^{\text{Det}}(c_j = 1) + P_{\text{ext}}^{\text{Det}}(c_j = 0) = 1$. Subscript $._{\text{ext}}$ denotes extrinsic information and the superscripts $.^{\text{Det}}$ and $.^{\text{Dec}}$ denote the information at the output of the MIMO detector and the channel decoder, respectively. $P_{\text{ext}}^{\text{Dec}}(c_i)$ is in fact the *a priori* information on bit c_i, coming from the SISO decoder. The summation in (2) is taken over the product of the conditional channel likelihood (the $\exp$ term) given a compound symbol $\boldsymbol{x}$, and the *a priori* probability on this symbol (the term $\prod P_{\text{ext}}^{\text{Dec}}$), fed back from the SISO decoder at the previous iteration. Notice that concerning this latter term, we exclude the *a priori* probability corresponding to the bit c_j itself, so as to respect the exchange of *extrinsic* information between the channel decoder and the MIMO detector. Also, notice that this term assumes independent coded bits c_i, which is true for random interleaving of large size. At the first iteration, where no *a priori* information is available on bits c_i, $P_{\text{ext}}^{\text{Dec}}(c_i)$ are set to $1/2$.

IV. CHANNEL ESTIMATION

Obviously, for signal detection we require the channel matrix $\boldsymbol{H}$ as well as the complex noise variance N_0. To estimate $\boldsymbol{H}$, we transmit some pilot symbols in each frame of transmitted data. As a frequency non-selective channel is considered, the identifiability condition is: $N_{ps} \geq M_T$ [21]. We first consider in Subsection IV-A the estimation of the channel based uniquely on pilot sequences, and next consider in Subsection IV-B semi-blind channel estimation based on the EM algorithm.

A. Pilot-only-based estimation

Let $\boldsymbol{x}_p(k)$ denote a compound pilot symbol, i.e., a vector of M_T pilot symbols, at the time sample k. The ML channel estimate $\hat{\boldsymbol{H}}$, which is equivalent to the least-squares solution, is:

$$\hat{\boldsymbol{H}} = \left(\sum_{k=1}^{N_{ps}} \boldsymbol{y}(k)\,\boldsymbol{x}_p^H(k)\right)\left(\sum_{k=1}^{N_{ps}} \boldsymbol{x}_p(k)\,\boldsymbol{x}_p^H(k)\right)^{-1}. \tag{3}$$

Let $\boldsymbol{R}_{\boldsymbol{x}_p} = \sum_{k=1}^{N_{ps}} \boldsymbol{x}_p(k)\boldsymbol{x}_p^H(k)$. The minimum least-squares estimation error is obtained iff the training sequences satisfy $\boldsymbol{R}_{\boldsymbol{x}_p} = N_{ps}\boldsymbol{I}$. When it is not possible to find such optimal sequences, the near-optimal sequences consist of those resulting in minimum $trace\{\boldsymbol{R}_{\boldsymbol{x}_p}^{-1}\}$ [21].

To estimate N_0, a simple way is to suppose $\hat{\boldsymbol{H}} \approx \boldsymbol{H}$ and to use the ML-based estimation:

$$\hat{N}_0 = \frac{1}{M_R N_{ps}}\left(\sum_{k=1}^{N_{ps}} \left\|\boldsymbol{y}(k) - \hat{\boldsymbol{H}}\,\boldsymbol{x}_p(k)\right\|^2\right). \tag{4}$$

However, since $\hat{\boldsymbol{H}}$ contains estimation errors, $\hat{N}_0$ is biased. We propose the following modified estimate that is unbiased for mutually orthogonal pilot vectors (see [22] for the proof):[2]

$$\hat{N}_0^{new} = \frac{N_{ps}}{M_R(N_{ps} - M_T)^2}\left(\sum_{k=1}^{N_{ps}} \left\|\boldsymbol{y}(k) - \hat{\boldsymbol{H}}\,\boldsymbol{x}_p(k)\right\|^2\right). \tag{5}$$

B. EM-based channel estimation

Let $\boldsymbol{\Theta} = [\boldsymbol{H}, N_0]$, the vector of parameters to be estimated. The ML estimate of $\boldsymbol{\Theta}$, given the vector of observations $\boldsymbol{Y}$, is:

$$\hat{\boldsymbol{\Theta}}^{ML} = \arg\ \max_{\boldsymbol{\Theta}}\ \log p(\boldsymbol{Y}|\boldsymbol{\Theta}) \tag{6}$$

The EM approach consists in estimating $\boldsymbol{\Theta}$ from (6), given the *incomplete* data $\boldsymbol{Y}$ and a vector $\boldsymbol{X}$, called *missing* data. The ensemble $(\boldsymbol{Y}, \boldsymbol{X})$ is called the *complete* data. In our channel estimation problem, $\boldsymbol{Y}$ is the vector of received signals on M_R antennas during N_s time samples, and $\boldsymbol{X}$ is the vector of transmitted compound symbols over this time interval. The EM update algorithm consists of two steps: computing the expected log-likelihood function of the complete data conditioned on incomplete data and the previous estimate of parameters (*E-step*), and then, maximizing the result with respect to the parameters (*M-step*). E- and M- steps for the $(m + 1)$th update are described below by (7) and (8), respectively.

$$Q\left(\boldsymbol{\Theta}, \hat{\boldsymbol{\Theta}}^{(m)}\right) = \mathrm{E}\left\{\log P(\boldsymbol{Y}, \boldsymbol{X}|\boldsymbol{\Theta})\,\big|\boldsymbol{Y}, \hat{\boldsymbol{\Theta}}^{(m)}\right\} \tag{7}$$

$$\hat{\boldsymbol{\Theta}}^{(m+1)} = \arg\ \max_{\boldsymbol{\Theta}}\ Q\left(\boldsymbol{\Theta}, \hat{\boldsymbol{\Theta}}^{(m)}\right) \tag{8}$$

1) Classical EM: Mixing data and pilots: Let us denote by $\boldsymbol{X}_d$ and $\boldsymbol{X}_p$ the transmitted compound data and pilot symbols in a frame, respectively, and by $\boldsymbol{Y}_d$ and $\boldsymbol{Y}_p$ the corresponding received vectors. To take into account pilot and data symbols in EM, we can consider $\{\boldsymbol{Y}_d, \boldsymbol{Y}_p\}$ as incomplete data and $\{\boldsymbol{X}_d, \boldsymbol{X}_p\}$ as missing data, as it is done in [11]. We call this approach Mix-EM. In this way, after some manipulations, from (7) and (8) we obtain the following update equations for the estimates of $\boldsymbol{H}$ and N_0 at $(m + 1)$th step [11].

$$\hat{\boldsymbol{H}}^{(m+1)} = \left(\sum_{k=1}^{N_s} \sum_{u=1}^{q} \boldsymbol{y}(k)\,\boldsymbol{x}_u^H\,APP_k(\boldsymbol{x}_u|\boldsymbol{\Theta}^{(m)})\right) \times$$
$$\left(\sum_{k=1}^{N_s} \sum_{u=1}^{q} \boldsymbol{x}_u\,\boldsymbol{x}_u^H\,APP_k(\boldsymbol{x}_u|\boldsymbol{\Theta}^{(m)})\right)^{-1} \tag{9}$$

$$\hat{N}_0^{(m+1)} = \frac{1}{M_R N_s}\left(\sum_{k=1}^{N_s} \sum_{u=1}^{q} APP_k(\boldsymbol{x}_u|\boldsymbol{\Theta}^{(m)}) \times \right.$$
$$\left. \left\|\boldsymbol{y}(k) - \hat{\boldsymbol{H}}^{(m+1)}\boldsymbol{x}_u\right\|^2\right) \tag{10}$$

---

[2]This estimate can further be refined by averaging over subsequent frames. Notice from (5) that for $N_{ps} \gg M_T$, we find the primary estimate (4). However, for relatively small N_{ps}, this modification is indispensable since the corresponding bias from (4) is too large. A relatively large biased noise variance estimate can degrade considerably the turbo-detector performance [23].

As it is seen, we make use of the information on transmitted symbols, obtained from SISO decoder, to re-estimate the channel by including data symbols in the estimation. Vector $\boldsymbol{x}_u$ is the uth among q possibly transmitted compound symbols, whose probability of transmission $APP_k(\boldsymbol{x}_u|\boldsymbol{\Theta}^{(m)})$, i.e., the probability that $\boldsymbol{x}(k) = \boldsymbol{x}_u$, is calculated using the *a posteriori* probabilities at the SISO decoder output at the end of the mth iteration:

$$APP_k(\boldsymbol{x}_u|\boldsymbol{\Theta}^{(m)}) \propto \prod_{i=1}^{BM_T} P_{post}^{\text{Dec}}(c_{u,i}) \qquad (11)$$

$P_{post}^{\text{Dec}}(c_{u,i})$ is the *a posteriori* probability corresponding to the ith bit of $\boldsymbol{x}_u$, $c_{u,i}$. For pilot symbols, which are known at receiver, $APP_k(\boldsymbol{x}_u|\boldsymbol{\Theta}^{(m)})$ equals either one or zero. Before being used in (9) and (10), the calculated APPs from (11) should first be normalized so as to satisfy $\sum_{u=1}^{q} APP_k(\boldsymbol{x}_u|\boldsymbol{\Theta}^{(m)}) = 1$.

Notice that in this way we obtain only an approximation of the exact APPs. In fact, $P_{post}^{\text{Dec}}(c_{u,i})$ consists of the probability of $c_{u,i}$ given the corresponding received symbol. This is related to the probability of $\boldsymbol{y}$ conditioned to $c_{u,i}$, through the Bayes theorem (see [11], [24] for more details). But the received signals $\boldsymbol{y}$ are correlated when partially conditioned. Our approach is thus sub-optimal in terms of the convergence speed and bit-error-rate, but it is much simpler than the exact calculation of APPs. However, we have compared our results to those of [11] that considers the exact APP calculation, and noticed that the performance difference is quite negligible.

2) Mix-EM: biased estimate: In what follows, we demonstrate that the estimate $\hat{\boldsymbol{H}}$ calculated from (9) is biased. To simplify our notations, we denote $APP_k(\boldsymbol{x}_u|\boldsymbol{\Theta}^{(m)})$ by $APP_{u,k}$ and do not specify the estimation update step. We also define $\overline{\boldsymbol{R}}_x$ and $\overline{\boldsymbol{R}}_{yx}$ as follows.

$$\overline{\boldsymbol{R}}_x = \sum_{k=1}^{N_s} \sum_{u=1}^{q} \boldsymbol{x}_u \boldsymbol{x}_u^H APP_{u,k} \qquad (12)$$

$$\overline{\boldsymbol{R}}_{yx} = \sum_{k=1}^{N_s} \sum_{u=1}^{q} \boldsymbol{y}(k) \boldsymbol{x}_u^H APP_{u,k} \qquad (13)$$

Equation (9) can now be written in the following form:

$$\hat{\boldsymbol{H}} = \overline{\boldsymbol{R}}_{yx} \overline{\boldsymbol{R}}_x^{-1} \qquad (14)$$

We define $\tilde{\boldsymbol{x}}(k) \triangleq \sum_{u=1}^{q} \boldsymbol{x}_u APP_{u,k}$ as the estimate of transmitted compound symbol $\boldsymbol{x}(k)$ based on the SISO decoder output information. Note that for $k = 1, \cdots, N_{ps}$, we have $\tilde{\boldsymbol{x}}(k) = \boldsymbol{x}_p(k)$. We can now write $\overline{\boldsymbol{R}}_x$ and $\overline{\boldsymbol{R}}_{yx}$ in (12) and (13) in terms of $\tilde{\boldsymbol{x}}$:[3]

$$\overline{\boldsymbol{R}}_{x,i,j} = \begin{cases} N_s & ; \quad i = j \\ \sum_{k=1}^{N_s} \tilde{\boldsymbol{x}}_i(k) \tilde{\boldsymbol{x}}_j^*(k) & ; \quad i \neq j \end{cases} \qquad (15)$$

[3]To obtain (15), we have used the approximation of (11), explained in the previous subsection.

$$\begin{aligned} \overline{\boldsymbol{R}}_{yx} &= \sum_{k=1}^{N_s} \boldsymbol{y}(k) \tilde{\boldsymbol{x}}^H(k) \\ &= \boldsymbol{H} \sum_{k=1}^{N_s} \boldsymbol{x}(k)\tilde{\boldsymbol{x}}^H(k) + \sum_{k=1}^{N_s} \boldsymbol{n}(k)\tilde{\boldsymbol{x}}^H(k) \\ &= \boldsymbol{H} \boldsymbol{R}_x' + \boldsymbol{\eta} \end{aligned} \qquad (16)$$

where, for instance, $\tilde{\boldsymbol{x}}_i(k)$ is the ith entry of the vector $\tilde{\boldsymbol{x}}(k)$, and,

$$\boldsymbol{R}_x' \triangleq \sum_{k=1}^{N_s} \boldsymbol{x}(k) \tilde{\boldsymbol{x}}^H(k). \qquad (17)$$

Also, $\boldsymbol{\eta}$ is the matrix of weighted noise samples with $\text{E}\{\boldsymbol{\eta}^H\boldsymbol{\eta}\} = N_0\boldsymbol{R}_x''$ where,

$$\boldsymbol{R}_x'' = \sum_{k=1}^{N_s} \tilde{\boldsymbol{x}}(k) \tilde{\boldsymbol{x}}^H(k). \qquad (18)$$

Now from (14), the estimated channel matrix can be written as:

$$\hat{\boldsymbol{H}} = \boldsymbol{H} \boldsymbol{R}_x' \overline{\boldsymbol{R}}_x^{-1} + \boldsymbol{\eta} \overline{\boldsymbol{R}}_x^{-1} \qquad (19)$$

As it can be seen from (15) and (17), $\boldsymbol{R}_x' \neq \overline{\boldsymbol{R}}_x$. So, from (19), $\hat{\boldsymbol{H}}$ is a biased estimate of $\boldsymbol{H}$. Observing (15) and (17) we see that this bias comes from the difference between $\boldsymbol{x}$ and $\tilde{\boldsymbol{x}}$, or in other words, from noise and the residual CAI after MIMO detection. For low SNR and at the first iterations of the turbo-detector, $\tilde{\boldsymbol{x}}$ is close to zero, the difference between $\boldsymbol{R}_x'$ and $\overline{\boldsymbol{R}}_x$ is important, and the channel coefficients are highly under-estimated. This is specially the case when the reception diversity is small. The residual CAI after MIMO detection is also due to the channel estimation errors $\tilde{\boldsymbol{H}}$. As a result, the bias will be more important for smaller number of pilot symbols. Notice that although the estimation is improved through iterations, the important bias at the critical (first) iterations prevents the turbo-detector to converge correctly. On the other hand, when the SNR is high enough, at the last iterations, $\tilde{\boldsymbol{x}} \approx \boldsymbol{x}$, and hence $\boldsymbol{R}_x' \approx \overline{\boldsymbol{R}}_x$, and the bias becomes negligible. For SNR$\to \infty$, Mix-EM becomes asymptotically unbiased.

3) Modifying Mix-EM: MU-EM: We separate the estimates based on pilot and data symbols and denote them by $\hat{\boldsymbol{H}}^p$ and $\hat{\boldsymbol{H}}^d$, respectively. Remember that $\hat{\boldsymbol{H}}^p$ is calculated from (3) and is unbiased. In addition, $\hat{\boldsymbol{H}}^d$ is calculated from (9) by replacing N_s by N_{ds}, or from (14) by replacing N_s by N_{ds} in the corresponding expressions (15)-(18). We propose a modification to Mix-EM that remains simple and can result in a considerable improvement in channel estimation. The increased computational complexity is quite justified, as we will see later. The core of our proposed modification is some simplifying approximations that we make before proceeding to the combination of $\hat{\boldsymbol{H}}^d$ and $\hat{\boldsymbol{H}}^p$. We assume that N_{ds} is large enough and approximate $\overline{\boldsymbol{R}}_x$, $\boldsymbol{R}_x'$, and $\boldsymbol{R}_x''$ by diagonal matrices. Indeed, the off-diagonal terms can be considered as empirical correlations between uncorrelated sequences and could be neglected. In fact, assuming large enough N_{ds} and uncorrelated transmit symbols, these correlations come from the residual CAI after MIMO detection. When enough receive diversity is available, this residual CAI is low enough and the

approximations work well. We consider hence:

$$\overline{\boldsymbol{R}}_x \approx N_{ds}\boldsymbol{I} \qquad (20)$$

$$\boldsymbol{R}'_x \approx diag \left(\sum_{k=1}^{N_{ds}} \boldsymbol{x}_1(k)\, \tilde{\boldsymbol{x}}_1^*(k)\,, \cdots, \sum_{k=1}^{N_{ds}} \boldsymbol{x}_{M_T}(k)\, \tilde{\boldsymbol{x}}_{M_T}^*(k) \right) \qquad (21)$$

$$\boldsymbol{R}''_x \approx diag \left(\sum_{k=1}^{N_{ds}} |\tilde{\boldsymbol{x}}_1(k)|^2\,, \cdots, \sum_{k=1}^{N_{ds}} |\tilde{\boldsymbol{x}}_{M_T}(k)|^2 \right) \qquad (22)$$

Using these approximations, from (19) we obtain:

$$\hat{\boldsymbol{H}}^d = \frac{1}{N_{ds}}\boldsymbol{H}\, diag \left(\sum_{k=1}^{N_{ds}} \boldsymbol{x}_1(k)\, \tilde{\boldsymbol{x}}_1^*(k), \cdots, \sum_{k=1}^{N_{ds}} \boldsymbol{x}_{M_T}(k)\, \tilde{\boldsymbol{x}}_{M_T}^*(k) \right) + \boldsymbol{\eta}' \qquad (23)$$

with $\boldsymbol{\eta}' = \frac{1}{N_{ds}}\boldsymbol{\eta}$. The (i,j)th entry of $\hat{\boldsymbol{H}}^d$ is then,

$$\hat{\boldsymbol{H}}_{ij}^d = \alpha_j \boldsymbol{H}_{ij} + \boldsymbol{\eta}'_{ij} \qquad (24)$$

where,

$$\alpha_j \triangleq \frac{1}{N_{ds}} \sum_{k=1}^{N_{ds}} \boldsymbol{x}_j(k)\, \tilde{\boldsymbol{x}}_j^*(k). \qquad (25)$$

Notice that α_j depends only on j, the index of the transmit antenna. In other words, $\hat{\boldsymbol{H}}_j^d$ (the jth column of $\hat{\boldsymbol{H}}^d$) is related to $\boldsymbol{H}_j$ (the jth column of $\boldsymbol{H}$) by a scalar factor α_j. Mathematically, it follows from the fact that from (23), $\boldsymbol{H}$ is multiplied by a diagonal matrix. This, in turn, is due to the approximations made in (20)-(22), thanks to which $\hat{\boldsymbol{H}}^d$ and $\boldsymbol{H}$ are tied column by column. The bias removing operation is hence much simpler than from the matrix multiplication expression of (19). The physical interpretation is that, as previously explained, by the diagonality assumptions on $\overline{\boldsymbol{R}}_x$, $\boldsymbol{R}'_x$, and $\boldsymbol{R}''_x$, we have in fact assumed perfect CAI cancellation after MIMO detection. So, the estimated $\hat{\boldsymbol{H}}_{ij}^d$, $i = 1,.., M_R$ would depend on the transmit symbols from the jth antenna only, and $\hat{\boldsymbol{H}}_{ij}^d$ is related to $\boldsymbol{H}_{ij}$ regardless of the receive antenna index i.

Now let us consider $\hat{\boldsymbol{H}}_j$ as the combination of $\hat{\boldsymbol{H}}_j^d$ and $\hat{\boldsymbol{H}}_j^p$ in the following form:

$$\hat{\boldsymbol{H}}_j = a_j\hat{\boldsymbol{H}}_j^d + b_j\hat{\boldsymbol{H}}_j^p \,. \qquad (26)$$

We have two degrees of freedom, i.e., the coefficients a_j and b_j, that we determine so as to minimize the estimation errors' variance, subject to the constraint of unbiased estimation:

$$\min \; \mathrm{E}\left\{ \|a_j\boldsymbol{\eta}'_j + b_j\boldsymbol{\eta}_j^p\|^2 \right\} \quad \text{subject to:} \;\; a_j\alpha_j + b_j = 1. \quad (27)$$

$\boldsymbol{\eta}^p$ is the matrix of estimation errors from (3). For mutually orthogonal pilot sequences, the variance of $\boldsymbol{\eta}_{ij}^p$ equals N_0/N_{ps}. Also, the variance of $\boldsymbol{\eta}'_{ij}$, $\sigma^2_{\boldsymbol{\eta}'_{ij}}$, is:

$$\sigma^2_{\boldsymbol{\eta}'_{ij}} = \frac{N_0}{N_{ds}}\beta_j^2 \;\; \text{with} \;\; \beta_j^2 = \frac{1}{N_{ds}}\boldsymbol{R}''_{x_{jj}} = \frac{1}{N_{ds}} \sum_{k=1}^{N_{ds}} |\tilde{\boldsymbol{x}}_j(k)|^2. \qquad (28)$$

Using the method of Lagrange multipliers, we can obtain the optimal coefficients a_j and b_j:

$$a_j = \frac{\alpha_j^*}{|\alpha_j|^2 + \frac{N_{ps}}{N_{ds}}\beta_j^2} \;\;, \;\; b_j = \frac{\frac{N_{ps}}{N_{ds}}\beta_j^2}{|\alpha_j|^2 + \frac{N_{ps}}{N_{ds}}\beta_j^2} \,. \qquad (29)$$

The only remaining problem is the calculation of α_j from (25), since we do not know the transmit symbols $\boldsymbol{x}(k)$. However, if we consider ε the average bit error probability on <u>coded bits</u>, we can calculate an estimate of α_j (see [1], [22]).[4] However, we have verified that we can take $\varepsilon = 0$ without any considerable effect on the obtained performance. This is equivalent to replace $\boldsymbol{x}_j$ in (25) by hard decisions on $\tilde{\boldsymbol{x}}_j$ (whatever the modulation).

• **Discussion on the proposed combination**

Let us verify our proposed combination in two extreme cases. Consider first the case of low SNR and the first iterations of the detector. In this case, the *a posteriori* probabilities at the decoder output are close to 1/2 and the estimated symbols $\tilde{\boldsymbol{x}}$ are close to zero. So, since we have little information on data symbols, the estimation should rely mostly on pilot symbols, i.e., we should have $\hat{\boldsymbol{H}} \approx \hat{\boldsymbol{H}}^p$. This is what our combination gives; here we have $\alpha_j \to 0$, and hence, from (29) we obtain $a_j \to 0$ and $b_j \to 1$. Now consider the case of high SNR and the concluding iterations of the detector. Here the decoder soft outputs are quite reliable and are close to one or zero. Hence, $\tilde{\boldsymbol{x}} \to \boldsymbol{x}$ and with this almost perfect knowledge on data symbols, $\hat{\boldsymbol{H}}^d$ and $\hat{\boldsymbol{H}}^p$ should have "equal weight" in the final estimate $\hat{\boldsymbol{H}}$. This is verified in the proposed combination; we have $\alpha_j \to 1$, and therefore, we obtain $a_j \to N_{ds}/N_s$ and $b_j \to N_{ps}/N_s$.

C. Th-HD estimation

As explained in the introduction, we consider this relatively simple semi-blind estimation approach to compare its performance with EM-based methods. In the Th-HD method, in addition to pilot symbols, we use the symbols detected with high enough reliability at each iteration. This is done by comparing the APPs to a threshold $0.5 < P_{TH} < 1$. Consider the probability $APP_i^{(m)}$ at iteration m, corresponding to the coded bit c_i. If $APP_i^{(m)} > P_{TH}$, we make the hard decision $\hat{c}_i^{(m+1)} = 1$; otherwise, if $APP_i^{(m)} < (1 - P_{TH})$, we make the hard decision $\hat{c}_i^{(m+1)} = 0$; and if none of these conditions are verified, we give up the compound symbol corresponding to c_i and do not consider it in channel estimation. If a hard decision is made on all BM_T constituting bits of a compound symbol, we use the resulting hard-detected compound symbol in channel estimation, in the same way as pilot symbols, i.e., using (3). The resulting channel estimate is then used in the next iteration of the detector. Note that if we take P_{TH} very close to 0.5, we effectively make hard decisions on all detected symbols and use them in channel estimation. This coincides with the so called decision-feedback channel estimation [15].

V. SIMULATION RESULTS

In this section we present a comparative study of the different estimation schemes described in Section IV. Since we consider the quasi-static channel model, the performance comparison is made in terms of frame error rate (FER) versus

---

[4]We can constitute a look-up table of ε for some values of SNR via simulations for the case of perfect channel knowledge, or better, for the case of pilot-only-based channel estimation.
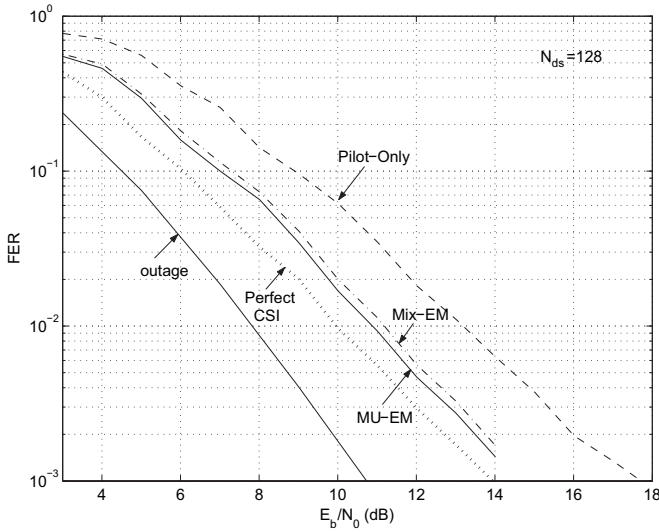
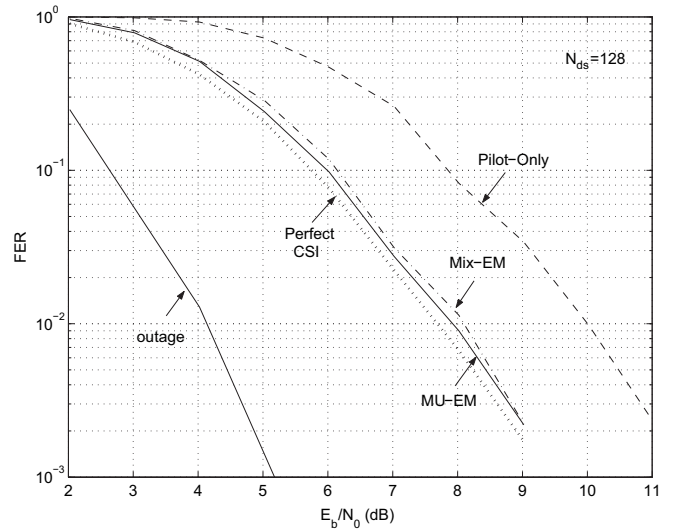Fig. 3. (2×2) MIMO channel, NRNSC code $(133,171)_8$, $N_{ps} = 4$ pilot compound symbols.

Fig. 4. (4×4) MIMO channel, NRNSC code $(5,7)_8$, $N_{ps} = 8$ pilot compound symbols.

E_b/N_0. E_b is the average received energy per *information* bit. Notice that we take into account the receiver array gain in our definition of E_b but exclude from it the pilot bits. Rate 1/2 NRNSC channel codes are considered. Since we compare FERs, the same $N_{ds} = 128$ is considered in all systems. Also, N_{ps} is chosen to be a little greater than the limit of channel identifiability, BM_T. Unless otherwise mentioned, the FER curves correspond to the fifth iteration of the turbo-detector. Rayleigh fading conditions are considered and the entries of $\boldsymbol{H}$ are considered to be IID and of unit variance.

A. Performance improvement by MU-EM

In this section we compare the receiver performance for different channel estimation methods, i.e., pilot-only based estimation, Mix-EM, and MU-EM. For the sake of comparison, we also present FER curves for the case of perfect channel knowledge, as well as the outage probability for QPSK modulation [25]. The other curve, labeled Th-HD, will be discussed in the next subsection.

Fig.3 shows performance curves for a two-transmit two-receive antenna structure, denoted by (2×2). We take $N_{ps} = 4$ that means that only 4/64=6.25% of the information data rate is dedicated to the transmission of pilot symbols. The difference between the slopes of FER curves and the outage probability curve is due to the limited interleaver size. Remember that the interleaver is of random type and is not optimized. We see that by using Mix-EM, a gain of 2 dB in SNR is achieved at FER=10^{-2}, as compared to the pilot-only-based estimation. The additional gain by using MU-EM is about 0.3 dB only.

We may expect that for an increased number of transmit antennas, the difference between Mix-EM and MU-EM performances is more considerable, since the CAI would be more important. We have presented in Fig.4 performance curves for a (4×4) system with $N_{ps} = 8$ (8/64=12.5% of data rate dedicated to pilots). A gain of 2 dB is again achieved at FER=10^{-2} by performing Mix-EM, as compared to pilot-only estimation. However, the difference between Mix-EM

and MU-EM is only about 0.18 dB which is even smaller than in (2×2) system. Also, surprisingly, we observe that the EM-based methods perform closely to the perfect CSI case. In fact, the increased diversity in (4×4) channel helps to improve the performance of the MIMO detector, and the residual CAI after detection will actually be lower, as compared to the (2×2) case. As a result, Mix-EM works well and the improved performance by MU-EM is less important.

This difference between Mix-EM and MU-EM performances becomes more considerable for shorter pilot sequences. This is shown in Fig.5 where $N_{ps} = 5$. The other parameters are the same as in Fig.4. We see that the gain obtained by MU-EM is now about 0.4 dB at FER=10^{-2}.

As we explained in Subsection IV-B.2, this gain is more important for a smaller number of iterations, where the residual CAI is more important. This can be seen from Fig.6 that considers the same conditions of Fig.5 but shows the FER after three iterations of the turbo-detector (instead of five). We see that the improvement by MU-EM is more considerable and is about 0.8 dB now.[5]

Let us go back to Fig.4. To see the impact of receive diversity on the EM-based estimation, we reduce the number of antennas to $M_R = 2$, while keeping $M_T = 4$. Results are shown in Fig.7. First, we see that Mix-EM performs worse and the resulting performance is comparable to that of pilot-only method. Second, the gain obtained by MU-EM is much more important and is about 1.3 dB at FER=10^{-2}. Third, we observe a more important difference between the FERs with perfect CSI and by MU-EM estimation, that is about 0.7 dB. In fact, for (4×2) system where we have less receive diversity as compared to (4×4) case, the residual CAI after MIMO detection is more important, and as a result, the APPs at the decoder output will be of less reliability and/or *contrast*. This shows a more important bias for Mix-EM, and thus, a more important gain is achieved using MU-EM. However, a

---

[5]Note that for more than three iterations, we obtain a negligible improvement in the detector performance for the cases of perfect-CSI and pilot-only-based estimation.
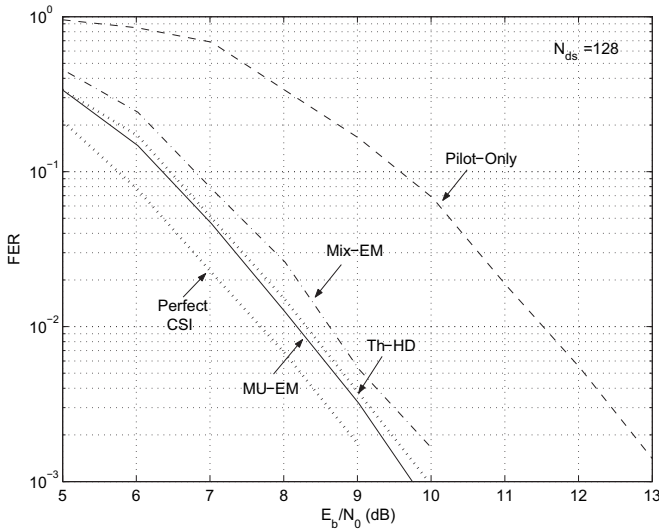
Fig. 5. (4×4) MIMO channel, NRNSC code $(5,7)_8$, $N_{ps} = 5$ pilot compound symbols.
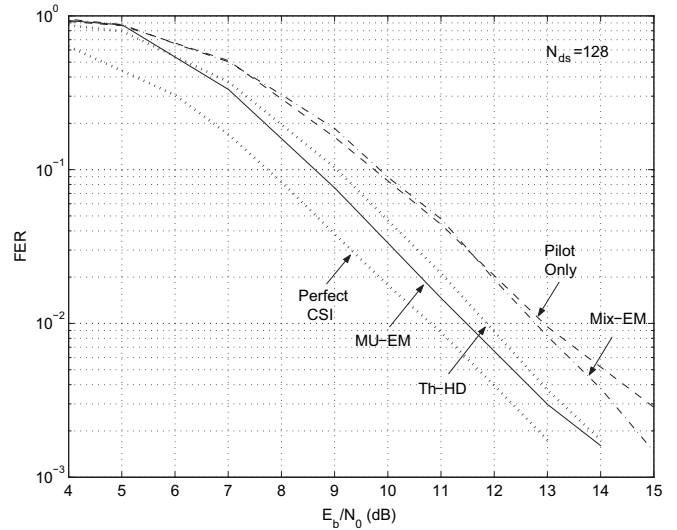

Fig. 7. (4×2) MIMO channel, NRNSC code $(5,7)_8$, $N_{ps} = 8$ pilot compound symbols.
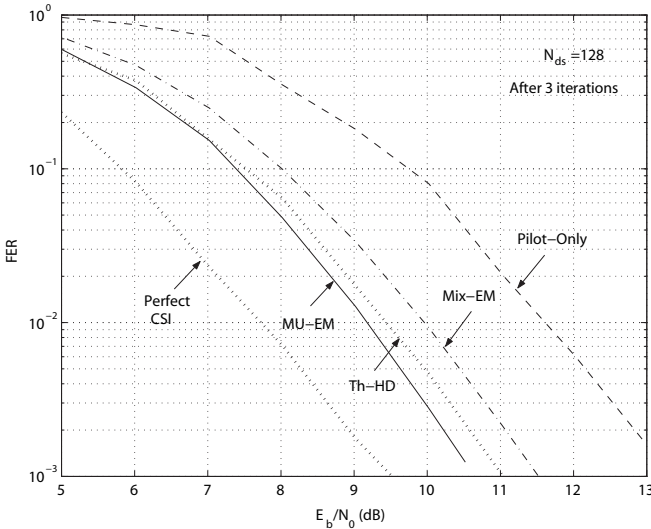

Fig. 6. (4×4) MIMO channel, NRNSC code $(5,7)_8$, $N_{ps} = 5$ pilot compound symbols, after 3 iterations.
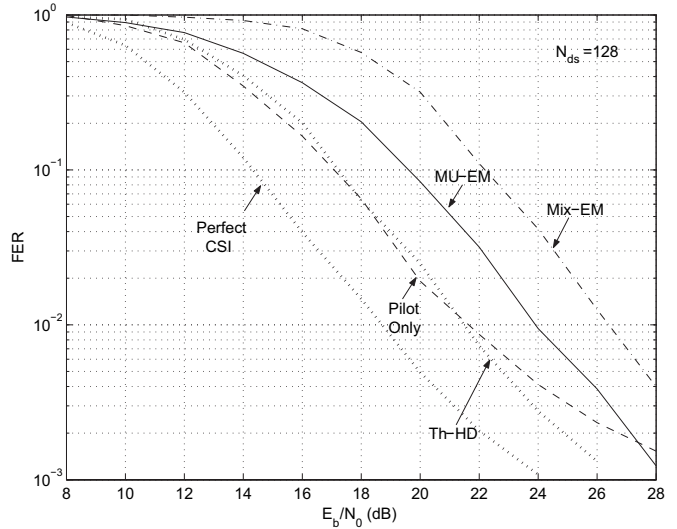

Fig. 8. (4×1) MISO channel, NRNSC code $(5,7)_8$, $N_{ps} = 8$ pilot compound symbols.

more important residual CAI affects MU-EM too, since it partially falsifies the basic simplifying assumptions used in the derivation of MU-EM expressions (see Subsection IV-B.3). Hence, MU-EM is less efficient than in (4×4) case, and the difference between its performance and that of perfect CSI is more important.

We have also considered the case of $M_T = 4$ and a single antenna at receiver, $M_R = 1$. Results, shown in Fig.8, are very interesting. For a (4×1) system and the range of SNRs considered, the residual CAI is important at the detector output due to the lack of receive diversity. This results in a bad estimate via Mix-EM and this can even result in the divergence of the turbo-detector. The performance of MU-EM is not satisfying either, although it performs better than Mix-EM. This is again because the assumptions on which MU-EM relies are not met, due to the important residual CAI. Here, the divergence of the turbo-detector occurs less frequently than for

Mix-EM.[6] Surprisingly, the pilot-only estimation has the best performance. So, for far less receive antennas than transmit antennas, the EM based channel estimation may actually be of no interest.

B. Performance of Th-HD method

Let us first study the impact of the APP threshold P_{TH} on the receiver performance. We have shown in Fig.9 curves of FER versus P_{TH} for the MIMO systems of (4×4), (4×2), and (4×1). The same parameters of the previous subsection are considered. For each system, E_b/N_0 is chosen so as to achieve an FER of about 10^{-2} after five iterations of the turbo-detector for the best P_{TH}. We see that the optimal P_{TH} resulting in the lowest FER depends on the underlying MIMO structure.

---

[6]For a given channel realization, the estimate via Mix-EM is worse and the probability that the estimate results in the divergence of the turbo-detector is higher. For more explanations, see [26].
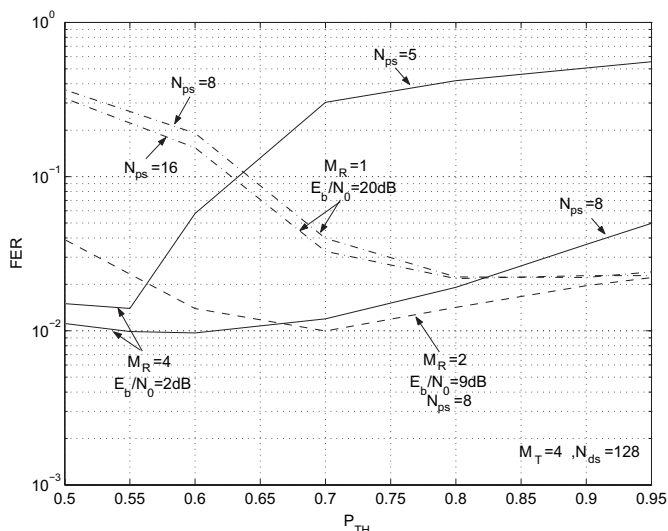
Fig. 9.   Impact of APP threshold P_{TH} on the performance of Th-HD estimation error, $0.501 \leq P_{TH} \leq 0.95$, NRNSC code $(5,7)_8$.

However, it has poor dependence on the number of pilots. When receive diversity is important, as in the (4×4) case, the optimal P_{TH} is close to 0.5 meaning that almost all APPs at the decoder output are reliable enough and their hard decisions coincide most likely with the corresponding transmitted bits. However, for lower receive diversity (smaller M_R), APPs at the decoder output are less reliable and the threshold should be placed higher. We then notice that the optimal P_{TH} depends on SNR: the higher the SNR is, the better the detection will be, and the more reliable will the APPs at the decoder output be, thus P_{TH} should be set lower.

To compare it with other estimation methods, we have shown in Figures 5, 6, 7, and 8, curves of FER for Th-HD, taking into account the optimal P_{TH} values from Fig.9, that is, 0.55, 0.7, and 0.9, respectively for (4×4), (4×2), and (4×1) systems.[7] From Fig.8 we see that for a (4×1) system that suffers from the lack of receive diversity, Th-HD gives a better performance than EM-based methods; yet, its performance is almost the same as the pilot-only method, so it is practically of no interest. For a (4×2) system, (Fig.7) Th-HD shows a considerable improvement, as compared to pilot-only and Mix-EM; yet, the best performance is achieved by MU-EM. For the (4×4) system, Th-HD gives a considerable improvement as compared to pilot-only, but the best performance is still obtained by MU-EM. For $N_{ps} = 8$, the performance obtained by Th-HD is almost equal to that of Mix-EM and the corresponding curve is not shown in Fig.4. For $N_{ps} = 5$ (Figures 5 and 6), Th-HD performs even better than Mix-EM.

VI. CONCLUSIONS

We considered semi-blind channel estimation at each iteration of an APP turbo-detector using the EM algorithm. Thanks to some pilot symbols that allow an initialization of the turbo-detector, we can obtain some *partial* information on the transmitted data symbols using the soft values at the decoder

output. Transmit symbols can be estimated using this partial information and the channel estimate can then be renewed using these symbols, together with pilot symbols. Naturally, a channel estimate based on this partial information can not be perfect. This imperfection not only appears in the form of an increased estimation errors' variance, but even, as we showed, in a bias in the resulting estimate $\hat{\boldsymbol{H}}^d$. The classical Mix-EM approach ignores this bias and simply combines data and pilot symbols to obtain the channel estimate. We proposed to consider separately $\hat{\boldsymbol{H}}^d$ and the pilot-only-based estimate $\hat{\boldsymbol{H}}^p$, and to make use of $\hat{\boldsymbol{H}}^p$ to correct the bias in $\hat{\boldsymbol{H}}^d$, while minimizing the resulting estimation errors' variance. The exact solution for such a combination is not straight forward and is computationally complex. In our proposed MU-EM approach, we made some simplifying assumptions that led us to consider a simple combination of $\hat{\boldsymbol{H}}^d$ and $\hat{\boldsymbol{H}}^p$ while adding negligible complexity to the initial algorithm. We saw that, as far as these assumptions are more or less verified, MU-EM works very well. We also studied the simple Th-HD semi-blind estimation scheme and made a comparative study of the different estimation methods for different MIMO structures. Based on the provided results and discussions, we can deduce the following concluding remarks.

• For $M_T \leq M_R$ where we have enough receive diversity available, semi-blind methods attain closely the ideal performance, i.e., with perfect channel knowledge. Relatively few pilot symbols are required to bootstrap these estimation schemes. Among these semi-blind methods, MU-EM gives the best performance, i.e., it leads to the best convergence of the iterative detector. The improvement obtained, compared to Mix-EM, is more considerable for fewer pilot symbols, as well as for a smaller number of iterations.

• For $M_T > M_R$, i.e., the case of asymmetric MIMO signal detection, the improvement achieved via MU-EM, as compared to Mix-EM, is quite considerable. In this case, the bias in the channel estimate via Mix-EM is important and makes it practically of no interest. For too few receive antennas, the performance of MU-EM is not satisfying either, although it still works much better than Mix-EM. The reason is that the assumptions we made in the derivation of the MU-EM formulation are not met. Here, the pilot-only approach gives the best performance. As a matter of fact, if we cannot obtain reliable enough information on data symbols, there is logically no interest to use a semi-blind approach. This is the case for too small M_R with respect to M_T, where symbol detection cannot be performed efficiently enough.

• The simple Th-HD method gives good performances, specially when enough receive diversity is available. Under this condition, Th-HD may be preferred to EM-based methods and in any case to Mix-EM, due to its lower computational complexity. However, the performance of Th-HD is very sensitive to the choice of the threshold P_{TH}. The optimal P_{TH} depends highly on the underlying MIMO structure as well as on the SNR, and should be chosen carefully.

• The EM algorithm is usually thought to be too computationally complex for implementation in a real system. However, it is well adapted to iterative receivers since the part of the complexity due to EM can be considered to be small to

---

[7]As explained, the optimal P_{TH} depends on SNR. The optimal values considered for each system correspond to the SNR leading to FER≈ 10^{-2} after five iterations of the detector.

moderate, as compared to that of the turbo-detector. Hence, in the cases where EM-based methods and in particular MU-EM work well, the increased complexity is quite justified given the improvement obtained in the receiver performance.

REFERENCES

[1] M. A. Khalighi and J. Boutros, "Modified unbiased EM-based channel estimation for MIMO turbo receivers," in *Proc. IEEE Symposium on Signal Processing and Information Technology (ISSPIT) 2004*, pp. 127–131,.

[2] E. Telatar, "Capacity of multi-antenna Gaussian channels," *European Trans. Telecommun.*, vol. 10, no. 6, pp. 585–595, Nov.-Dec. 1999.

[3] G. J. Foschini and M. J. Gans, "On limits of wireless communications in a fading environment when using multiple antennas," *Wireless Personal Commun.*, vol. 6, no. 3, pp. 311–335, Mar. 1998.

[4] M. A. Khalighi, K. Raoof, and G. Jourdain, "Capacity of wireless communication systems employing antenna arrays, a tutorial study," *Wireless Personal Commun.*, vol. 23, no. 3, pp. 321–352, Dec. 2002.

[5] J. K. Cavers, "An analysis of pilot symbol assisted modulation for Rayleigh fading channels," *IEEE Trans. Veh. Technol.*, vol. 40, no. 4, pp. 686–693, Nov. 1991.

[6] B. Hassibi and B. M. Hochwald, "How much training is needed in multiple-antenna wireless links?" *IEEE Trans. Inf. Theory*, vol. 49, no. 4, pp. 951–963, Apr. 2003.

[7] E. De Carvalho and D. T. M. Slock, "Cramer-Rao bounds for semi-blind, blind, and training antenna arrays, a tutorial study," in *Proc. International Workshop on Signal Processing Advances in Wireless Communications (SPAWC) 1997*, pp. 129–132.

[8] G. B. Giannakis, Y. Hua, P. Stoica, and L. Tong, *Signal Processing Advances in Wireless and Mobile Communications*. Upper Saddle River, NJ: Prentice Hall PTR, 2001, vol. 1, Trends in Channel Estimation and Equalization.

[9] A. Dempster, N. Laird, and D. Rubin, "Maximum-likelihood from incomplete data via the EM algorithm," *J. Royal Statistical Society*, vol. 39, no. 1, pp. 1–38, Jan. 1977.

[10] T. K. Moon, "The expectation-maximization algorithm," *IEEE Signal Processing Mag.*, vol. 13, no. 6, pp. 47–60, Nov. 1996.

[11] J. Boutros, F. Boixadera, and C. Lamy, "Bit-interleaved coded modulations for multiple-input multiple-output channels," in *Proc. IEEE International Symposium on Spread Spectrum Techniques and Applications (ISSSTA) 2000*, vol. 1, pp. 123–126.

[12] A. O. Berthet, B. S. Unal, and R. Visoz, "Iterative decoding of convolutionally encoded signals over multipath Rayleigh fading channels," *IEEE J. Sel. Areas Commun.*, vol. 19, no. 9, pp. 1729–1743, Sept. 2001.

[13] E. Zehavi, "8-PSK trellis codes for a Rayleigh channel," *IEEE Trans. Commun.*, vol. 40, no. 3, pp. 873–884, May 1992.

[14] G. Caire, G. Taricco, and E. Biglieri, "Bit-interleaved coded modulation," *IEEE Trans. Inf. Theory*, vol. 44, no. 3, pp. 927–946, May 1998.

[15] R. Visoz and A. O. Berthet, "Iterative decoding and channel estimation for space-time BICM over MIMO block fading multipath AWGN channel," *IEEE Trans. Commun.*, vol. 51, no. 8, pp. 1358–1367, Aug. 2003.

[16] M. Kobayashi, J. Boutros, and G. Caire, "Successive interference cancellation with SISO decoding and EM channel estimation," *IEEE J. Sel. Areas Commun.*, vol. 19, no. 8, pp. 1425–1428, Aug. 2001.

[17] M. Sellathurai and S. Haykin, "Turbo-BLAST for wireless communications: theory and experiments," *IEEE Trans. Signal Processing*, vol. 50, no. 10, pp. 2538–2546, Oct. 2002.

[18] J. Boutros, N. Gresset, and L. Brunel, "Turbo coding and decoding for multiple antenna channels," in *Proc. International Symposium on Turbo Codes and Related Topics 2003*.

[19] M. Dong and L. Tong, "Optimal design and placement of pilot symbols for channel estimation," *IEEE Trans. Signal Processing*, vol. 50, no. 12, pp. 3055–3069, Dec. 2002.

[20] L. R. Bahl, J. Cocke, F. Jelinek, and J. Raviv, "Optimal decoding of linear codes for minimizing symbol error rate," *IEEE Trans. Inf. Theory*, vol. 20, no. 2, pp. 284–287, Mar. 1974.

[21] J. Balakrishnan, M. Rupp, and H. Viswanathan, "Optimal channel training for multiple antenna systems," *Conference on Multiacess, Mobility and Teletraffic for Wireless Communications 2000*.

[22] M. A. Khalighi and J. Boutros, "Channel estimation in turbo-BLAST detectors using EM algorithm," in *Proc. International Conference on Acoustics, Speech and Signal Processing (ICASSP) 2005*, vol. 3, pp. 1037–1040.

[23] M. A. Khalighi, "The effect of mismatched SNR on the performance of Log-MAP turbo-detector," *IEEE Trans. Veh. Technol.*, vol. 52, no. 5, pp. 1386–1397, Sept. 2003.

[24] F. Simoens and M. Moeneclaey, "Improved embedded parameter estimation for bit-interleaved coded modulation with iterative decoding," in *Proc. International Conference on Acoustics, Speech and Signal Processing (ICASSP) 2005*, vol. 3, pp. 489–492.

[25] J. B. Anderson, "Array gain and capacity for known random channels with multiple element arrays at both ends," *IEEE J. Sel. Areas Commun.*, vol. 18, no. 11, pp. 2172–2178, Nov. 2000.

[26] M. A. Khalighi, "Iterative equalization and decoding with imperfect channel estimates," in *Proc. International Conference on Telecommunications (ICT) 2002*, vol. 1, pp. 88–92.

Mohammad-Ali Khalighi was born in Kerman, Iran in 1975. He received his B.Sc. and M.Sc. degrees from Sharif University of Technology, Tehran, in 1995 and 1997, respectively, and his Ph.D. from Institut National Polytechnique de Grenoble (INPG), Grenoble, France in 2002, in Electrical Engineering. From 2002 to 2005, he has been with the Laboratoire des Images et des Signaux (LIS), Grenoble, École Nationale Supérieure des Télécommunications (ENST), Paris, and Institut d' Électronique et de Télécommunications de Rennes, as a post-doctoral research fellow. He joined Institut Fresnel, Marseille, in 2005 as an assistant professor. His main research interests include signal detection and channel estimation in high data rate communication systems.

Joseph Jean Boutros was born in Beirut, Lebanon, in 1967. He received the electrical engineering degree in 1992 and the Ph.D. degree in 1996 from École Nationale Supérieure des Télécommunications (ENST), Paris, France. Since September 1996, he is with the Communications and Electronics Department at ENST as an Associate Professor. His fields of interest are codes on graphs, iterative decoding, joint source-channel coding, space-time coding and lattice sphere packings.