Comparison of EM-Based Algorithms for MIMO Channel Estimation

Xavier Wautelet, Cédric Herzet, Antoine Dejonghe, Jérôme Louveaux, *Member, IEEE*, and Luc Vandendorpe, *Fellow, IEEE*

***Abstract*—Iterative channel estimation can improve the channel-state information (CSI) with respect to noniterative estimation. New iterative channel estimators based on the expectation-maximization (EM) algorithm are proposed in this paper. A first estimator, called the unbiased EM (UEM), is designed to unbias the EM estimates. A second estimator is then put forward, which is based on the expectation-conditional-maximization (ECM) algorithm, and its complexity is lower than that of the EM. An unbiased ECM (UECM) estimator is also proposed. Although the unbiasedness of the UEM and UECM estimators is not rigorously proved, the use of these names is explained in the paper. The new estimators are compared with well-known ones, such as the EM, the decision-directed (DD), and the data-aided (DA) estimators. Simulations are reported for a turbo receiver operating over frequency-selective multiple-input multiple-output channels. It is shown that the UEM channel estimator outperforms the EM, and that the ECM-based estimators are very close to the EM-based ones.**

***Index Terms*—Channel estimation, expectation-maximization (EM) algorithm, iterative estimation, multiple-input multiple-output (MIMO) systems, turbo detection.**

I. Introduction

THE DEMAND for high-quality, high-bit-rate communications grows steadily. A good knowledge of the channel parameters is essential to achieve an efficient detection of the transmitted signal. The required channel estimation is often of the data-aided (DA) type [1, pp. 782–787]–[4], i.e., it is carried out with the help of pilot symbols. However, the length of the required training sequence may become huge when the channel impulse response (CIR) consists of many taps, such as in wideband multiple-input multiple-output (MIMO) channels. It results in a waste of bandwidth and power. Instead of using the pilot symbols only, the information carried by the observations corresponding to the data symbols can also be used to help the channel estimator. It can lead either to better channel-parameter estimates or to a smaller number of required pilot symbols. A first approach is to use decision-directed (DD) [1, pp. 772–782] estimators, but a part of the information is lost when hard decisions are taken. A second approach is to use "soft decisions" [5]–[14] which provide confidence measures about the decisions. In soft-decision estimators, the channel estimation is progressively improved through iterations [6], [7], [11].

Several papers have recently been devoted to an iterative refinement of the channel-state information (CSI), e.g., [5]–[14]. Various algorithms have been considered in these papers. Among them, there are soft-DD variants of known techniques, such as the least-square (LS) or the Kalman estimators [10]. Another technique is the so-called expectation-maximization (EM) algorithm [15], which has already been considered in [5]–[8] for channel estimation. It is an attractive choice because it provides a theoretical framework to ensure the convergence towards a stationary point of the likelihood function under fairly general conditions [16]. With the available data only, i.e., the incomplete data set according to the EM terminology, the maximum-likelihood (ML) estimates are sometimes computationally too demanding. With another data set, the so-called complete data set, the ML estimates can be easily computed, but not all the elements of the complete data set are known. The EM algorithm provides a framework to iteratively calculate probabilities on the unknown elements of the complete data set and estimate the unknown parameters to be estimated.

When the channel and the transmitted symbols are unknown, two different applications of the EM algorithm are conceivable. First, as in [5]–[8], the target can be the channel estimation. The EM algorithm enables iteratively refining the channel estimates, thanks to the calculation of symbol *a posteriori* probabilities (APPs). Secondly, as proposed by [12]–[14], the EM algorithm can be used to directly estimate the symbols, since this is the final goal of the receiver. In this latter case, the symbol decisions are iteratively refined, computing APPs on the channel taps. In this paper, we will consider EM-based methods for channel estimation in MIMO frequency-selective communications systems. We only focus on the first solution, because the second one usually leads to trellis-based solutions when trellis coding is used [12], and trellis-based solutions are intractable when the CIR length is large and/or when the constellation size of the considered modulation is high [17].

A major problem with the EM channel-tap estimation is that the estimates are biased, as shown in [6]. This bias can severely degrade the receiver performance. We propose in this paper an unbiased EM (UEM) channel estimator. We show that the UEM estimator outperforms the classical EM estimator. The EM and

Paper approved by A. Lozano, the Editor for Wireless Communication of the IEEE Communications Society. Manuscript received May 24, 2004; revised September 22, 2005; February 10, 2006; and May 18, 2006. This work was supported in part by the Federal Office for Scientific, Technical and Cultural Affairs (OSTC, Belgium) under IAP Contract No. P5/11. The work of A. Dejonghe was supported by the Belgian National Science Foundation. This paper was presented in part at the Third International Symposium on Turbo Codes and Related Topics, Brest, France, September 2003.

X. Wautelet, C. Herzet, J. Louveaux, and L. Vandendorpe are with the Communications and Remote Sensing Laboratory, Université catholique de Louvain, B-1348 Louvain-la-Neuve, Belgium (e-mail: wautelet@tele.ucl.ac.be; herzet@tele.ucl.ac.be; louveaux@tele.ucl.ac.be; vandendorpe@tele.ucl.ac.be).

C. Herzet was with the Communications and Remote Sensing Laboratory, Université catholique de Louvain, B-1348 Louvain-la-Neuve, Belgium. He is now with the University of California, Berkeley, CA 94720 USA.

A. Dejonghe is with the Interuniversity Micro-Electronics Center (IMEC), B-3001 Leuven, Belgium (e-mail: dejonghe@imec.be).

Digital Object Identifier 10.1109/TCOMM.2006.887507

UEM estimators require a matrix inversion. In order to avoid this matrix inversion, and thus to reduce the estimator complexity, we also propose to estimate the channel with the expectation-conditional-maximization (ECM) algorithm [18], [19], which decreases the complexity of the maximization step of the iterative estimation process. Like the EM algorithm, the ECM algorithm leads to a biased channel-tap estimate. We derive here an unbiased ECM (UECM) algorithm for channel estimation. Although it is not rigorously proved in the paper that the UEM and UECM estimators are unbiased, a justification of these estimators is given.

In addition to the channel taps, the noise variance is also needed to perform soft-in soft-out (SISO) equalization (see [20, eq. (14)]). There are two sources of bias in the noise-variance estimation. The first one is due to the joint estimation with the channel-tap values. We show that even with a DA ML estimator, the noise variance is biased when joint CIR/noise-variance estimation is performed, but it is asymptotically unbiased as the burst length increases. We show that this bias can be removed. The second source of bias is the imperfect knowledge of the symbols. We show that this bias can be divided by two.

Besides the theoretical part of the paper, we run simulations to compare the performance of the new UEM, ECM, and UECM algorithms with the one of the already known methods such as the DA ML, DD ML, and EM algorithms and with an approximate solution of the EM algorithm. We compare them in the context of the estimation of frequency-selective MIMO channels. A communications system based on bit-interleaved coded modulation (BICM) [21] is considered. The MIMO channel is frequency-selective and quasi-static, i.e., the channel taps remain constant over one burst length. A fractionally spaced turbo receiver made up of three main parts is considered. The first part is the channel estimator. The second part is the equalizer/demodulator, which is based on the minimum mean-square error (MMSE) criterion [17]. The third part is the classical SISO Bahl–Cocke–Jelinek–Raviv (BCJR) [22] decoder. With appropriate interleaving/deinterleaving, the channel estimator, the equalizer/demodulator, and the decoder exchange soft information about the coded bits and estimates of the channel taps and noise variance. The performance of the different algorithms are compared on the basis of the bit-error rate (BER) and mean-square error (MSE) of the estimators.

The rest of this paper is organized as follows. In Section II, the channel estimators are derived. In Section III, the transmitter, the channel model, and the receiver for the turbo MIMO setup described above are described. The simulation results are presented in Section IV, and the conclusions are drawn in Section V. The calculation about the bias of the noise-variance estimator is given in the Appendix.

II. Channel-Estimation Algorithms

We focus here on iterative algorithms for quasi-static channel estimation. That means that the channel is constant inside each frame.

We consider a MIMO system with n_T transmit antennas and n_R receive antennas. Each transmitted frame is made up of $n_T L_s$ complex symbols $s_k^{(i)} (k = 1, \ldots, L_s)$. The symbol $s_k^{(i)}$ is sent from antenna $i (i = 1, \ldots, n_T)$ at time kT, where T denotes the symbol period. The lowpass-equivalent CIR (including the pulse-shaping filter) between transmit antenna i and receive antenna $j (j = 1, \ldots, n_R)$ is denoted by $h^{(i,j)}(t)$. We assume that $h^{(i,j)}(t)$ can be truncated without loss of accuracy. We only keep its values $\forall i, \forall j$ for $-L_1 T \leq t \leq L_2 T$. The channel length is denoted by $L \triangleq L_1 + L_2 + 1$. At receive antenna j, $r^{(j)}(t)$ denotes the received signal, and $n^{(j)}(t)$ 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. Let $F_{\max}$ be the highest frequency of the useful part of the received signal. Samples are taken at the rate M_s/T after lowpass filtering at $M_s/2T$. The oversampling factor M_s is chosen such that $M_s/2T > F_{\max}$. For $l = 1 - L_1, \ldots, L_s + L_2$ and $m = 0, \ldots, M_s - 1$, the received samples $r_{l,m}^{(j)} \triangleq r^{(j)}(lT + mT/M_s)$ are sufficient statistics [1, pp. 234–245] for symbol detection and channel estimation. We also define the channel taps $h_{l',m}^{(i,j)} \triangleq h^{(i,j)}(l'T + mT/M_s)$ for $l' = -L_1, \ldots, L_2$ and the noise samples $n_{l,m}^{(j)} \triangleq n^{(j)}(lT + mT/M_s)$. The noise is assumed to be white in the space and time domains, i.e., it is independent $\forall j, \forall l, \forall m$. The MIMO channel equation is thus

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

In our channel estimators, we will only consider the observation samples which are free from the interference of the adjacent frames, i.e., $r_{l,m}^{(j)}$ for $L_2 < l \leq L_s - L_1 \, \forall m, \forall j$. A compact representation may be obtained by stacking the samples. The vector $\underline{r}_m^{(j)}$ of the mth polyphase component of the received samples at antenna j is then

$$\underline{r}_m^{(j)} = \left[r_{1+L_2,m}^{(j)} \cdots r_{L_s-L_1,m}^{(j)} \right]_{(L_r \times 1)}^T \tag{2}$$

where $L_r = L_s + 1 - L$. The corresponding noise vector $\underline{n}_m^{(j)}$ can be written as

$$\underline{n}_m^{(j)} = \left[n_{1+L_2,m}^{(j)} \cdots n_{L_s-L_1,m}^{(j)} \right]_{(L_r \times 1)}^T. \tag{3}$$

Let $\underline{\underline{S}}^{(i)}$ be the following Toeplitz matrix containing the symbols transmitted from antenna i:

$$\underline{\underline{S}}^{(i)} = \begin{bmatrix} s_L^{(i)} & s_{L-1}^{(i)} & \cdots & s_1^{(i)} \\ s_{L+1}^{(i)} & s_L^{(i)} & \cdots & s_2^{(i)} \\ \ddots & \ddots & \ddots & \ddots \\ s_{k+L_1}^{(i)} & s_{k+L_1-1}^{(i)} & \cdots & s_{k-L_2}^{(i)} \\ \ddots & \ddots & \ddots & \ddots \\ s_{L_s}^{(i)} & s_{L_s-1}^{(i)} & \cdots & s_{L_r}^{(i)} \end{bmatrix}_{(L_r \times L)}. \tag{4}$$

The matrix containing all the transmitted symbols is $\underline{\underline{S}}$

$$\underline{\underline{S}} = \left[\underline{\underline{S}}^{(1)} \cdots \underline{\underline{S}}^{(n_T)} \right]. \tag{5}$$

Finally, the samples of the mth polyphase component of the CIR corresponding to the receive antenna j are stacked in the vector $\underline{h}_m^{(j)}$

$$\underline{h}_m^{(j)} = \left[h_{-L_1,m}^{(1,j)} \dots h_{L_2,m}^{(1,j)} h_{-L_1,m}^{(2,j)} \dots h_{L_2,m}^{(n_T,j)}\right]^T_{(n_T L \times 1)}. \quad (6)$$

The observation model becomes

$$\underline{r}_m^{(j)} = \underline{\underline{S}} \underline{h}_m^{(j)} + \underline{n}_m^{(j)}. \quad (7)$$

A. DA ML Estimator

A training sequence enables to carry out channel acquisition. Since the channel taps are assumed constant over the burst duration and the $n_T L_p$ pilot symbols are known, the channel estimation can easily be performed according to the ML criterion. Even if the channel-tap estimation is performed jointly with the noise-variance (σ_n^2) estimation, the channel-tap estimates do not depend on the noise-variance estimate $\hat{\sigma}_{n\mathrm{DA}}^2$. Let $\underline{\underline{S_p}}$ contain the transmitted training sequence of length L_p, and let $\underline{r_p}_m^{(j)}$ be the corresponding received samples

$$\underline{r_p}_m^{(j)} = \underline{\underline{S_p}} \underline{h}_m^{(j)} + \underline{n}_m^{(j)} \quad (8)$$

where vector $\underline{r_p}_m^{(j)}$ is of length $L_{rp} = L_p - L + 1$. Since the noise is assumed to be white both in the space and time domains, the channel-estimation problem according to the ML criterion can be split up into $n_R \times M_s$ separate subproblems. For each polyphase component m and each receive antenna j, we get the following expression for the estimate of the CIR $\hat{\underline{h}}_{m,\mathrm{DA}}^{(j)}$:

$$\hat{\underline{h}}_{m,\mathrm{DA}}^{(j)} = \left(\underline{\underline{S_p}}^H \underline{\underline{S_p}}\right)^{-1} \underline{\underline{S_p}}^H \underline{r_p}_m^{(j)}. \quad (9)$$

Note that in (9), only the data-free samples are kept in $\underline{r_p}_m^{(j)}$. That means that the received samples depend on the pilot symbols only and do not depend on unknown data symbols. In this case, (9) also corresponds to the LS estimate of the channel [23]. The estimator (9) is the only efficient estimator of the channel taps, i.e., it is unbiased, and its variance is minimal [23].

The pilot symbols which are optimal for minimizing the estimator MSE are such that the matrix $\underline{\underline{S_p}}^H \underline{\underline{S_p}}$ is diagonal [25]. Furthermore, the inversion of this matrix becomes very easy. For more details about the training sequence generation, the reader is referred to [25]. $(n_T + 1)L - 1$ is the minimum training-sequence length L_p for the channel-tap estimation, because it enables the inversion of matrix $\underline{\underline{S_p}}^H \underline{\underline{S_p}}$ in (9).

When the CIR estimate has been computed, the noise variance can also be estimated according to the ML criterion. The noise-variance estimate $\hat{\sigma}_{n,\mathrm{DA}}^2$ is then

$$\begin{aligned}\hat{\sigma}_{n,\mathrm{DA}}^2 = \frac{1}{n_R M_s L_{rp}} &\sum_{j=1}^{n_R} \\ &\times \sum_{m=0}^{M_s-1} \left[\left(\underline{r_p}_m^{(j)} - \underline{\underline{S_p}}\hat{\underline{h}}_{m,\mathrm{DA}}^{(j)}\right)^H \right. \\ &\left. \times \left(\underline{r_p}_m^{(j)} - \underline{\underline{S_p}}\hat{\underline{h}}_{m,\mathrm{DA}}^{(j)}\right)\right]. \end{aligned} \quad (10)$$

In (10), the noise variance is simply evaluated by measuring the squared difference between the received signal $\underline{r_p}_m^{(j)}$ and its estimated useful part $\underline{\underline{S_p}}\hat{\underline{h}}_{m,\mathrm{DA}}^{(j)}$. The noise-variance estimator is biased because of the joint estimation with the CIR. This bias is similar to the bias which appears when one tries to jointly estimate the mean and variance of a random variable. The detailed derivation about that bias of the noise-variance estimate can be found in the Appendix. It is shown that the unbiased estimate is

$$\hat{\sigma}_{n,\mathrm{UDA}}^2 = \frac{L_{rp}}{L_{rp} - n_T L} \hat{\sigma}_{n,\mathrm{DA}}^2. \quad (11)$$

The difference between (10) and (11) is small if the training sequence is much larger than the CIR, i.e., $L_{rp} \gg n_T L$, but it is significant otherwise. Looking at the denominator of (11), we must have $L_{rp} > n_T L$. Since $L_{rp} = L_p - L + 1$, the minimum training-sequence length L_p for the joint acquisition of the channel taps and noise variance is $(n_T + 1)L$.

The equalizer performance is very sensitive to channel-estimation errors, as it will be shown in Section IV. The DA ML estimation only provides a poor estimate of the channel parameters. Of course, many pilot symbols could be sent to improve the estimation quality, but a lot of bandwidth and of power would be wasted to send those symbols.

B. EM Estimator

The DA ML estimator does not exploit all the information available about the channel because it makes use of the data-free observation samples only. To improve the channel-estimate quality, the channel can be estimated on the basis of the whole received burst, including both the training sequence and the data. As shown in (9) and (10), it is easy to compute the ML channel estimate if the transmitted symbols are known. However, when the transmitted sequence is unknown and encoded, the likelihood function exploiting the code is much more difficult to compute.

The EM algorithm [15] enables iteratively solving this problem. The EM algorithm has been exploited previously to estimate the channel [5]–[8], but it has usually been used to estimate the channel taps only. Here, the EM algorithm is explained again in the case of the joint estimation of the channel taps and noise variance. Let us define the vector of parameters to be estimated as $\mathcal{B}$, which contains the noise variance σ_n^2 and the channel taps $\underline{h}_m^{(j)}$ $\forall j, \forall m$. Using the EM formalism, the received samples $\underline{r}_m^{(j)}$ $\forall j, \forall m$ form the incomplete data set $\mathcal{R}$, a set from which it is difficult to calculate the ML estimate $\hat{\mathcal{B}}$. The received samples $\underline{r}_m^{(j)}$ $(\forall m, \forall j)$ together with the transmitted symbols $\underline{\underline{S}}$ form the complete data set $\mathcal{Z}$, a set from which it is easier to calculate the ML channel-parameter estimates. The EM algorithm provides a framework to iteratively estimate the channel parameters when the symbols $\underline{\underline{S}}$ are unavailable at the receiver. At iteration n, the expectation step, the first step of this method, is the computation of the function $\mathcal{Q}(\tilde{\mathcal{B}}, \hat{\mathcal{B}}^{(n-1)})$, which depends on the estimates $\hat{\mathcal{B}}^{(n-1)}$ at the previous iteration and on the trial value $\tilde{\mathcal{B}}$ for the parameters to be estimated

$$\mathcal{Q}\left(\tilde{\mathcal{B}}, \hat{\mathcal{B}}^{(n-1)}\right) = \int_{\mathcal{Z}} p\left(\mathcal{Z}|\mathcal{R}, \hat{\mathcal{B}}^{(n-1)}\right) \ln p(\mathcal{Z}|\tilde{\mathcal{B}})\, d\mathcal{Z}. \quad (12)$$

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

$$\hat{\mathcal{B}}^{(n)} = \arg \max_{\tilde{\mathcal{B}}} \left\{ \mathcal{Q}\left(\tilde{\mathcal{B}}, \hat{\mathcal{B}}^{(n-1)}\right) \right\}. \tag{13}$$

It can be shown [16] that sequence $\hat{\mathcal{B}}^{(n)}$ converges under fairly general conditions towards a stationary point of the likelihood function $p(\mathcal{R}|\tilde{\mathcal{B}})$. Moreover, at each EM iteration, this likelihood does not decrease. If initial estimate $\hat{\mathcal{B}}^{(0)}$ is close enough to $\mathcal{B}$, we may thus hope that sequence $\hat{\mathcal{B}}^{(n)}$ converges to the global maximum of $p(\mathcal{R}|\tilde{\mathcal{B}})$, i.e., the ML estimate of $\mathcal{B}$.

Since the received samples $\mathcal{R}$ are known, (12) can be rewritten as

$$\mathcal{Q}\left(\tilde{\mathcal{B}}, \hat{\mathcal{B}}^{(n-1)}\right) = \int_{\underline{\underline{S}}} p\left(\underline{\underline{S}}|\mathcal{R}, \hat{\mathcal{B}}^{(n-1)}\right) \ln\left[p(\mathcal{R}|\underline{\underline{S}}, \tilde{\mathcal{B}}) p(\underline{\underline{S}}|\tilde{\mathcal{B}})\right] d\underline{\underline{S}}. \tag{14}$$

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

$$\mathcal{Q}'\left(\tilde{\mathcal{B}}, \hat{\mathcal{B}}^{(n-1)}\right) = \int_{\underline{\underline{S}}} p\left(\underline{\underline{S}}|\mathcal{R}, \hat{\mathcal{B}}^{(n-1)}\right) \ln p(\mathcal{R}|\underline{\underline{S}}, \tilde{\mathcal{B}})\, d\underline{\underline{S}} \tag{15}$$

with

$$\begin{aligned}\ln p(\mathcal{R}|\underline{\underline{S}}, \tilde{\mathcal{B}}) &= -n_R M_s L_r \ln\left(\pi \tilde{\sigma}_n^2\right) \\ &\quad - \frac{1}{\tilde{\sigma}_n^2} \sum_{j=1}^{n_R} \sum_{m=0}^{M_s-1} \left[\left(\underline{r}_m^{(j)} - \underline{\underline{S}}\tilde{\underline{h}}_m^{(j)}\right)^H \right. \\ &\quad \left. \times \left(\underline{r}_m^{(j)} - \underline{\underline{S}}\tilde{\underline{h}}_m^{(j)}\right)\right]\end{aligned} \tag{16}$$

where $\tilde{\underline{h}}_m^{(j)}$ and $\tilde{\sigma}_n^2$ are, respectively, the trial values for $\underline{h}_m^{(j)}$ and σ_n^2. The maximization step gives the following equation for the channel-tap estimation at iteration n:

$$\hat{\underline{h}}_{m,\mathrm{EM}}^{(j)(n)} = \left(E\left[\underline{\underline{S}}^H \underline{\underline{S}}|\mathcal{R}, \hat{\mathcal{B}}^{(n-1)}\right]\right)^{-1} E\left[\underline{\underline{S}}|\mathcal{R}, \hat{\mathcal{B}}^{(n-1)}\right]^H \underline{r}_m^{(j)} \tag{17}$$

and the following expression for the noise-variance estimation:

$$\begin{aligned}\hat{\sigma}_{n,\mathrm{EM}}^{2(n)} &= \frac{1}{n_R M_s L_r} \sum_{j=1}^{n_R} \sum_{m=0}^{M_s-1} \left[\underline{r}_m^{(j)^H} \underline{r}_m^{(j)}\right. \\ &\quad + \hat{\underline{h}}_{m,\mathrm{EM}}^{(j)^H} E\left[\underline{\underline{S}}^H \underline{\underline{S}}|\mathcal{R}, \hat{\mathcal{B}}^{(n-1)}\right] \hat{\underline{h}}_{m,\mathrm{EM}}^{(j)} \\ &\quad \left. -2\Re\left\{\underline{r}_m^{(j)^H} E\left[\underline{\underline{S}}|\mathcal{R}, \hat{\mathcal{B}}^{(n-1)}\right] \hat{\underline{h}}_{m,\mathrm{EM}}^{(j)}\right\}\right].\end{aligned} \tag{18}$$

Fortunately, the joint estimation of the channel parameters does not lead to a (nonlinear) system coupled between the noise variance and the channel taps. Indeed, the channel-tap estimation does not depend on the noise-variance estimation. We can, therefore, estimate the taps, then the noise variance.

In this part of the paper, we assume that a part of the receiver, called the APP calculator, is able to output the probabilities $p(\underline{\underline{S}}|\mathcal{R}, \hat{\mathcal{B}}^{(n-1)})$ required in (15), and thus, in the conditional expectations in (17) and (18). The resulting scheme is depicted in Fig. 1. The EM estimation is iterative. At the first iteration, an initial estimate of the CSI is provided to the APP calculator. This initial estimate can be computed with the help of pilot symbols and (9) and (11), for instance. For the next EM iterations, (17) and (18) are computed, using the APPs on the symbols which are delivered by the APP calculator. The channel estimator can then deliver a novel refined CSI to the APP calculator, and so on.

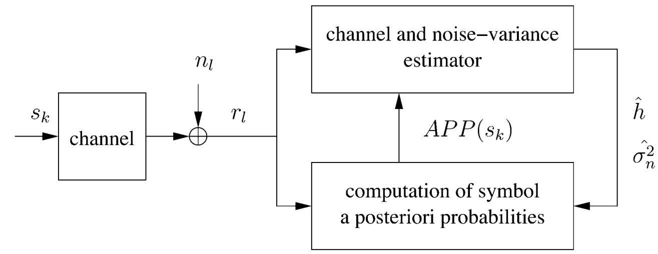


Fig. 1. EM-based channel-estimator scheme.

C. LS and UEM Estimators

In this section, we show that the EM CIR estimator can be viewed as an LS estimator. The LS estimator [23] for channel estimation minimizes the square error $J(\hat{\underline{r}}_m^{(j)})$ of the reconstructed received signal $\hat{\underline{r}}_m^{(j)}$

$$J\left(\hat{\underline{r}}_m^{(j)}\right) = \left(\underline{r}_m^{(j)} - \hat{\underline{r}}_m^{(j)}\right)^H \left(\underline{r}_m^{(j)} - \hat{\underline{r}}_m^{(j)}\right) \tag{19}$$

where $\hat{\underline{r}}_m^{(j)} = \underline{\underline{S}}\hat{\underline{h}}_{m,\mathrm{LS}}^{(j)}$. When the symbols are known, the solution that minimizes $J(\hat{\underline{r}}_m^{(j)})$ is equivalent to the ML DA estimator (9). However, the data symbols of the frame are unknown. We only assume that *a priori* information $\mathcal{P}$ about the symbol distribution is available. In order to take advantage of these *a priori* probabilities, the LS criterion has to be modified. We choose to minimize $J_2(\hat{\underline{h}}_m^{(j)})$, which is the average of $J(\hat{\underline{r}}_m^{(j)})$ over the possible sequences

$$J_2\left(\hat{\underline{h}}_m^{(j)}\right) = E_{\underline{\underline{S}}}\left[J\left(\hat{\underline{r}}_m^{(j)}\right); \mathcal{P}\right] \tag{20}$$

where the notation $E_{\underline{\underline{S}}}[\ldots]$ means that the expectation is taken with respect to the distribution of $\underline{\underline{S}}$, that is to say $\mathcal{P}$. The notation $E[\ldots; \mathcal{P}]$ emphasizes the fact that the symbols are not necessarily equiprobable. The solution $\hat{\underline{h}}_{m,\mathrm{LS}}^{(j)}$ that minimizes $J_2(\hat{\underline{h}}_m^{(j)})$ is

$$\hat{\underline{h}}_{m,\mathrm{LS}}^{(j)} = (E[\underline{\underline{S}}^H \underline{\underline{S}}; \mathcal{P}])^{-1} E[\underline{\underline{S}}; \mathcal{P}]^H \underline{r}_m^{(j)}. \tag{21}$$

It has been shown already [6] that the EM channel-tap estimator is biased and that this bias can degrade the BER. The mean of the EM estimates is very difficult to calculate because of the dependence of the APPs $p(\underline{\underline{S}}|\mathcal{R}, \hat{\mathcal{B}}^{(n-1)})$ on the received samples $\mathcal{R}$ and on the previous estimate $\hat{\mathcal{B}}^{(n-1)}$. On the contrary, the mean of the LS estimator for *a priori* symbol distribution $\mathcal{P}$ is easy to find. On the basis of (21), the expectation of the LS channel-tap estimate is

$$E_{\underline{r}_m^{(j)}}\left[\hat{\underline{h}}_{m,\mathrm{LS}}^{(j)};\mathcal{P}\right]=(E[\underline{\underline{S}}^H\underline{\underline{S}};\mathcal{P}])^{-1}E[\underline{\underline{S}};\mathcal{P}]^H E\left[\underline{r}_m^{(j)};\mathcal{P}\right]. \quad (22)$$

Assuming that the noise is zero-mean and using (7), we get

$$E_{\underline{r}_m^{(j)}}\left[\hat{\underline{h}}_{m,\mathrm{LS}}^{(j)};\mathcal{P}\right]=(E[\underline{\underline{S}}^H\underline{\underline{S}};\mathcal{P}])^{-1}E[\underline{\underline{S}};\mathcal{P}]^H E[\underline{\underline{S}};\mathcal{P}]\underline{h}_m^{(j)}. \quad (23)$$

The channel-estimate mean is thus not the true channel-tap value. The bias factor is $(E[\underline{\underline{S}}^H\underline{\underline{S}};\mathcal{P}])^{-1}E[\underline{\underline{S}};\mathcal{P}]^H E[\underline{\underline{S}};\mathcal{P}]$. The LS estimator can be unbiased by multiplying (21) by the bias inverse. We then get the unbiased LS estimator

$$\hat{\underline{h}}_{m,\mathrm{ULS}}^{(j)}=(E[\underline{\underline{S}};\mathcal{P}]^H E[\underline{\underline{S}};\mathcal{P}])^{-1}E[\underline{\underline{S}};\mathcal{P}]^H\underline{r}_m^{(j)}. \quad (24)$$

Actually, the LS estimator is not very useful when the symbols are zero-mean and *a priori* equiprobable, and when this information is not updated. The estimate is then zero. However, the LS estimator can be used in a "turbo-like" fashion, i.e., the *a priori* distribution used by the estimator is the output of the symbol APP calculator. *A priori* symbol distribution $\mathcal{P}$ is then updated through the iterations, and it is equal to the APPs $p(\underline{\underline{S}}|\mathcal{R},\hat{\mathcal{B}}^{(n-1)})$. In this case, the LS estimator appears to be equal to the EM estimator.

Based on the analogy between the EM and the LS estimators, we propose a new version of the EM algorithm

$$\hat{\underline{h}}_{m,\mathrm{UEM}}^{(j)}=\left(E\left[\underline{\underline{S}}|\mathcal{R},\hat{\mathcal{B}}^{(n-1)}\right]^H E\left[\underline{\underline{S}}|\mathcal{R},\hat{\mathcal{B}}^{(n-1)}\right]\right)^{-1}\times E\left[\underline{\underline{S}}|\mathcal{R},\hat{\mathcal{B}}^{(n-1)}\right]^H\underline{r}_m^{(j)}. \quad (25)$$

Although it has not been rigorously proved that (25) is an unbiased estimator, it will be referred to as the UEM algorithm in the following.

If the EM noise-variance estimator (18) used updated symbol *a priori* probabilities $\mathcal{P}$ instead of $p(\underline{\underline{S}}|\mathcal{R},\hat{\mathcal{B}}^{(n-1)})$, its mean could be easily calculated, too. It is shown in the Appendix that even the DA ML noise-variance estimator is biased, due to the joint estimation with the channel taps. We put aside this bias by assuming that the channel taps are known. Using $\mathcal{P}$, which is updated through the iterations, instead of $p(\underline{\underline{S}}|\mathcal{R},\hat{\mathcal{B}}^{(n-1)})$, the mean of the modified EM estimator $\hat{\sigma}^2_{n,\mathrm{EM}'}$ is given by

$$E_{\mathcal{R}}[\hat{\sigma}^2_{n,\mathrm{EM}'};\mathcal{P}] = \sigma_n^2+\frac{2}{n_R M_s L_r}\sum_{j=1}^{n_R}\sum_{m=0}^{M_s-1}\left[\underline{h}_m^{(j)^H}\times(E[\underline{\underline{S}}^H\underline{\underline{S}};\mathcal{P}]-E[\underline{\underline{S}};\mathcal{P}]^H E[\underline{\underline{S}};\mathcal{P}])\underline{h}_m^{(j)}\right]. \quad (26)$$

This bias is due to the imperfect knowledge of the symbols. Indeed, in the known-symbol case, $E[\underline{\underline{S}}^H\underline{\underline{S}};\mathcal{P}] = E[\underline{\underline{S}};\mathcal{P}]^H E[\underline{\underline{S}};\mathcal{P}]$, and the bias of $\hat{\sigma}^2_{n,\mathrm{EM}'}$ is null. The bias of $\hat{\sigma}^2_{n,\mathrm{EM}'}$ can be removed, but the "unbiased" EM estimator of the noise variance

$$\hat{\sigma}^2_{n,\mathrm{UEM}'}=\sum_{j=1}^{n_R}\sum_{m=0}^{M_s-1}\frac{\underline{r}_m^{(j)^H}\underline{r}_m^{(j)}-\underline{h}_m^{(j)^H}E[\underline{\underline{S}}^H\underline{\underline{S}}|\mathcal{P}]\underline{h}_m^{(j)}}{n_R M_s L_r} \quad (27)$$

is useless because it can sometimes provide negative values. We propose to use another estimator, the half-biased EM (HEM) estimator

$$\hat{\sigma}^{2(n)}_{n,\mathrm{HEM}}=\sum_{j=1}^{n_R}\sum_{m=0}^{M_s-1}\left[\left[\underline{r}_m^{(j)}-E[\underline{\underline{S}}|\mathcal{R},\hat{\mathcal{B}}^{(n-1)}]\hat{\underline{h}}_{m,\mathrm{EM}}^{(j)}\right]^H\times\frac{\underline{r}_m^{(j)}-E\left[\underline{\underline{S}}|\mathcal{R},\hat{\mathcal{B}}^{(n-1)}\right]\hat{\underline{h}}_{m,\mathrm{EM}}^{(j)}}{n_R M_s L_r}\right] \quad (28)$$

which can easily be shown to exhibit half the bias of $\hat{\sigma}^2_{n,\mathrm{EM}'}$ when updated $\mathcal{P}$ is used instead of $p(\underline{\underline{S}}|\mathcal{R},\hat{\mathcal{B}}^{(n-1)})$.

D. Variants of the EM Estimator

1) The ECM Estimator: The inversion of matrix $E[\underline{\underline{S}}^H\underline{\underline{S}}|\mathcal{R},\hat{\mathcal{B}}^{(n-1)}]$ is a demanding task in the EM estimator. This inversion can be avoided by means of the ECM algorithm [18], [19]. With the EM algorithm, all the taps of the CIR are jointly estimated (17). That leads to $n_R M_s$ linear systems, each one with $n_T L$ equations and $n_T L$ unknowns, to be solved during the M step. The basic idea of the ECM algorithm is to simplify the M step: one channel tap is estimated at a time, assuming that 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 $\tilde{\underline{h}}_{l,m}^{(i,j)(n)}$ for the calculation of the estimate of $h_{l,m}^{(i,j)}$ at iteration n, $\hat{h}_{l,m}^{(i,j)(n)}$

$$\tilde{\underline{h}}_{l,m}^{(i,j)(n)}=\left[\hat{h}_{-L_1,m}^{(1,j)(n)}\ldots\hat{h}_{l-1,m}^{(i,j)(n)}0\hat{h}_{l+1,m}^{(i,j)(n-1)}\ldots\hat{h}_{L_2,m}^{(n_T,j)(n-1)}\right]^T. \quad (29)$$

The E step of the ECM algorithm is the same as the EM one, but the M step is different. During the M step of an ECM iteration, each tap is estimated in turn, as follows:

$$\hat{h}_{l,m,\mathrm{ECM}}^{(i,j)(n)}=\frac{\left\{E\left[\underline{\underline{S}}|\mathcal{R},\hat{\mathcal{B}}^{(n-1)}\right]^H\underline{r}_m^{(j)}\right\}_{Li+l}}{\left\{E\left[\underline{\underline{S}}^H\underline{\underline{S}}|\mathcal{R},\hat{\mathcal{B}}^{(n-1)}\right]\right\}_{Li+l,Li+l}}-\frac{\left\{E\left[\underline{\underline{S}}^H\underline{\underline{S}}|\mathcal{R},\hat{\mathcal{B}}^{(n-1)}\right]\tilde{\underline{h}}_{l,m}^{(i,j)(n)}\right\}_{Li+l}}{\left\{E\left[\underline{\underline{S}}^H\underline{\underline{S}}|\mathcal{R},\hat{\mathcal{B}}^{(n-1)}\right]\right\}_{Li+l,Li+l}} \quad (30)$$

where $\{\underline{\underline{A}}\}_{b,c}$ denotes the element located at row $b-L_2$ and column $c-L_2$ of matrix A, and $\{\underline{a}\}_b$ denotes the element located at row $b-L_2$ of vector $\underline{a}$. For each iteration n, there are $n_R M_s n_T L$ linear equations such as (30), each one with only one unknown: $\hat{h}_{-L_1,m}^{(i,j)}$ is estimated, then $\hat{h}_{-L_1+1,m}^{(i,j)}$, etc. The solution is straightforward and the matrix inversion is avoided.

TABLE I
ESTIMATOR COMPLEXITY PER ITERATION

estimator	# multiplications	# divisions
EM	$(n_T L)^2(2L_s + 4n_T L + 2)$ $+n_T L L_s M_s n_R$	$2n_T L$
ECM	$(n_T L)^2(L_s + M_s n_R)$ $+n_T L L_s(M_s n_R + 1)$	$n_T L$
Diag EM	$n_T L L_s(M_s n_R + 2)$	$n_T L$

Table I shows that the EM is roughly twice as complex as the ECM when the frame length L_s is large.

2) The UECM Estimator: Like the EM channel estimator, the ECM estimator is biased. We propose an estimator that we call the UECM estimator. Like above, the UECM unbiasedness can be shown when the estimator uses updated symbol *a priori* distribution $\mathcal{P}$ instead of the APPs. For $\hat{h}_{l,m}^{(i,j)}$ at iteration n, the UECM estimate is

$$\hat{h}_{l,m,\mathrm{UECM}}^{(i,j)(n)} = \frac{\left\{E[\underline{\underline{S}}]^H \underline{r}_m^{(j)}\right\}_{Li+l} - \left\{E[\underline{\underline{S}}]^H E[\underline{\underline{S}}]\underline{\tilde{h}}_{l,m,\mathrm{UECM}}^{(i,j)(n)}\right\}_{Li+l}}{\left\{E[\underline{\underline{S}}]^H E[\underline{\underline{S}}]\right\}_{Li+l,Li+l}}. \quad (31)$$

Using $\mathcal{P}$ instead of $p(\underline{\underline{S}}|\mathcal{R}, \hat{\mathcal{B}}^{(n-1)})$, the unbiasedness of the UECM estimator can be proven recursively.

First, it is shown that the UECM estimate $\hat{h}_{l,m,\mathrm{UECM}}^{(i,j)(n)}$ is unbiased if the estimates in $\underline{\tilde{h}}_{l,m}^{(i,j)(n)}$ are unbiased. Let us calculate the mathematical expectation of $\hat{h}_{l,m,\mathrm{UECM}}^{(i,j)(n)}$, assuming that the noise is zero-mean and using (31) and (7)

$$E\left[\hat{h}_{l,m,\mathrm{UECM}}^{(i,j)(n)}\right] = \frac{\left\{E[\underline{\underline{S}}]^H E[\underline{\underline{S}}]\left(\underline{h}_m^{(j)} - E\left[\underline{\tilde{h}}_{l,m}^{(i,j)(n)}\right]\right)\right\}_{Li+l}}{\{E[\underline{\underline{S}}]^H E[\underline{\underline{S}}]\}_{Li+l,Li+l}}. \quad (32)$$

At this stage of the demonstration, we assume that the elements of $\underline{\tilde{h}}_{l,m}^{(i,j)(n)}$ are unbiased. $E[\underline{\tilde{h}}_{l,m}^{(i,j)(n)}] = \underline{h}_m^{(j)} - \underline{e}^{(Li+l)} h_{l,m}^{(i,j)}$, where $\underline{e}^{(Li+l)}$ denotes a length-$n_T L$ vector filled with zeros except for the $Li + l$th element, which is 1. Then (32) yields

$$E\left[\hat{h}_{l,m,\mathrm{UECM}}^{(i,j)(n)}\right] = \frac{\{E[\underline{\underline{S}}]^H E[\underline{\underline{S}}]\}_{Li+l,Li+l}}{\{E[\underline{\underline{S}}]^H E[\underline{\underline{S}}]\}_{Li+l,Li+l}} h_{l,m}^{(i,j)} \quad (33)$$

which shows that the UECM estimate $\hat{h}_{l,m,\mathrm{UECM}}^{(i,j)(n)}$ is unbiased, provided that the previous estimates, i.e., $\underline{\tilde{h}}_{l,m}^{(i,j)(n)}$ are unbiased, too.

If the initial estimates $\underline{\tilde{h}}_{-L_1,m}^{(i,j)(1)}$ are unbiased, then $\hat{h}_{-L_1,m,\mathrm{UECM}}^{(i,j)(1)}$ is unbiased, and thus, $\underline{\tilde{h}}_{-L_1+1,m}^{(i,j)(1)}$ only contains unbiased estimates. The same reasoning can be applied till any $\hat{h}_{l,m,\mathrm{UECM}}^{(i,j)(n)}$. The only problem left is to find unbiased initial estimates. They can be obtained through a DA ML estimation at the first iteration, for instance.

3) The Diagonal Approximation: In [6], it is proposed to keep only the diagonal elements of matrix $E[\underline{\underline{S}}^H \underline{\underline{S}}|\mathcal{R}, \hat{\mathcal{B}}^{(n-1)}]$ to provide another low-complexity solution. This "diagonal approximation" leads to a linear complexity ($\propto n_T L$) instead of a cubic one ($\propto n_T^3 L^3$) for the matrix inversion in (17). Table I shows that the EM with the diagonal approximation is much less complex than the EM and the ECM. Unfortunately, its performance reported in Section IV is much weaker, too. Actually, even if the symbols are *a priori* independent, they are correlated when they are conditioned on the observations $\mathcal{R}$ and the channel estimate $\hat{\mathcal{B}}^{(n-1)}$.

4) The DD Estimator: The EM, UEM, ECM, and UECM 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, which is well known already [1, pp. 772–782]. The equations for that DD estimator are identical to (9)

$$\underline{\hat{h}}_{m,\mathrm{DD}}^{(j)} = (\underline{\underline{\tilde{S}}}^H \underline{\underline{\tilde{S}}})^{-1} \underline{\underline{\tilde{S}}}^H \underline{r}_m^{(j)} \quad (34)$$

and (10)

$$\hat{\sigma}_{n,\mathrm{DD}}^2 = \sum_{j=1}^{n_R} \sum_{m=0}^{M_s-1} \frac{1}{n_R M_s L_r} \left(\underline{r}_m^{(j)} - \underline{\underline{\tilde{S}}}\underline{\hat{h}}_m^{(j)}\right)^H \left(\underline{r}_m^{(j)} - \underline{\underline{\tilde{S}}}\underline{\hat{h}}_m^{(j)}\right) \quad (35)$$

but the pilot-symbol matrix $\underline{\underline{S}}_p$ is replaced by $\underline{\underline{\tilde{S}}}$, which is filled with the hard decisions on the symbols. The DD version of the UECM is equal to the DD ECM. Their equation can be easily derived from (30). Finally, the "diagonal approximation" can also be applied to the DD ML estimator to lower the complexity of (34).

III. THE TURBO MIMO SYSTEM

A. The System Model

In order to compare the performance of the various channel estimators, the estimation algorithms are applied to a MIMO system with a turbo space–time equalizer. The transmitter scheme is depicted in Fig. 2. 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 frame of interleaved coded bits is then split into n_T subblocks (corresponding to the n_T transmit antennas). Within each of these subblocks, the bits are grouped and mapped to one of the M possible complex symbols in the considered multilevel/phase constellation $\mathcal{S}$. On each transmit antenna i ($i = 1, \ldots, n_T$), the resulting frame of complex symbols $s_k^{(i)} \in \mathcal{S}(k = 1, \ldots, L_s)$ is transmitted over the wireless channel. The symbols are zero mean and have variance σ_s^2. After pulse shaping with a square-root Nyquist filter $g(t)$, the n_T symbols $s_k^{(1)}, \ldots, s_k^{(n_T)}$ are sent at time kT.

We assume that the transmitter does not have any knowledge of the CSI, so that the various antennas send signals with identical powers. They also use the same shaping filter, modulation, and mapping rule.

The channel is assumed to be quasi-static, i.e., the channel remains static over one burst length. The frequency-selective MIMO fading channel is described by the tapped-delay line model. The different paths are characterized by delays and gains

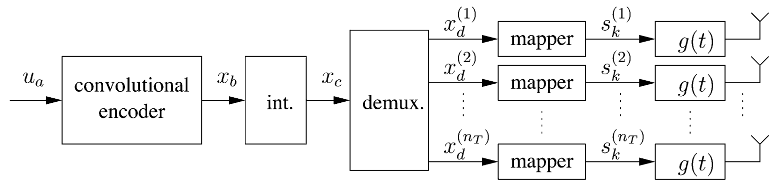


Fig. 2. Transmitter scheme.

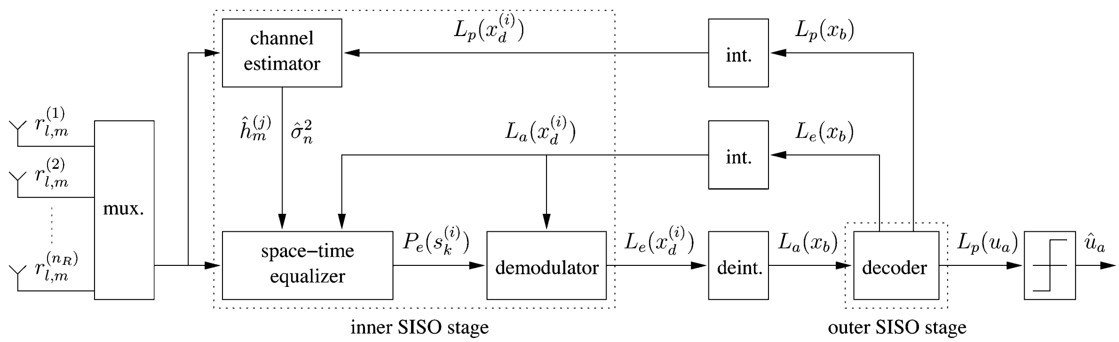


Fig. 3. Receiver scheme.

(independent zero-mean complex Gaussian random variables with tap-specific variances). The lowpass-equivalent CIR (including the pulse-shaping filter) between transmit antenna i and receive antenna $j\,(j = 1, \ldots, n_R)$ is denoted by $h^{(i,j)}(t)$. We assume that $h^{(i,j)}(t)$ can be truncated without loss of accuracy. We only keep its values $\forall i, \forall j$ from time $-L_1 T$ till $L_2 T$.

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. 3, 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 ratios (LLRs). L_a, L_e, and L_p, respectively, denote the *a priori*, extrinsic, and *a posteriori* bit LLRs.

The equalizer/demodulator, the inner SISO stage, has to mitigate intersymbol interference (ISI) and properly demodulate the symbols. Because of its low complexity, we use the MMSE-based equalizer described in [20]. This equalizer is fractionally spaced, which avoids the cascade of global matched filter and noise-whitening filter. Its design relies on an approximation [17] of the optimal MMSE receiver [24]. On the basis of the received samples, the channel parameter estimates, and the bit *a priori* probabilities, the equalizer outputs symbol extrinsic probabilities $P_e(s_k^{(i)})\,(k = 1, \ldots, L_s, i = 1, \ldots, n_T)$. Then the demodulator computes the bit extrinsic LLRs. According to the turbo principle, the equalizer exchanges soft extrinsic information with the SISO decoder, which is implemented by means of the BCJR algorithm [22].

B. Channel Estimation Applied to the Turbo MIMO System

We assumed in Section II that the receiver is able to compute APPs on the symbols. In fact, the receiver should be doubly iterative to compute those APPs. Inside each EM iteration n, the receiver should perform several turbo equalization/decoding iterations, keeping the channel estimate $\hat{\mathcal{B}}^{(n-1)}$. It has not yet been proved, but it is widely believed that after convergence of this turbo-equalization process, the probabilities at the decoder output would be equal to $p(\underline{\underline{S}}|\mathcal{R}, \hat{\mathcal{B}}^{(n-1)})$. For complexity reasons, the considered receiver performs only one pass through the equalizer/decoder inside each EM iteration. So, at EM iteration $n, p(\underline{\underline{S}}|\mathcal{R}, \hat{\mathcal{B}}^{(n-1)})$, and thus $E[\underline{\underline{S}}|\mathcal{R}, \hat{\mathcal{B}}^{(n-1)}]$ and $E[\underline{\underline{S}}^H \underline{\underline{S}}|\mathcal{R}, \hat{\mathcal{B}}^{(n-1)}]$, are approximated by taking the APPs delivered by the decoder after one turbo equalization/decoding iteration using the channel estimate $\hat{\mathcal{B}}^{(n-1)}$. Moreover, the joint probability $p(s_k^{(i)}, s_{k'}^{(i')}|\mathcal{R}, \hat{\mathcal{B}}^{(n-1)})$ is approximated by the product of the marginal probabilities $p(s_k^{(i)}|\mathcal{R}, \hat{\mathcal{B}}^{(n-1)})p(s_{k'}^{(i')}|\mathcal{R}, \hat{\mathcal{B}}^{(n-1)})\ \forall k' \neq k$ or $\forall i' \neq i$. Because of these approximations, the conditions for the proof [16] of the EM-algorithm convergence are not fulfilled. The convergence of the EM to the ML estimate can thus not be ensured.

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. That supplies an initial CSI to the equalizer. With appropriate interleaving/deinterleaving, the equalizer/demodulator and the decoder exchange soft information about the coded bits. For the next EM iterations, the EM algorithm or one of its variants is applied, using the training sequence and the APPs on the symbols delivered by the decoder. Note that in the case of the DD channel estimator, the hard decisions are taken on the basis of the *a posteriori* LLRs at the decoder output. The channel estimator can then deliver a novel refined CSI to the equalizer, and so on.

IV. Simulation Results

To assess the relative performance of the various channel-estimation algorithms, simulations have been conducted over two different MIMO channels. The channel estimator estimates the CIR and the noise variance. Unless otherwise stated, the

UDA and HEM noise-variance estimators are, respectively, used with the DA and UEM channel estimators.

In each considered simulation scenario, the channel remains static during a frame period. In order to cover the whole statistics of channel values, we average the results over a sufficient number of channel realizations. At each considered E_b/N_0, the simulations have been stopped after at least 100 frames corrupted by one or several bit errors. The BER has been measured, as well as the normalized MSE of the CIR estimator. The normalized MSE is the mean value of the square distance between the channel-tap estimate and its true value, divided by the mean total power of the CIR. The normalized MSE is the average of the normalized square error (NSE)

$$\text{NSE} = \frac{\sum_{j=1}^{n_R}\sum_{i=1}^{n_T}\sum_{m=0}^{M_s}\sum_{l=-L_1}^{L_2}\left|\hat{h}_{l,m}^{(i,j)} - h_{l,m}^{(i,j)}\right|^2}{n_R n_T M_s L \sum_{j=1}^{n_R}\sum_{i=1}^{n_T}\sum_{m=0}^{M_s}\sum_{l=-L_1}^{L_2} E\left[\left|h_{l,m}^{(i,j)}\right|^2\right]} \quad (36)$$

over many frames.

The first considered channel is a four-input four-output flat Rayleigh-fading channel. All the delays between a transmit antenna and a receive one are identical, but the fading coefficients are independently distributed. For this particular channel only, the estimator does not estimate the concatenation of the physical channel with the pulse-shaping filter but the physical channel only, i.e., 16 complex coefficients. To cope with the pulse-shaping filter, a filter matched to it is placed in front of the receiver. The sampling is performed thanks to a synchronizer which is assumed to be ideal. There is no oversampling in this case: $M_s = 1$.

The other parameters of the simulation setup are the following. The modulation is 8-phase-shift keying (PSK) with Gray mapping. The interleaver is random. A rate-0.5 convolutional encoder with generator polynomials $[23_8, 35_8]$ is considered. The frames are made up of 2000 information bits, which correspond to 1336 8-PSK symbols. We assume that we are in acquisition mode, which means that the initial CSI estimates are obtained thanks to 4×5 pilot symbols placed at the beginning of each burst. The training sequence is designed according to [25] in order to present interesting orthogonality properties.

Fig. 4 reports the normalized MSE obtained after 10 iterations of the whole receiver, as a function of the average E_b/N_0 per receive antenna. The curve labeled "CRB" is the Cramer–Rao bound (CRB) when all the symbols of the frame are known, that is to say the DA CRB. The MSE of the UEM-HEM estimator, i.e., the UEM channel estimator with the HEM noise-variance estimator, reaches that bound for E_b/N_0 above 5–6 dB. The DD estimator is also close to the CRB. The MSEs of the three other EM-based estimators are between those of the DD and DA estimators. All the iterative algorithms seem to perform better than the DA one, whose estimates are not iteratively refined. Regarding the noise variance, the EM estimator overestimates it more than the HEM one, as expected. For a $E_b/N_0 = 7$ dB, the bias of the EM-EM and UEM-EM estimators is 3 dB at iteration 2 and around 0.5–1.5 dB at iteration 10, whereas the bias of the

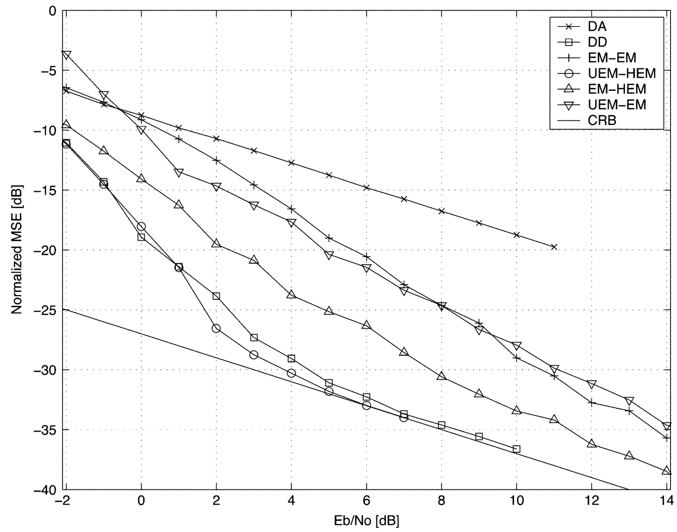


Fig. 4. Normalized MSE of the channel-tap estimates after 10 iterations for a 4×4 flat Rayleigh-fading channel.

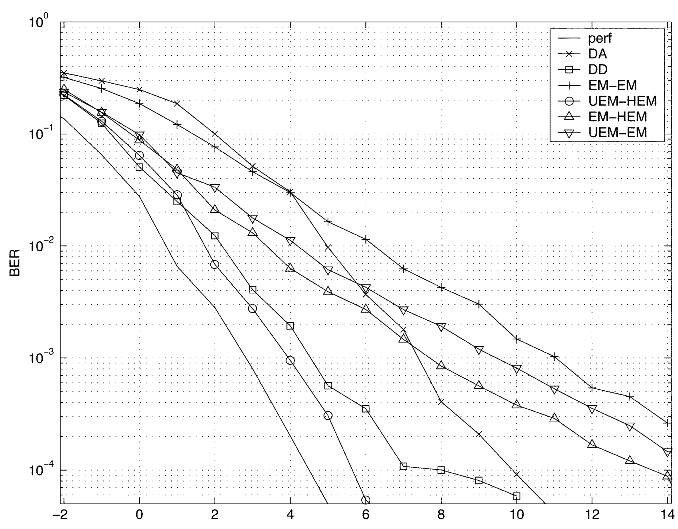


Fig. 5. BER after 10 iterations for a 4×4 flat Rayleigh-fading channel with several estimation methods.

EM-HEM and UEM-HEM estimators is only 2 dB at iteration 2 and around 0.05 dB at iteration 10.

What is more interesting than the MSE is the receiver BER. The BER at iteration 10 is plotted in Fig. 5. The curve labeled "perf" shows the BER when the true CIR and noise variance are provided to the equalizer. It can be considered as a lower bound on the results obtained with channel-estimation algorithms. It should be mentioned that when pilots are used (all curves but "perf"), the E_b should be corrected because the power is spread over pilot and information bits. This correction is not made in the figures. There is thus a small penalty for the corresponding curves of Fig. 5 that should be shifted to the right by 0.065 dB. In Fig. 5, the DD estimator performs very well, except for BERs below 10^{-4}. The best results are obtained by the UEM-HEM receiver. The EM-EM, EM-HEM, and UEM-EM algorithms are even worse than the DA ML method, in spite of their better

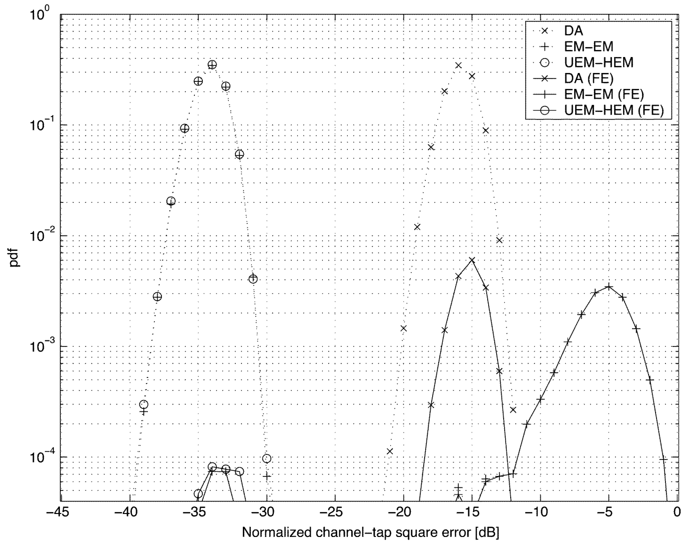


Fig. 6. Distribution of the NSE of the channel-tap estimates after 10 iterations for a 4×4 flat Rayleigh-fading channel. $E_b/N_0 = 7$ dB. Dotted curves: global pdf. Solid curves: pdf when an FE occurs.

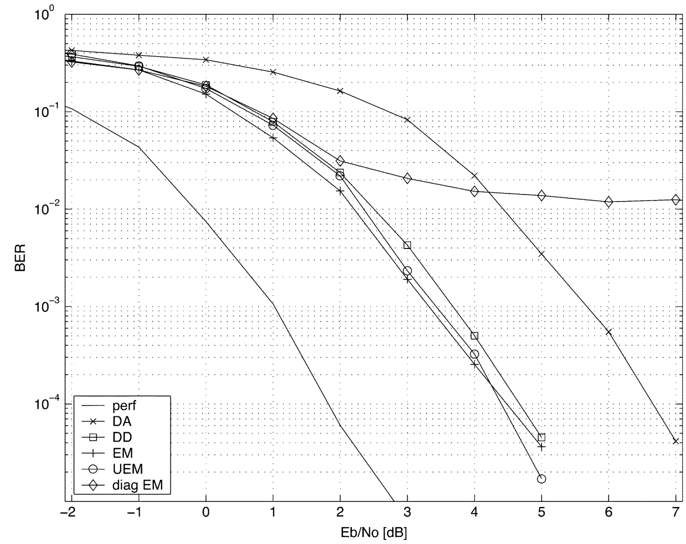


Fig. 7. BER after six iterations for a 4×4 frequency-selective Rayleigh-fading channel (GSM TU profile).

TABLE II
GSM TU CHANNEL MODEL

Path Delay (μs)	0.0	0.2	0.5	1.6	2.3	5.0
Path Power (dB)	−3.0	0.0	−2.0	−6.0	−8.0	−10.0

MSEs. Except for this major difference, the BERs in Fig. 5 are coherent with the MSEs observed in Fig. 4.

To understand the poor BERs obtained with the EM-EM, EM-HEM, and UEM-EM receivers, we have to look at the distributions of the NSE of the channel-tap estimates, which is defined in (36). In Fig. 6, dotted curves represent p(NSE), i.e., the probability density function (pdf) of the NSE. Let FE denote the frame-error event, which occurs when the decision made on the frame is corrupted by one or several information-bit errors. p(FE) is thus equal to the frame-error rate (FER). The solid curves in Fig. 6 represent $p(\text{NSE}|\text{FE})p(\text{FE})$. In another way, they represent (up to a scaling factor) the pdf of the square error knowing that the frame is in error. Those pdfs have been measured for the DA, EM-EM, and UEM-HEM estimators over 200 000 frames at iteration 10 and $E_b/N_0 = 7$ dB. The peaks of the DA and UEM-HEM square-error distributions are around the MSE of the corresponding estimators. However, the square error of the EM-EM estimator is distributed in another way. It appears that the EM estimator sometimes diverges. We say that an estimator diverges when the estimate square error at the last iteration is larger than the square error of the initial estimate. The maximum of the EM-EM square-error distribution is located around −34 dB like the UEM-HEM one, but there is also a lower peak around −5 dB. This second peak is formed by estimates that have diverged from their initial values, since the NSE of the DA estimator is smaller than −12 dB. Around this "divergence peak," the EM dotted and solid curves overlap. That means that when the EM estimator diverges, the frame is erroneously detected. Although the EM estimator diverges for only 1.6% of the frames in Fig. 6, this divergence is responsible for 97% of the frame errors. The DD estimator, whose square-error distribution is not reported here, diverges sometimes too, but it happens 25 times less often than the EM-EM estimator. It can be concluded that for the flat Rayleigh MIMO channel and the transmission setup used here, the UEM estimator provides excellent results. The EM performance is disappointing, but for flat-fading channels it has been proposed in [6] to combine the estimate from the training sequence and the estimate from the EM estimator. In some cases, this "combining" method can improve the EM performance. The extension of this method to frequency-selective channels is a topic for future research.

Secondly, the same scheme has also been investigated for a four-input four-output frequency-selective channel model called "GSM typical urban (TU)." The code, the frame length, the modulation, and the mapping rule are the same as in the flat-fading case. We assume again that we are in acquisition mode. The number of pilot symbols is 4×55 and the penalty left out of account in Fig. 7 is 1.15 dB. Table II shows the path power and delay profiles of this channel. Each one of the 4×4 physical multipath impulse responses has taps selected according to these profiles. The symbol period T is equal to 3.7 μs. Each tap is modeled as an independent (of all others) zero-mean complex Gaussian random variable. The space–time equalizer span is equal to 11 symbol periods and $M_s = 2$.

The results after six iterations are given in Fig. 7. All the considered iterative algorithms perform better than the DA ML method, except the EM with the diagonal approximation, whose curve is labeled "diag EM." Among the lower complexity algorithms, i.e., the ECM and the EM with the diagonal approximation, the ECM is, by far, better. For the sake of the figure readability, the curve corresponding to the ECM has not been plotted, but it is between the EM and the DD curves. Note that no optimization of the training-sequence length has been performed. It has been chosen as the smallest number of pilot symbols which still enables CIR and noise-variance acquisition, i.e., 4×55 pilot symbols, in this case.

V. CONCLUSION

We have proposed an UEM algorithm for CIR estimation. We have also proposed a lower complexity channel-estimation method based on the ECM algorithm and an unbiased version of this latter technique. These algorithms have been integrated in a turbo receiver in order to compare them with known estimators, like the EM one [5]–[8] and the classical DA ML criterion. Simulations conducted over MIMO channels show that in most cases, the iterative CIR estimation techniques outperform the classical DA ML channel estimator. It is worth refining the channel-parameter estimates through the iterations, because it lowers the CIR-estimate MSE, and above all, it lowers the BER. The DD estimator performance is often very good. However, the proposed UEM algorithm provides the best results among all considered estimators in both simulated scenarios. The ECM and the UECM methods, respectively, give results very close to the EM and the UEM techniques. The "diagonal approximations" of the EM and UEM algorithms have a lower complexity than the ECM-based techniques, but these ECM-based methods outperform the algorithms with the diagonal approximation.

The algorithms considered in this paper do not handle a channel variation inside a burst length. Further work will be done on this topic. The optimization of the number of pilot symbols is also a topic for further research.

APPENDIX

The DA ML noise-variance estimate (10) is unbiased if the channel taps are known. However, by analogy with the joint estimation of the mean and 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 (9) and the fact that $(\underline{\underline{S}}_p^H \underline{\underline{S}}_p)^H = \underline{\underline{S}}_p^H \underline{\underline{S}}_p$, it is straightforward to show that

$$\left(\underline{r}_m^{(j)} - \underline{\underline{S}}_p \underline{h}_m^{(j)}\right)^H \left(\underline{r}_m^{(j)} - \underline{\underline{S}}_p \underline{h}_m^{(j)}\right) \tag{37}$$

$$= \left(\underline{r}_m^{(j)} - \underline{\underline{S}}_p \hat{\underline{h}}_m^{(j)}\right)^H \left(\underline{r}_m^{(j)} - \underline{\underline{S}}_p \hat{\underline{h}}_m^{(j)}\right) \tag{38}$$

$$+ \left(\underline{\underline{S}}_p \hat{\underline{h}}_m^{(j)} - \underline{\underline{S}}_p \underline{h}_m^{(j)}\right)^H \left(\underline{\underline{S}}_p \hat{\underline{h}}_m^{(j)} - \underline{\underline{S}}_p \underline{h}_m^{(j)}\right). \tag{39}$$

With this relation, the expectation of (10) can be expressed as

$$E\left[\hat{\sigma}_{n,\mathrm{DA}}^2\right] = \frac{n_R M_s L_r \sigma_n^2}{n_R M_s L_r} - \sum_{j=1}^{n_R} \sum_{m=0}^{M_s-1} \frac{E\left[\left(\hat{\underline{h}}_m^{(j)} - \underline{h}_m^{(j)}\right)^H \underline{\underline{S}}_p^H \underline{\underline{S}}_p \left(\hat{\underline{h}}_m^{(j)} - \underline{h}_m^{(j)}\right)\right]}{n_R M_s L_r}. \tag{40}$$

If the pilot symbols are chosen according to [25], we get

$$\underline{\underline{S}}_p^H \underline{\underline{S}}_p = L_r \sigma_s^2 \underline{\underline{I}}. \tag{41}$$

The variance of the DA ML channel-tap estimator can be easily calculated. This variance is

$$E\left[\left(\hat{h}_{l,m}^{(i,j)} - h_{l,m}^{(i,j)}\right)^H \left(\hat{h}_{l,m}^{(i,j)} - h_{l,m}^{(i,j)}\right)\right] = \frac{\sigma_n^2}{L_r \sigma_s^2}. \tag{42}$$

Using the last two relations, (40) becomes

$$\begin{aligned} E\left[\hat{\sigma}_{n,\mathrm{DA}}^2\right] &= \sigma_n^2 - \frac{n_R M_s n_T L}{n_R M_s L_r} L_r \sigma_s^2 \frac{\sigma_n^2}{L_r \sigma_s^2} \\ &= \frac{L_r - n_T L}{L_r} \sigma_n^2. \end{aligned} \tag{43}$$

The bias shown by (43) is negligible when $L_r \gg n_T L$. The estimator is asymptotically unbiased. Usually, L_r is not much larger than $n_T L$ during the acquisition stage, because there are only a few pilot symbols. To unbias the estimator, (10) has to be multiplied by $(L_r)/(L_r - n_T L)$.

REFERENCES

[1] H. Meyr, M. Moeneclaey, and S. A. Fechtel, "Digital communication receivers: Synchronization, channel estimation and signal processing," in *Wiley Series in Telecommunications and Signal Processing*. New York: Wiley, 1998.

[2] T. L. Marzetta, "Blast training: Estimating channel characteristics for high-capacity space-time wireless," in *Proc. 37th Annu. Allerton Conf. Commun., Control, Comput.*, Monticello, IL, Sep. 1999, pp. 958–966.

[3] J. Baltersee, G. Fock, and H. Meyr, "Achievable rate of MIMO channels with data-aided channel estimation and perfect interleaving," *IEEE J. Sel. Areas Commun.*, vol. 19, no. 9, pp. 2358–2368, Dec. 2001.

[4] D. Samardzija and N. Mandayam, "Pilot-assisted estimation of MIMO fading channel response and achievable data rates," *IEEE Trans. Signal Process.*, vol. 51, no. 11, pp. 2882–2890, Nov. 2003.

[5] H. Zamiri-Jafarian and S. Pasupathy, "EM-based recursive estimation of channel parameters," *IEEE Trans. Commun.*, vol. 47, no. 9, pp. 1297–1302, Sep. 1999.

[6] 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. 6, pp. 1450–1460, Aug. 2001.

[7] A. O. Berthet, B. S. Ünal, and R. Visoz, "Iterative decoding of convolutionally encoded signals over multipath Rayleigh fading channels," *IEEE J. Sel. Areas Commun.*, vol. 19, no. 7, pp. 1729–1743, Sep. 2001.

[8] C. H. Aldana, E. de Carvalho, and J. M. Cioffi, "Channel estimation for multicarrier multiple input single output systems using the EM algorithm," *IEEE Trans. Signal Process.*, vol. 51, no. 12, pp. 3280–3292, Dec. 2003.

[9] M. Tüchler, R. Otnes, and A. Schmidbauer, "Performance of soft iterative channel estimation in turbo equalization," in *Proc. Int. Conf. Commun.*, New York, NY, May 2002, pp. 1858–1862.

[10] R. Otnes and M. Tüchler, "Soft iterative channel estimation for turbo equalization: Comparison of channel estimation algorithms," in *Proc. Int. Conf. Commun. Syst.*, Singapore, Nov. 2002, pp. 72–76.

[11] R. Otnes and M. Tüchler, "On iterative equalization, estimation, and decoding," in *Proc. Int. Conf. Commun.*, Anchorage, AK, May 2003, pp. 2958–2962.

[12] C. N. Georghiades and J. C. Han, "Sequence estimation in the presence of random parameters via the EM algorithm," *IEEE Trans. Commun.*, vol. 45, no. 3, pp. 300–308, Mar. 1997.

[13] Y. Li, C. N. Georghiades, and G. Huang, "Iterative maximum-likelihood sequence estimation for space-time coded systems," *IEEE Trans. Commun.*, vol. 49, no. 6, pp. 948–951, Jun. 2001.

[14] B. Lu, X. Wang, and Y. G. Li, "Iterative receivers for space-time block-coded OFDM systems in dispersive fading channels," *IEEE Trans. Wireless Commun.*, vol. 1, no. 2, pp. 213–225, Mar. 2002.

[15] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum-likelihood from incomplete data via the EM algorithm," *J. Roy. Statist. Soc.*, ser. B, vol. 39, no. 1, pp. 1–38, Jan. 1977.

[16] C. F. J. Wu, "On the convergence properties of the EM algorithm," *Ann. Statist.*, vol. 11, no. 1, pp. 95–103, 1983.

[17] M. Tüchler, A. Singer, and R. Koetter, "Minimum mean squared error equalization using *a priori* information," *IEEE Trans. Signal Process.*, vol. 50, no. 3, pp. 673–683, Mar. 2002.

[18] X. L. Meng and D. B. Rubin, "Maximum likelihood estimation via the ECM algorithm: A general framework," *Biometrika*, vol. 80, pp. 267–278, 1993.

[19] G. J. McLachlan and T. Krishnan, "The EM algorithm and extensions," in *Wiley Series in Probability and Statistics*. New York: Wiley, 1997.

[20] X. Wautelet, A. Dejonghe, and L. Vandendorpe, "MMSE-based fractional turbo receiver for space-time BICM over frequency-selective MIMO fading channels," *IEEE Trans. Signal Process.*, vol. 52, no. 6, pp. 1804–1809, Jun. 2004.

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

[22] 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. IT-20, no. 2, pp. 284–287, Mar. 1974.

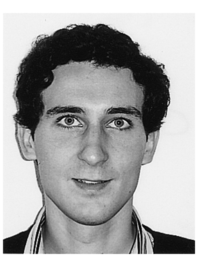
[23] J. M. Mendel, "Lessons in estimation theory for signal processing, communications, and control," in *Prentice-Hall Signal Processing Series*. Englewood Cliffs, NJ: Prentice-Hall, 1995.

[24] X. Wang and H. V. Poor, "Iterative (turbo) soft interference cancellation and decoding for coded CDMA," *IEEE Trans. Commun.*, vol. 47, no. 7, pp. 1046–1061, Jul. 1999.

[25] G. Caire and U. Mitra, "Structured multiuser channel estimation for block-synchronous DS/CDMA," *IEEE Trans. Commun.*, vol. 49, no. 9, pp. 1605–1617, Sep. 2001.

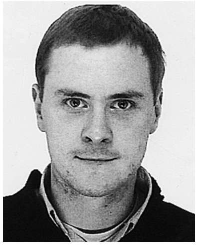
Xavier Wautelet was born in Namur, Belgium, in 1978. He received the electrical engineering degree and the Ph.D. degree from the Université catholique de Louvain (UCL), Louvain-la-Neuve, Belgium, in 2001 and 2006, respectively.

He has been a Research Assistant with the Communications and Remote Sensing Laboratory of UCL since 2001. His Ph.D. thesis dealt with turbo equalization and iterative channel estimation for wireless MIMO channels. Since May 2006, he has worked on a project on DSL systems in collaboration with Alcatel Bell, Antwerpen, Belgium.



Cédric Herzet was born in Verviers, Belgium in 1978. He received the electrical engineering degree and the Ph.D. degree in applied science from the Université catholique de Louvain (UCL), Louvain-la-Neuve, Belgium, in 2001 and 2006, respectively. His graduation thesis concerned multiuser turbo receivers for wideband code-division multiple-access transmissions. His Ph.D. thesis dealt with code-aided synchronization for digital burst communications.

From August 2001 to April 2006, he was a Research Assistant in the Communications and Remote Sensing Laboratory of UCL in the digital communications group (DIGICOM). From Augustus 2001 to December 2001, he worked on the topic of VDSL systems in collaboration with Alcatel Bell, Antwerp, Belgium. From January 2002 to April 2006, his research concerned synchronization and parameter estimation in iterative receivers. From May to August 2006, he was a Postdoctoral researcher with the Ecole Normale Supérieure de Cachan, France. He is currently a Fulbright Postdoctoral researcher at the University of California, Berkeley.



Antoine Dejonghe received the M.Sc. degree in electrical engineering (*summa cum laude*) and the Ph.D. degree from the Université Catholique de Louvain (UCL), Louvain-la-Neuve, Belgium, in 2000 and 2004, respectively.

He was with the Communications and Remote Sensing Laboratory of UCL from 2000 to 2004, as a Research Fellow of the Belgian National Fund for Scientific Research (FNRS). Since 2004, he has been a Senior Research Engineer in the Wireless Research Group of the Interuniversity Microelectronics Center (IMEC), Leuven, Belgium. His work in the past focused on digital communication system design, with a special emphasis on iterative detection and decoding schemes relying on the so-called turbo principle. He is currently carrying and coordinating research activities in the field of cross-layer optimization for energy-aware multimedia wireless communications, building on SDR-based baseband engines and reconfigurable analog front-ends. He is also looking at the extension of these concepts to cognitive radio systems.

Jérome Louveaux (M'98) was born in Leuven, Belgium, in 1974. He received the electrical engineering degree and the Ph.D. degree from the Université catholique de Louvain (UCL), Louvain-la-Neuve, Belgium, in 1996 and 2000, respectively.

From 2000 to 2001, he was a Visiting Scholar in the Electrical Engineering Department, Stanford University, Stanford, CA. From 2004 to 2005, he was a Postdoctoral Researcher with the Technische Universiteit Delft, Delft, The Netherlands, working on the European project U-BROAD. He is currently an Assistant Professor in the Electrical Engineering Deptartment of UCL. His research interests are in signal processing for digital communications, mainly synchronization, estimation and equalization issues for high-bit-rate transmission over wired channels.

Dr. Louveaux has served as an Associate Editor for the IEEE COMMUNICATIONS LETTERS since 2003. He is a corecipient of the "Prix biennal Siemens 2000" and the "Prix Scientifique Alcatel 2005."

Luc Vandendorpe (M'93–SM'99–F'06) was born in Mouscron, Belgium, in 1962. He received the electrical engineering Degree (summa cum laude) and the Ph.D. degree from the Université catholique de Louvain (UCL), Louvain-la-Neuve, Belgium, in 1985 and 1991, respectively.

Since 1985, he has been with the Communications and Remote Sensing Laboratory of UCL, where he first worked in the field of bit-rate-reduction techniques for video coding. In 1992, he was a Visiting Scientist and Research Fellow at the Telecommunications and Traffic Control Systems Group of the Delft Technical University, The Netherlands, where he worked on Spread Spectrum Techniques for Personal Communications Systems. From October 1992 to August 1997, he was a Senior Research Associate of the Belgian National Science Foundation (NSF) at UCL, and an invited Assistant Professor. Presently, he is a Professor and Head of the EE department of UCL. He is mainly interested in digital communication systems: equalization, joint detection/synchronization for CDMA, OFDM (multicarrier), MIMO, and turbo-based communications systems (UMTS, xDSL, WLAN, etc.) and joint source/channel (de)coding.

In 1990, Dr. Vandendorpe was a corecipient of the Biennal Alcatel-Bell Award from the Belgian NSF for a contribution in the field of image coding. In 2000, he was a corecipient (with J. Louveaux and F. Deryck) of the Biennal Siemens Award from the Belgian NSF for a contribution on filterbank-based multicarrier transmission. In 2004, he was a co winner (with J. Czyz) of the Face Authentication Competition, FAC 2004. He is or has been a TPC member for IEEE VTC Fall 1999, IEEE Globecom 2003 Communications Theory Symposium, the 2003 Turbo Symposium, IEEE VTC Fall 2003, IEEE SPAWC 2005, and IEEE SPAWC 2006. He is co-technical chair (with P. Duhamel) for IEEE ICASSP 2006. He was an Editor of the IEEE TRANSACTIONS ON COMMUNICATIONS for Synchronization and Equalization between 2000 and 2002, and Associate Editor of the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS between 2003 and 2005. He was Chair of the IEEE Benelux joint chapter on Communications and Vehicular Technology between 1999 and 2003. He was an elected member of the Signal Processing for Communications committee between 2000 and 2005. Currently, he is an elected member of the Sensor Array and Multichannel Signal Processing committee of the Signal Processing Society, and an Associate Editor of the IEEE TRANSACTIONS ON SIGNAL PROCESSING.