Frederik Simoens, Henk Wymeersch
Nele Noels, Heidi Steendam, Marc Moeneclaey
TELIN Dept., Ghent University
St. Pietersnieuwstraat 44, B - 9000 Gent
e-mail:firstName.lastName@telin.UGent.be

# TURBO CHANNEL ESTIMATION FOR BIT-INTERLEAVED CODED MODULATION

***Abstract*— This contribution deals with code-aided channel estimation for bit-interleaved coded modulation (BICM) for single- and multi-antenna systems. Based on the Expectation-Maximization algorithm, a low-complexity estimator is derived. The receiver iterates between estimation and detection and operates by accepting soft information from the detector. Through computer simulations the proposed estimator, and variations thereof, are investigated in terms of computational complexity, BER performance and mean square error (MSE) performance. Furthermore, the impact of the symbol mapping is investigated.**

## I. Introduction

Bit-interleaved coded modulation (BICM) is a popular technique that combines error-correcting codes with high-order modulation schemes. Although originally developed for fading single-input single-output (SISO) channels [1], [2], BICM has quickly found its way in state-of-the-art multi-antenna systems (MIMO systems) [3]. It is now a serious contender against space-time (ST) codes, which exploit the spatial diversity in rich scattering MIMO environments at a cost of reduced transmission rates. BICM, on the other hand, can combine high data rates while still maintaining high diversity [4].

In its original form, BICM combines a convolutional encoder and a symbol mapper, separated by a pseudo-random interleaver. The interleaver spreads adjacent coded bits over different symbols, thus increasing the time-diversity. Later, BICM was extended by means of the turbo-principle [5]: by considering the symbol mapper as an additional code, one can iterate at the receiver between the demapper and the decoder. This is known as BICM-ID. It was shown in [6] that good mappings in conventional systems (such as Gray mapping), can be outperformed in BICM-ID by different, non-standard mappings, albeit only above a certain SNR and at the cost of increased computational complexity. Extension to MIMO systems is straightforward: the convolutional encoder, bit-interleaver and symbol-mapper are now followed by a serial-to-parallel (S/P) converter, which passes the coded symbols to the different antennas in a round-robin fashion. This form of vertical encoding has the potential of achieving full diversity as each information bit can be spread across all the transmit antennas. Again the mapping strategy is a crucial design parameter [7], [8]. Many variations on this basic theme has since been developed (e.g. [9]).

While BICM-ID systems have been shown to achieve excellent performance in a great variety of scenarios, the impact of channel estimation should be considered [10]. Standard channel estimation techniques are based on the presence of pilot (training) sequences at some point in the transmitted frame. As these pilots effectively reduce the transmission rate, they are wasteful. Hence, the number of pilot symbols should be low (compared to the overall frame length). Since the transmitted bits are commonly protected by powerful error-correcting codes, it makes sense to exploit the code properties during estimation. In fact, many technical papers have been devoted to the problem of code-aided channel estimation. Depending on the channel type (i.e., multi-path vs. flat fading and time-varying vs. static channels), different techniques are applied: iterative receivers with different coding schemes have been proposed using the Expectation Maximization (EM) algorithm for joint channel estimation and data detection in a MIMO context [3], [11]–[13]. The EM algorithm is utilized in [11] to perform joint noise variance and channel estimation for fast fading channels with space-time coding and MAP decoding. An MMSE channel estimator that exploits knowledge of the transmitted symbols via a modified Viterbi decoder is considered in [12] and [13]. In [3], joint MIMO channel estimation and MAP ST-BICM decoding is performed. Extensions to ST-OFDM have been proposed in [14]. Rather ad-hoc methods have been considered for iterative decoding and MIMO channel estimation, such as frequency selective channel estimation using decision feedback for space-time codes with Viterbi decoding [15] and MAP decoding [16].

Our mathematical framework is developed for flat-

fading MIMO channels, but is also applicable to SISO channels since the latter are just a special case of the former. For short burst data transmission, we can assume the channel to remain static during each burst. We derive an estimation algorithm, based on the iterative EM algorithm, that should, in principle, be able to achieve optimal Maximum Likelihood (ML) performance. In its exact formulation, the algorithm is still prohibitively complex. We will introduce two sub-optimal variations of the complex EM estimator. The first treats the data as uncoded during the estimation process. The second variation is obtained by making a simple approximation and results in an estimator that is truly *embedded* in the iterative detector. The proposed estimators combine excellent performance with fairly low computational overhead. This is verified through Monte Carlo simulations.

## II. System Model

We consider a system with $N_T$ transmit and $N_R$ receive antennas. The transmitter encodes a stream of $N_b$ bits using a rate $R$ convolutional encoder. The resulting $N_c = N_b/R$ coded bits are fed to an interleaver (denoted by $\Pi$) and mapped to an $M$-point signaling constellation $\Omega$. This mapping operates on blocks of $\log_2 M$ bits and results in a sequence of $N_S N_T = N_c/\log_2 M$ coded symbols. These are then multiplexed with $N_P N_T$ pilot symbols, delivered to the S/P converter. At each discrete time instant, we transmit $N_T$ symbols, i.e., one per transmit antenna. Hence, the resulting $N_T \times 1$ transmit-vector at time $k$, $\mathbf{a}_k$, depends on $N_T.\log_2 M$ bits. We denote these bits as $a_k[m]$, $m = 1, \ldots, N_T.\log_2 M$. The transmit-vector $\mathbf{a}_k$ is transmitted over the (unknown) MIMO channel with equivalent, discrete-time baseband $N_R \times N_T$ channel response matrix $\mathbf{H}$. This matrix is modeled as a block-fading channel with $h_{ij} = [\mathbf{H}]_{ij}$ a complex-valued Gaussian random variable; its real and imaginary parts are statistically independent with zero mean and variance equal to $1/2$, so that $E\left[|h_{ij}|^2\right] = 1$. We assume $\mathbf{H}$ is constant during a burst of $N_S + N_P$ symbol vectors, but can change independently from burst to burst. Adding thermal white Gaussian noise yields

$$\mathbf{r}_k = \mathbf{H}\mathbf{a}_k + \mathbf{n}_k \tag{1}$$

with $\mathbf{n}_k$ an $N_R \times 1$ vector of complex iid AWGN samples, with independent real and imaginary part, each having variance $\sigma^2$. After reception of the entire frame, we can write:

$$\mathbf{R} = \mathbf{H}\mathbf{A}+\mathbf{N} \tag{2}$$

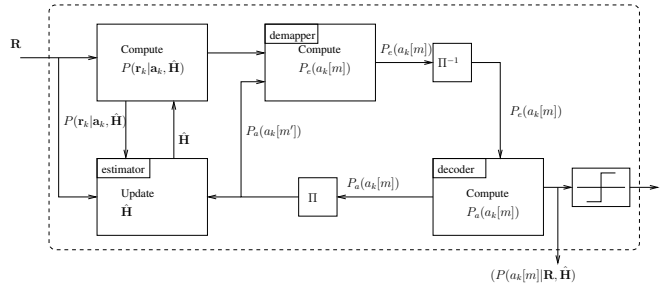


Fig. 1. BICM-ID detector with embedded estimator

$\mathbf{R} = [\mathbf{r}_0, \ldots, \mathbf{r}_{N_P+N_S-1}]$, $\mathbf{A} = [\mathbf{a}_0, \ldots, \mathbf{a}_{N_P+N_S-1}]$, $\mathbf{N} = [\mathbf{n}_0, \ldots, \mathbf{n}_{N_P+N_S-1}]$. The goal of the receiver is to recover the information bits. In order to do this, the channel matrix $\mathbf{H}$ needs to be estimated.

## III. BICM-ID detector

The detector iterates between demapping and decoding according to the turbo principle. For SISO channels, such iterative receivers are well known [17]. The extension to BICM MIMO systems has been addressed in [3], [18], so we will only cover the parts essential to this paper. Using turbo terminology, $P^e(x)$, $P^a(x)$ and $P^p(x) = P\left(x \middle| \mathbf{R}, \hat{\mathbf{H}}\right)$ refer to the *a priori*, *extrinsic* and *a posteriori* probabilities of the random variable $x$, respectively.

The operation of the iterative turbo detector is described in Algorithm 1 and depicted in Fig. 1. At each iteration, we discern two stages: demapping and decoding. First, the extrinsic probabilities of the coded bits $a_k[m]$ are determined from $\mathbf{r}_k$ and a priori knowledge of $\{a_k[m']\}_{m' \neq m}$, using the function $f_M(.)$. Once the extrinsic probabilities of all the coded bits are known, they are delivered to the decoder. The decoder (represented by the function $f_D(.)$) updates the a priori information of the coded bits. As previously mentioned, details of $f_M$ and $f_D$ can be found in technical literature.

After a number of iterations ($I_{max}$), the a posteriori probabilities of the coded bits are determined as the product of the corresponding extrinsic and a priori probabilities.

## IV. Estimation through the EM algorithm

Assume we want to estimate a parameter vector $\mathbf{b}$ from an observation $\mathbf{r}$ in the presence of a so-called nuisance parameter vector $\mathbf{a}$, with distribution $p(\mathbf{a})$. The maximum likelihood (ML) estimate ($\hat{\mathbf{b}}_{ML}$) of $\mathbf{b}$ maximizes the log-likelihood function (LLF):

$$\hat{\mathbf{b}}_{ML} = \arg\max_{\mathbf{b}} \{\ln p(\mathbf{r}|\mathbf{b})\} \tag{3}$$

**Algorithm 1** Turbo detector

```
input: R, Ĥ
initialize: P^a(a_k[m]) = 1/2, ∀ unknown bits
for j = 0 to I_max − 1 do
    for all k and m do
        //demapping
        P^e(a_k[m]) = f_M(r_k, Ĥ, {P^a(a_k[m′])}_{m′≠m})
    end for
    for all k and m do
        //decoding
        P^a(a_k[m]) = f_D({P^e(a_k′[m′])})
    end for
end for
output: P^p(a_k[m]) ∝ P^a(a_k[m]) × P^e(a_k[m])
```

where

$$p(\mathbf{r}|\mathbf{b}) = \int_{\mathbf{a}} p(\mathbf{r}|\mathbf{a},\mathbf{b})\, p(\mathbf{a})\, d\mathbf{a}. \tag{4}$$

denotes the likelihood function. Often $p(\mathbf{r}|\mathbf{b})$ is very difficult to calculate. The EM algorithm is a method that iteratively solves (3), without the explicit calculation of (4). Defining the complete data $\mathbf{x}$, related to the observation $\mathbf{r}$ through a many-to-one mapping, the EM algorithm breaks up in two parts: the Expectation part (Eq. 5) and the Maximization part (Eq. 6):

$$Q\left(\mathbf{b}, \hat{\mathbf{b}}^{(i)}\right) = \int_{\mathbf{x}} p\left(\mathbf{x}\,\middle|\,\mathbf{r}, \hat{\mathbf{b}}^{(i)}\right) \ln p(\mathbf{x}|\mathbf{b}) d\mathbf{x} \tag{5}$$

$$\hat{\mathbf{b}}^{(i+1)} = \arg\max_{\mathbf{b}} \left\{ Q\left(\mathbf{b}, \hat{\mathbf{b}}^{(i)}\right)\right\}. \tag{6}$$

It has been shown that $\hat{\mathbf{b}}^{(i)}$ converges to a stationary point of the LLF under fairly general conditions [19]. The EM algorithm can easily be extended to acquire the Maximum a Posteriori (MAP) estimate of $\mathbf{b}$, by taking the a priori distribution $p(\mathbf{b})$ into account in (5).

## V. Turbo Estimation

*Initialization:* In order to converge to the ML estimate, the EM algorithm requires proper initialization. We resort to a data-aided (DA) ML estimator to provide the initial estimate $\hat{\mathbf{H}}^{(0)}$ of $\mathbf{H}$ [20]:

$$\hat{\mathbf{H}}^{(0)} = \mathbf{R}\mathbf{A}_P^H \left(\mathbf{A}_P\mathbf{A}_P^H\right)^{-1} \tag{7}$$

where $\mathbf{A}_P$ denotes the $N_R \times (N_S + N_P)$ matrix obtained by replacing all (unknown) coded symbols in $\mathbf{A}$ by 0. Note that $\mathbf{A}_P^H \left(\mathbf{A}_P\mathbf{A}_P^H\right)^{-1}$ can be precomputed and stored at the receiver-side. Performance of this estimator depends on the structure of the training sequence. It was shown in [20] that orthogonal training sequences (i.e. where $\mathbf{A}_P\mathbf{A}_P^H$ is a diagonal matrix) yield estimates with the lowest MSE. Hence, we initialize the EM algorithm using (7) with a limited number of orthogonal pilot symbols.

The EM algorithm to be described involves including Algorithm 1 into an additional for-loop: at each EM iteration $i$ (for $i = 0, \ldots, I_{EM} - 1$) we first perform Algorithm 1 and then execute the following two steps.

*Expectation Step:* In the context of the MIMO transmission system, $\mathbf{a}$ and $\mathbf{b}$ from section IV represent the transmitted symbols and the channel gains, respectively: $\mathbf{r} \leftrightarrow \mathbf{R}$, $\mathbf{a} \leftrightarrow \mathbf{A}$ and $\mathbf{b} \leftrightarrow \mathbf{H}$ [1]. We select as complete data $\mathbf{x} = [\mathbf{R}, \mathbf{A}]$. Considering the fact that $\mathbf{H}$ and $\mathbf{A}$ are independent, (5) is transformed into:

$$Q\left(\mathbf{H}, \hat{\mathbf{H}}^{(i)}\right) = E_{\mathbf{A}}\left[\ln p(\mathbf{R}|\,\mathbf{A}, \mathbf{H})|\,\mathbf{R}, \hat{\mathbf{H}}^{(i)}\right] \tag{8}$$

Taking into account the properties of the noise, the LLF $\ln p(\mathbf{R}|\,\mathbf{A}, \mathbf{H})$ can be written as follows (up to irrelevant constants):

$$\begin{aligned} &\ln p(\mathbf{R}|\,\mathbf{A}, \mathbf{H}) = \\ &-tr\left(\mathbf{H}\mathbf{A}\mathbf{A}^H\mathbf{H}^H\right) + 2\Re\left(tr\left(\mathbf{R}\mathbf{A}^H\mathbf{H}^H\right)\right) \end{aligned} \tag{9}$$

where $tr(\mathbf{X})$ denotes the trace of the square matrix $\mathbf{X}$. Substituting (9) into (8) yields:

$$\begin{aligned} &Q\left(\mathbf{H}, \hat{\mathbf{H}}^{(i)}\right) = \\ &-tr\left(\mathbf{H}\overline{\mathbf{A}\mathbf{A}^H}\mathbf{H}^H\right) + 2\Re\left(tr\left(\mathbf{R}\overline{\mathbf{A}}^H\mathbf{H}^H\right)\right) \end{aligned}$$

with

$$\begin{aligned} \overline{\mathbf{A}\mathbf{A}^H} &= E_{\mathbf{A}}\left[\mathbf{A}\mathbf{A}^H\,\middle|\,\mathbf{R}, \hat{\mathbf{H}}^{(i)}\right] \\ \overline{\mathbf{A}} &= E_{\mathbf{A}}\left[\mathbf{A}|\,\mathbf{R}, \hat{\mathbf{H}}^{(i)}\right]. \end{aligned}$$

Since $\overline{\mathbf{A}\mathbf{A}^H} = E_{\mathbf{A}}\left[\sum_k \mathbf{a}_k\mathbf{a}_k^H\,\middle|\,\mathbf{R}, \hat{\mathbf{H}}^{(i)}\right]$, both $\overline{\mathbf{A}\mathbf{A}^H}$ and $\overline{\mathbf{A}}$ can be determined based on the a posteriori probabilities of the transmit vectors $\mathbf{a}_k$ (i.e., $P^p(\mathbf{a}_k) = P\left(\mathbf{a}_k\,\middle|\,\mathbf{R}, \hat{\mathbf{H}}^{(i)}\right)$):

$$P^p(\mathbf{a}_k) = \gamma P\left(\mathbf{r}_k\,\middle|\,\mathbf{a}_k, \hat{\mathbf{H}}^{(i)}\right) \prod_m P^a(a_k[m]) \tag{10}$$

[1] Actually, a vectorized representation of $\mathbf{R}$, $\mathbf{A}$ and $\mathbf{H}$

for some normalizing constant $\gamma$. The a priori probabilities $P^a(a_k[m])$ in (10) are provided by the iterative detector. In passing, we mention that in several papers $P^p(\mathbf{a}_k)$ is approximated as $P^p(\mathbf{a}_k) \approx \prod_{m=1}^{N_T \log_2 M} P^p(a_k[m])$ or as $P^p(\mathbf{a}_k) \approx \prod_{l=1}^{N_T} P^p(a_{k,l})$ with $a_{k,l} \in \Omega$ denoting the $l$-th symbol and $a_k[m] \in \{0,1\}$ the $m$-th bit of symbolvector $\mathbf{a}_k$. This approach inevitably gives rise to some degradation as compared to the correct expression (10).

*Maximization step:* An exact solution of (6) is given by:

$$\hat{\mathbf{H}}^{(i+1)} = \mathbf{R}\overline{\mathbf{A}^H}\left(\overline{\mathbf{A}\mathbf{A}^H}\right)^{-1}. \tag{11}$$

Notice the resemblance with the DA estimator (7). In fact, the training symbols from (7) are now replaced with the a posteriori first and second order moments of the data symbols. Observe that $\left(\overline{\mathbf{A}\mathbf{A}^H}\right)$ is an $N_T \times N_T$ matrix, so $\left(\overline{\mathbf{A}\mathbf{A}^H}\right)^{-1}$ is generally easy to compute.

## VI. Low-complexity approximation: turbo estimation

The main drawback of the proposed channel estimator is its high computational complexity: indeed, each time we update $\hat{\mathbf{H}}$, Algorithm 1 has to be executed. Hence, if $\hat{\mathbf{H}}$ is updated $I_{EM}$ times, the complexity is $I_{EM}$ times that of Algorithm 1, which is generally unacceptable. We will propose two ways of reducing this complexity.

*Uncoded estimation:* One approach to avoid unnecessary decoding stages, is to ignore the presence of the code in the iterative estimation algorithm. This way, the *exact* a posteriori probabilities of the symbol vectors are obtained from the received signal by formally replacing in expression (10), $P^a(a_k[m])$ with $1/2$ when $a_k[m]$ is an (unknown) coded bit and $P^a(a_k[m]) = 1$ or $0$ when $a_k[m]$ is part of a pilot symbol (-vector). The iterative estimator will again converge to a stationary point of the *'uncoded'* likelihood function. This 'uncoded' likelihood function is again given by (4), but the expectation in (4) involves averaging with respect to all possible (uncoded) symbols, instead of all possible codewords. The estimate of $\mathbf{H}$ is then delivered to the iterative detector.

*Embedded estimation:* In order to exploit all available information with no significant additional computational cost, we propose to *embed* the estimation steps in the turbo detector. The idea is to move the initialization steps in Algorithm 1 outside the for-loop and simultaneously reduce $I_{max}$. The additional computational complexity compared to Algorithm 1 is now only marginal. For instance, if we set $I_{max} = 1$, the only additional complexity, in the best case (i.e., when the estimate of $\mathbf{H}$ converges after one or two EM iterations), lies in computing $P^p(\mathbf{a}_k)$ in (10) and the matrix computation in (11). Although this approach is not optimal in the sense of determining the ML estimate of $\mathbf{H}$, we will show through computer simulations that the performance of this practical algorithm is still excellent. The resulting algorithm is shown in Algorithm 2.

**Algorithm 2** Embedded estimator

```
input: R, Ĥ^(0)
initialize: P^a(a_k[m]) = 1/2, ∀ unknown bits
for i = 0 to I_EM − 1 do
    perform Algorithm 1 using Ĥ^(i) without initialization steps
    determine Ĥ^(i+1) using (11)
end for
output: P^p(a_k[m]) ∝ P^e(a_k[m]) × P^e(a_k[m])
```

## VII. Performance results

We have carried out computer simulations using a recursive systematic rate 1/2 convolutionally encoder with octal generators (27,33) and a random interleaver. We consider two channel types: a SISO AWGN channel using 16QAM signaling, and a flat-fading MIMO channel with BPSK signaling. The considered performance measures are the bit-error-rate (BER) and the mean square error (MSE) of the estimates. The BER will be compared to the BER under perfect synchronization, while the MSE will be compared to the modified Cramer Rao Bound (MCRB), which is a lower bound on the MSE of any unbiased estimator [21].

### A. SISO AWGN

Let us consider a SISO AWGN channel with both Gray and optimized non-Gray mapped 16QAM symbols. Frames consist of 256 symbols and initial estimates for the iterative estimation algorithms are obtained by means of only 8 pilot symbols. The BER is plotted in Fig. 2 for several estimation schemes. When the channel is perfectly known, the non-Gray mapped scheme outperforms the Gray-mapped scheme for $E_b/N_0 > 5\,\text{dB}$. However, the benefit from choosing an optimized mapping is lost when estimating the parameters based on pilot symbols only. Non-Gray BICM-ID schemes appear to be more sensitive to channel errors, than their Gray-mapped counterparts. This emphasizes the impor-

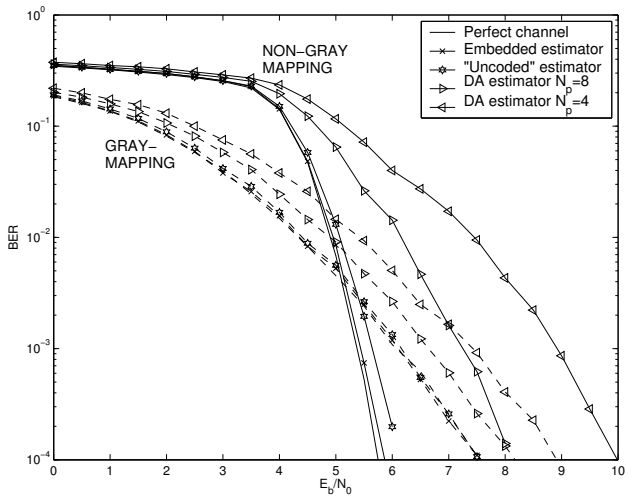


Fig. 2. BER of joint estimation and detection over AWGN channel for Gray (dashed lines) and non-Gray mapping (full lines).

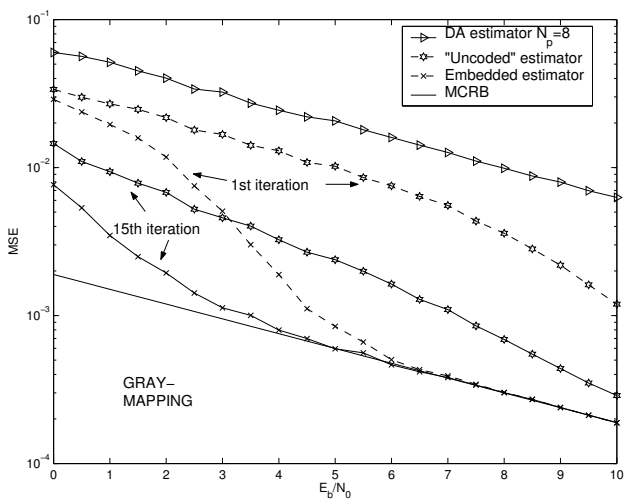


Fig. 3. MSE of joint estimation and detection over AWGN channel for Gray mapping.

tance of accurate channel estimation algorithms for configurations using advanced mappings. Using the embedded turbo-estimator from section VI yields an error rate within 0.1 dB of the perfectly synchronized performance. Compared to the data-aided estimator using 8 pilot symbols, a gain of more than 2 dB is achieved for Non-Gray mappings. The 'uncoded' estimator yields a degradation of 0.5 dB compared to perfect channel knowledge for non-Gray mappings, whereas for Gray mappings hardly any degradation is observed.

Let us now examine the MSE performance, as shown in Fig. 3 and Fig. 4 for Gray and non-Gray mappings, respectively. We observe that the 'uncoded' estimator upper bounds the embedded es-

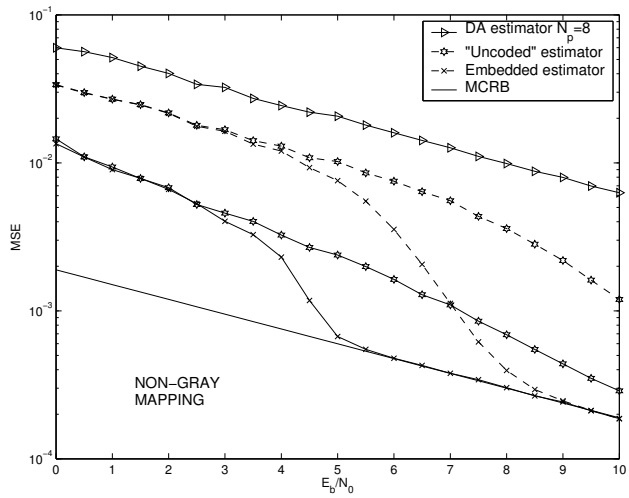


Fig. 4. MSE of joint estimation and detection over AWGN channel for non-Gray mapping.

timator. This is not surprising since the embedded estimator has more information (i.e., from the decoder) at its disposal compared to the 'uncoded' estimator. At low SNR, the benefit from using the code is negligible and the embedded and 'uncoded' estimator converge. At high SNR, the MSE of the embedded estimator approaches the MCRB after convergence, whereas the MSE of the 'uncoded' estimator does not attain the MCRB for the considered SNR range. Note also that the MCRB is reached at lower SNR for Gray mappings compared to non-Gray mappings. This can be explained by Fig. 2, where we notice that the code becomes effective only above a certain SNR-threshold for non-Gray mappings ('waterfall region'). Hence, the estimator only benefits from decoder output at high SNR.

### *B. MIMO*

Similar results are obtained for channel estimation in multiple antenna configurations. Fig. 5 and 6 illustrate the BER and MSE performance of a BICM MIMO system with $N_T = 4$ transmit and $N_R = 2$ receive antennas using BPSK signaling. Frames consist of 1024 bits (i.e., $N_S = 256$) and initial estimates are obtained by means of $N_P = 8$ pilot symbol vectors. We observe in Fig. 6 that the MCRB is reached, above a certain SNR threshold ($E_b/N_0 > 5\,\text{dB}$), using the embedded estimator. This means that for $E_b/N_0 > 5\,\text{dB}$ the embedded estimator is as accurate as an estimator that knows all the coded symbols exactly. The suboptimal 'uncoded' estimator converges to the MCRB for higher SNR. From the BER curves in Fig. 5, we observe that the 'uncoded' estimator results in approximately 1

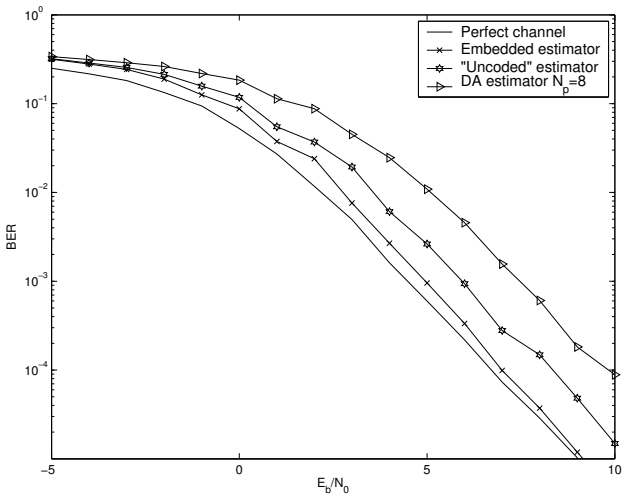


Fig. 5. BER of joint estimation and detection over MIMO channel ($N_T = 4$, $N_R = 2$).

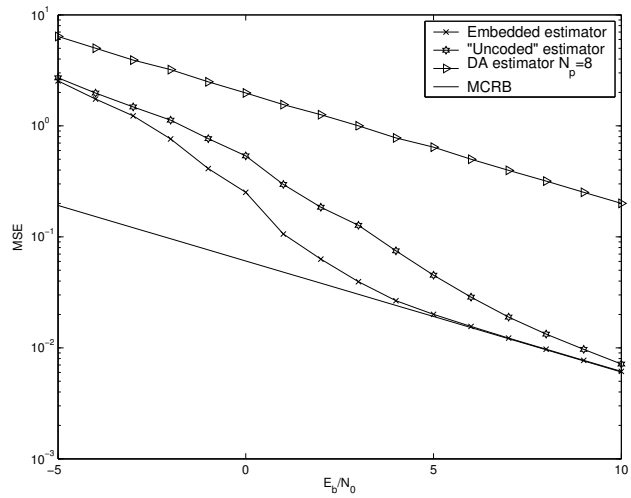


Fig. 6. MSE of joint estimation and detection over MIMO channel ($N_T = 4$, $N_R = 2$, 15 iterations).

dB degradation, whereas the embedded estimator degrades the BER by only 0.1 dB compared to perfect channel knowledge. Again, a significant gain (2.5 dB) is achieved compared to data-aided estimation using 8 pilot symbols per transmit antenna.

## VIII. Conclusions

In this paper, we addressed code-aided channel estimation in a (multi-antenna) BICM set-up. Employing the Expectation Maximization (EM) algorithm, an optimal (in the ML sense) estimator is derived. This estimator iterates between detection and estimation by accepting at each iteration soft information from the detector.

In order to reduce the computational overhead related to estimation, two low-complexity estimators are proposed, both based on the aforementioned optimal EM estimator. The first one treats the data as uncoded during estimation, while the second embeds estimation stages in the detection stages. Both BER and MSE analysis demonstrate the near-optimal performance of the embedded estimator.

## Acknowledgments

This work has been supported by the Interuniversity Attraction Poles Program P5/11- Belgian Science Policy. The first author also gratefully acknowledges the support from the National Fund for Scientific Research (FWO-Vlaanderen).

## References

[1] E. Zehavi, "8-PSK Trellis Codes for a Rayleigh Channel," *IEEE Trans. Comm.*, vol. 40, no. 5, pp. 873–884, May 1992.

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

[3] J.J. Boutros, F. Boixadera and C. Lamy, "Bit-interleaved coded modulations for multiple-input multiple-output channels," in *Proceedings of the IEEE 6th Int. Symp. on Spread-Spectrum Tech. & Appli.*, New Jersey, USA, Sept. 2000, pp. 123–126.

[4] G.J. Foshini and M.J. Gans, "On limits of wireless communication in a fading environment when using multiple amtennas," *Wireless Personal Communications*, vol. 6, no. 3, pp. 311–335, Mar. 1998.

[5] X. Li, A. Chindapol and A.R. Ritcey, "Bit-Interleaved Coded Modulation With Iterative Decoding and 8PSK Signaling," *IEEE Trans. Comm.*, vol. 50, no. 8, pp. 1250–1257, Aug. 2002.

[6] S.Y. Le Goff, "Signal Constellations for Bit-Interleaved Coded Modulation," *IEEE Trans. Inform. Theory*, vol. 49, no. 1, pp. 307–313, Jan. 2003.

[7] J. Boutros, N. Gresset and L. Brunel, "Turbo Coding and Decoding for Multiple Antenna Channels," in *International Symposium on Turbo Codes and Related Topics*, Brest, Sept. 2003.

[8] F. Simoens, H. Wymeersch and M. Moeneclaey, "Spatial Mapping for MIMO systems," in *IEEE Information Theory Workshop proceedings (to appear)*, San Antonio, Oct. 2004.

[9] T. Muharemovic, A. Gatherer, W. Ebel, S. Hosur, D. Hocevar and E. Huang, "Space-time codes with bit interleaving," in *Proc. IEEE GLOBECOM*, 2001.

[10] Y. Huang and J.A. Ritcey, "16-QAM BICM-ID in Fading Channels With Imperfect Channel State Information," *IEEE Trans. Wireless. Comm.*, vol. 2, no. 5, pp. 1000–1007, Sept. 2003.

[11] Z. Baranski, A.M. Haimovich and J. Garcia-Frias, "EM-based iterative receiver for space-time coded modulation with noise variance estimation," in *IEEE Globecom*, nov 2002, pp. 355–359.

[12] C. Cozzo and B.L. Hughes, "Joint Channel Estimation and Data Detection in Space-Time Communications," *IEEE Trans. Comm.*, vol. 51, no. 8, pp. 1266–1270, aug 2003.

[13] Y. Li, C.N. Georghiades and G. Huang, "Iterative Maximum-Likelihood Sequence Estimation for Space-Time

Coded Systems," *IEEE Trans. Comm.*, vol. 49, no. 6, pp. 948–951, June 2001.

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

[15] A. Grant, "Joint Decoding and Channel Estimation for Linear MIMO Channels," in *In proceedings of Wireless Comm. and Networking Conf.*, Chicago, USA, Sep 2000, pp. 1009–1012.

[16] R. Visoz and A.O. Berthet, "Iterative Decoding and Channel Estimation for Space-Time BICM over MIMO Block Fading Multipath AWGN Channel," *IEEE Trans. Comm.*, vol. 51, no. 8, pp. 1358–1367, Aug 2003.

[17] X. Li and J.A. Ritcey, "Trellis-coded modulation with bit interleaving and iterative decoding," *IEEE Journal on Selected Areas in Comm.*, vol. 17, no. 4, Apr. 1999.

[18] A. Stefanov and T.M. Duman, "Turbo-Coded Modulation for Systems with Transmit and Receive Antenna Diversity over Block Fading Channels: System Model, Decoding Approaches, and Practical Considerations," *IEEE Journal on Select. areas Comm.*, vol. 19, no. 5, pp. 958–968, may 2001.

[19] A.P. Dempster, N.M. Laird and D.B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *Journal of the Royal Statistical Society*, vol. 39, no. 1, pp. pp. 1–38, 1977, Series B.

[20] T. L. Marzetta, "BLAST Training: Estimating Channel Characteristics for High Capacity Space-Time Wireless," in *Ann. Allerton Conf.*, Monticello, Sept 1999.

[21] M. Moeneclaey, "On the true and the modified Cramer-Rao bounds for the estimation of a scalar parameter in the presence of nuisance parameters," *IEEE Trans. Comm.*, vol. 46, no. 11, pp. 1536–1544, Nov. 1998.