WIRELESS COMMUNICATIONS AND MOBILE COMPUTING
Wirel. Commun. Mob. Comput. 2007; **7**:81–89
Published online 9 January 2006 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/wcm.319

EM-based iterative receiver for coded MIMO systems in unknown spatially correlated noise

Wei Mo*,†, Zhengdao Wang and Aleksandar Dogandžić
Department of Electrical and Computer Engineering, Iowa State University, Ames, IA 50010, U.S.A.

Summary

We present iterative channel estimation and decoding schemes for multi-input multi-output (MIMO) Rayleigh block fading channels in spatially correlated noise. An expectation-maximization (EM) algorithm is utilized to find the maximum likelihood (ML) estimates of the channel and spatial noise covariance matrices, and to compute soft information of coded symbols which is sent to an error-control decoder. The extrinsic information produced by the decoder is then used to refine channel estimation. Several iterations are performed between the above channel estimation and decoding steps. We derive modified Cramer–Rao Bound (MCRB) for the unknown channel and noise parameters, and show that the proposed EM-based channel estimation scheme achieves the MCRB at medium and high SNRs. For a bit error rate of 10^{-6} and long frame length, there is negligible performance difference between the proposed scheme and the ideal coherent detector that utilizes the true channel and noise covariance matrices. Copyright © 2006 John Wiley & Sons, Ltd.

KEY WORDS: MIMO; channel estimation; expectation-maximization; turbo code; CRB; BER

1. Introduction

Multi-input multi-output (MIMO) fading channel estimation is a major challenge for multiple antenna systems because the detection of information symbols depends critically on the availability of full or partial channel state information. Recently, there has been an increasing interest in iterative channel estimation and data decoding [2,10,16], where data decision obtained from the decoding, either hard or soft, is used as an additional information to refine the channel estimation. In References [2,10], maximum likelihood (ML) and maximum *a posteriori* (MAP) methods are used to estimate the channel via expectation-maximization (EM) algorithms [5]. EM algorithms have also been applied for symbol detection, see References [4,12]. Least-squares (LS) estimation together with hard and soft decision feedback is studied in Reference [16]. All of these methods assume that the additive noise is both temporally and spatially white. Channel estimation for MIMO systems in spatially correlated noise has been studied in References [7,11], where deterministic ML and simple non-iterative data decoding methods were proposed.

In this paper (see also Reference [13]), we propose an iterative channel estimation (via EM algorithm) and decoding scheme for spatially correlated noise with unknown covariance matrix. Instead of MAP estimation in Reference [2], which requires knowledge of second-order statistical properties of the

*Correspondence to: Wei Mo, Department of Electrical and Computer Engineering, Iowa State University, Ames, IA 50010, U.S.A.
†E-mail: mowei@iastate.edu

channel at the receiver, we estimate both the channel and the spatial noise covariance without prior knowledge of the channel statistical properties. This work generalizes our results for single-input multi-output (SIMO) systems in Reference [6] to the coded MIMO scenario. We also develop an iterative receiver which alternates between deterministic ML channel estimation with soft decision feedback and error-control decoding.

The system model is introduced in Section 2. In Section 3, we derive the EM algorithm for estimating the unknown channel and noise parameters. Section 4 discussed the design of the iterative space-time receiver. We discuss the initial values and Cramér–Rao bounds of the channel estimation in Section 5. Simulation results are presented in Section 6 and Section 7 concludes the paper.

2. System Modeling

We consider a coded MIMO system having n_T transmit and n_R receive antennas in a frequency-flat block fading environment. We will use turbo code as an example of the error control code. Other codes, such as low-density parity-check (LDPC) codes, can also be used. The discrete-time transmitter model is shown in Figure 1.

Suppose that a block of L space-time codewords $\mathbf{X}$ of size $n_\mathrm{T} \times K$ each are transmitted. The lth received *space-time data matrix* $\mathbf{Y}_l$ can be modeled as

$$\mathbf{Y}_l = \mathbf{H} \cdot \mathbf{X}_l + \mathbf{E}_l, \quad l = 1, 2, \ldots, L \tag{1}$$

where $\mathbf{H}$ is an unknown $n_\mathrm{R} \times n_\mathrm{T}$ channel response matrix; $\mathbf{X}_l$ is the lth transmitted space-time codeword; $\mathbf{E}_l = [\boldsymbol{e}_l(1) \cdots \boldsymbol{e}_l(K)]$ is the lth noise matrix, where $\boldsymbol{e}_l(k)$ is temporally white and circularly symmetric zero-mean complex Gaussian noise vector with unknown spatial covariance matrix $\boldsymbol{\Sigma}$. It models co-channel interference (CCI) and receiver noise. This is a standard model for a communication channel, subject to (unstructured) interference and jamming, for example [6,7,11].

In this paper, we assume that M-ary phase shift keying (PSK) modulation and space-time orthogonal design (cf. [1,8,17]) are used. However, after minor

Fig. 1. Discrete-time transmitter model.

modifications, the proposed method can also be applied to other modulation schemes and general space-time codes. In Reference [6], we discuss similar modifications for the SIMO case; the extension to the MIMO case is straightforward, leading to algorithms with higher computational complexity compared with those presented herein. Without loss of generality, assume that each space-time codeword $\mathbf{X}_l$ is a linear function of K' transmitted symbols $\mathcal{S}_l = \{s_1^{(l)}, \ldots, s_{K'}^{(l)}\}$:

$$\mathbf{X}_l = \sum_{k=1}^{K'} \left(\mathrm{Re}\left\{ s_k^{(l)} \right\} \mathbf{A}_k + j \cdot \mathrm{Im}\left\{ s_k^{(l)} \right\} \mathbf{B}_k \right) \tag{2}$$

where $\mathrm{Re}\{\cdot\}$ and $\mathrm{Im}\{\cdot\}$ denote the real and imaginary parts, and $\mathbf{A}_k$ and $\mathbf{B}_k$ are fixed real-valued $n_\mathrm{T} \times K$ 'elementary' code matrices, satisfying the orthogonality conditions as follows [8,17]:

$$\begin{array}{lll} \mathbf{A}_k\mathbf{A}_k^\mathrm{T} = \mathbf{I}_{n_\mathrm{T}}, & \mathbf{B}_k\mathbf{B}_k^\mathrm{T} = \mathbf{I}_{n_\mathrm{T}}, & \mathbf{A}_k\mathbf{B}_t^\mathrm{T} = \mathbf{B}_t\mathbf{A}_k^\mathrm{T} \\ \mathbf{A}_k\mathbf{A}_t^\mathrm{T} = -\mathbf{A}_t\mathbf{A}_k^\mathrm{T}, & \mathbf{B}_k\mathbf{B}_t^\mathrm{T} = -\mathbf{B}_t\mathbf{B}_k^\mathrm{T}, & k \neq t \end{array} \tag{3}$$

so that

$$\mathbf{X}_l\mathbf{X}_l^\mathrm{H} = \sum_{k=1}^{K'} \left| s_k^{(l)} \right|^2 \cdot \mathbf{I}_{n_\mathrm{T}} = K'\mathbf{I}_{n_\mathrm{T}} \tag{4}$$

where $(\cdot)^\mathrm{T}$ and $(\cdot)^\mathrm{H}$ denote the transpose and the conjugate transpose, respectively and symbols $s_k^{(l)}$ from M-ary PSK constellation are assumed to have unit energy. The number of transmitted symbols K' represented by one space-time codeword is usually less than the codeword length K when $n_\mathrm{T} > 2$, and is equal to K when $n_\mathrm{T} = 2$.

To allow unique estimation of the channel $\mathbf{H}$ (i.e., to resolve the phase ambiguity associated with PSK modulation), we further assume that L_p known pilot space-time codewords $\mathbf{X}_{\mathrm{p},\ell}$, $\ell = 1, \ldots, L_\mathrm{p}$, are inserted at the beginning of the block, and denote the corresponding data matrices received by the array as $\mathbf{Y}_{\mathrm{p},\ell}$. Usually, L_p is a small number (say 2) and the pilot symbols alone do not provide good channel estimation. We adopt the block fading assumption implying that the channel $\mathbf{H}$ and noise covariance matrix $\boldsymbol{\Sigma}$ remain constant within each *block* of $(L + L_\mathrm{p})$ codewords, that is $K(L + L_\mathrm{p})$ time intervals, and change from one block to another independently.

Since turbo codes need long *frame length* to achieve good error performance, and the length of one block is limited by the coherent time of the fading channel, the turbo encoder implements coding *across* R blocks. Therefore, one turbo code frame is composed of $R(L+L_\mathrm{p})$ space-time codewords. At the receiver, the turbo decoder needs the estimates of the channels and the noise covariance matrices for all R blocks. A channel interleaver is used to spread the effect of imperfect channel estimates across the whole R blocks.

3. EM Algorithm for Channel Estimation

In this section, we derive an EM-based channel and noise covariance estimator from one block. The proposed method incorporates extrinsic information about the transmitted symbols from the turbo decoder through prior symbol probabilities. The estimates of the channel and noise covariance matrix will be used to update the extrinsic information about the transmitted symbols used by the turbo decoder.

Given a block of received data $[\mathbf{Y}_{\mathrm{p},1},\ldots,\mathbf{Y}_{\mathrm{p},L_\mathrm{p}},\mathbf{Y}_1,\ldots,\mathbf{Y}_L]$, the pilot space-time codewords $[\mathbf{X}_{\mathrm{p},1},\ldots,\mathbf{X}_{\mathrm{p},L_\mathrm{p}}]$, and the prior probabilities of the space-time codewords $\mathbf{X}_1,\ldots,\mathbf{X}_L$, we wish to find the ML estimates of the channel $\mathbf{H}$ and the noise covariance matrix $\boldsymbol{\Sigma}$ for this block.

The EM algorithm is a general iterative method for computing ML estimates in the scenarios where ML estimation cannot be easily performed by directly maximizing the likelihood function for the observed data [5]. Each EM iteration consists of maximizing the expected complete-data log-likelihood function, where the expectation is computed with respect to the conditional distribution of the unobserved data given the observed data. A good choice of unobserved data allows easy maximization of the expected complete-data log-likelihood.

For our channel estimation problem, the unknown space-time codewords $\{\mathbf{X}_l\}_{l=1}^{L}$ are modeled as the unobserved (or missing) data. By generalizing our results for SIMO systems in Reference [6], we obtain the following EM iteration:

Step I:

$$\mathbf{H}^{(i+1)}=\frac{1}{(L+L_\mathrm{p})K'}\times\left[\sum_{l=1}^{L}\mathbf{Y}_l\mathrm{E}_{\mathbf{X}_l|\mathbf{Y}_l}\left(\mathbf{X}_l^\mathrm{H};\mathbf{H}^{(i)},\boldsymbol{\Sigma}^{(i)}\right)+\sum_{\ell=1}^{L_\mathrm{p}}\mathbf{Y}_{\mathrm{p},\ell}\mathbf{X}_{\mathrm{p},\ell}^\mathrm{H}\right] \tag{5}$$

Step II:

$$\boldsymbol{\Sigma}^{(i+1)}=\mathbf{R}_{yy}-\frac{K'}{K}\cdot\mathbf{H}^{(i+1)}\left(\mathbf{H}^{(i+1)}\right)^\mathrm{H} \tag{6}$$

where

$$\mathbf{R}_{yy}=\frac{1}{(L+L_\mathrm{p})K}\left[\sum_{l=1}^{L}\mathbf{Y}_l\mathbf{Y}_l^\mathrm{H}+\sum_{\ell=1}^{L_\mathrm{p}}\mathbf{Y}_{\mathrm{p},\ell}\mathbf{Y}_{\mathrm{p},\ell}^\mathrm{H}\right] \tag{7}$$

Note that both Steps I and II contain both the expectation and maximization steps. To ensure positive definiteness (with probability one) of the estimates of $\boldsymbol{\Sigma}$, the following condition needs to be satisfied:

$$(L+L_\mathrm{p})K\geq(n_\mathrm{T}+n_\mathrm{R}) \tag{8}$$

see Reference [14, Theorems 10.1.1 and 3.1.4]. Since there is a one-to-one mapping between the set of information symbols $\mathcal{S}_l$ and the codeword $\mathbf{X}_l$, conditioning on $\mathcal{S}_l$ is equivalent to conditioning on $\mathbf{X}_l$. Following a derivation similar to Reference [11, Equation 7], the likelihood function of $\mathbf{X}_l$, $\mathbf{H}$, and $\boldsymbol{\Sigma}$ can then be written as

$$\begin{aligned}f(\mathbf{Y}_l\,|\,\mathbf{X}_l;\mathbf{H},\boldsymbol{\Sigma})&=f(\mathbf{Y}_l\,|\,\mathcal{S}_l;\mathbf{H},\boldsymbol{\Sigma})\\&=\mathrm{const}\cdot\prod_{k=1}^{K'}\exp\Big\{2\mathrm{Re}\Big((\mathrm{Re}(\mathrm{Tr}[\mathbf{Y}_l^\mathrm{H}\boldsymbol{\Sigma}^{-1}\mathbf{H}\mathbf{A}_k])\\&\quad+j\cdot\mathrm{Im}(\mathrm{Tr}[\mathbf{Y}_l^\mathrm{H}\boldsymbol{\Sigma}^{-1}\mathbf{H}\mathbf{B}_k]))s_k^{(l)}\Big)\Big\}\\&=\mathrm{const}\cdot\prod_{k=1}^{K'}f_k\left(\mathbf{Y}_l\,|\,s_k^{(l)};\mathbf{H},\boldsymbol{\Sigma}\right)\end{aligned} \tag{9}$$

where const denotes the terms that do not depend on $s_k^{(l)}$. The second equality in the above expression follows by applying the orthogonality conditions in Equation (3) which leads to the decoupling of the likelihood function for the space-time codeword into the product of the likelihood functions $f_k(\mathbf{Y}_l\,|\,s_k^{(l)};\mathbf{H},\boldsymbol{\Sigma})$ for the information symbols $s_k^{(l)}$, where the normalizing constants have been omitted. Assume that the information symbols $s_k^{(l)}$, $k=1,\ldots,K'$, $l=1,\ldots,L$, are independent and have the prior probability mass functions $p(s_k^{(l)})=p(s_k^{(l)}=s_m)$,

$m = 1, \ldots, M$, then Step I of the EM iteration can be simplified as

$$\mathbf{H}^{(i+1)} = \frac{1}{(L+L_\mathrm{p})K'} \left[\sum_{l=1}^{L} \sum_{k=1}^{K'} \mathbf{Y}_l \times \left(\mathrm{Re}\left(\mathrm{E}_{s_k^{(l)}|\mathbf{Y}_l}\left[s_k^{(l)}; \boldsymbol{H}^{(i)}, \boldsymbol{\Sigma}^{(i)} \right] \right) \boldsymbol{A}_k^\mathrm{H} - j \cdot \mathrm{Im}\left(\mathrm{E}_{s_k^{(l)}|\mathbf{Y}_l}\left[s_k^{(l)}; \boldsymbol{H}^{(i)}, \boldsymbol{\Sigma}^{(i)} \right] \right) \boldsymbol{B}_k^\mathrm{H} \right) + \sum_{\ell=1}^{L_\mathrm{p}} \boldsymbol{Y}_{\mathrm{p},\ell} \boldsymbol{X}_{\mathrm{p},\ell}^\mathrm{H} \right] \tag{10}$$

where

$$\mathrm{E}_{s_k^{(l)}|\mathbf{Y}_l}\left[s_k^{(l)}; \mathrm{H}^{(i)}, \boldsymbol{\Sigma}^{(i)} \right] = \frac{\sum_{m=1}^{M} s_m p\left(s_k^{(l)} = s_m \right) f_k\left(\mathbf{Y}_l | s_m; \mathbf{H}^{(i)}, \boldsymbol{\Sigma}^{(i)} \right)}{\sum_{n=1}^{M} p\left(s_k^{(l)} = s_n \right) f_k\left(\mathbf{Y}_l | s_n; \mathbf{H}^{(i)}, \boldsymbol{\Sigma}^{(i)} \right)} \tag{11}$$

and Step II remains the same. The prior probabilities $p(s_k^{(l)})$ comes from the error control decoder.

Most of the computations are in the calculation of Equation (10). Given $\mathrm{E}_{s_k^{(l)}|\mathbf{Y}_l}[s_k^{(l)}; \mathbf{H}^{(i)}, \boldsymbol{\Sigma}^{(i)}]$, we need $2n_\mathrm{R}K \cdot LK' + 2n_\mathrm{R}n_\mathrm{T}K \cdot K'$ multiplications to compute $\mathbf{H}^{(i+1)}$. According to Equation (9), $2n_\mathrm{R}K$ multiplications are needed for computing $\mathrm{E}_{s_k^{(l)}|\mathbf{Y}_l}[s_k^{(l)}; \mathbf{H}^{(i)}, \boldsymbol{\Sigma}^{(i)}]$. If $K' = K = n_\mathrm{R} = n_\mathrm{T} \triangleq n$, then the total computational complexity can be expressed as $\mathcal{O}(4n^2)$ per symbol per iteration.

4. Iterative Space-Time Receiver

The proposed iterative receiver model is shown in Figure 2. It consists of two modules: a bank of R channel estimators and demodulators (developed in the previous section) and a turbo decoder. The soft information about the information symbols is exchanged between them. For simplicity, we have not shown the interleaver and deinterleaver in the diagram. In the following, we also assume that the interleaving and deinterleaving operations are performed as needed.

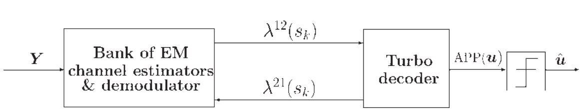


Fig. 2. The receiver with iterative channel estimation and decoding.

The received data $\mathbf{Y}$ are first divided into R blocks $\{[\mathbf{Y}_{\mathrm{p},1}^r, \ldots, \mathbf{Y}_{\mathrm{p},L_\mathrm{p}}^r, \mathbf{Y}_1^r, \ldots, \mathbf{Y}_L^r]\}_{r=1}^R$ each of length $(L + L_\mathrm{p})K$, and then fed into R channel estimators. Based on the pilot codewords and the prior probabilities of the information symbols, each channel estimator estimates the channel $\hat{\mathbf{H}}_r$ and noise covariance matrix $\hat{\boldsymbol{\Sigma}}_r$, then computes the posterior log-probabilities of the information symbols as follows

$$\begin{aligned} \Lambda_r^1\left[s_k^{(l)} \right] &= \mathrm{const} + \log p\left(s_k^{(l)} \right) + \log f_1\left(\mathbf{Y}_l^r | s_k^{(l)}; \hat{\mathbf{H}}_r, \hat{\boldsymbol{\Sigma}}_r \right) \\ &= \mathrm{const} + \lambda_r^{21}\left[s_k^{(l)} \right] + \lambda_r^{12}\left[s_k^{(l)} \right] \\ r = 1, \ldots, R, &\quad l = 1, \ldots, L, \quad k = 1, \ldots, K' \end{aligned} \tag{12}$$

where const denotes the terms independent of $s_k^{(l)}$. The second term $\lambda_r^{21}[s_k^{(l)}]$ represents the prior log-probability of the information symbol $s_k^{(l)}$, which is computed by the turbo decoder in the previous iteration, and then fed back to the channel estimator. For the first iteration, we assume equally likely symbols, that is no prior information available. The third term $\lambda_r^{12}[s_k^{(l)}]$ in Equation (12) represents the *extrinsic* information produced by the channel estimator and demodulator, based on the received data $\mathbf{Y}_r$, pilot codewords, and the prior information of all other symbols in the block. All the extrinsic information metrics $\{\lambda_r^{12}[s_k^{(l)}]\}_{r=1,k=1,l=1}^{R,K',L}$ are reassembled together, and sent into the turbo decoder, as the prior information for the decoding.

Using the extrinsic information of the information symbols coming from channel estimators and the structure of the turbo codes, the turbo decoder computes the posterior log-probability of each symbol as:

$$\begin{aligned} \Lambda_r^2\left[s_k^{(l)} \right] &= \mathrm{const} + \lambda_r^{12}\left[s_k^{(l)} \right] + \log p\left(s_k^{(l)} \middle| \left\{ \lambda_{r'}^{12}\left[s_{k'}^{(l')} \right] \right\}_{r'=1,k'=1,l'=1,(r',k',l')\neq(r,k,l)}^{R,K',L}; \text{code constraints} \right) \\ &= \mathrm{const} + \lambda_r^{12}\left[s_k^{(l)} \right] + \lambda_r^{21}\left[s_k^{(l)} \right] \end{aligned} \tag{13}$$

It is seen from Equation (13) that the output of the turbo decoder consists of the prior information $\lambda_r^{12}[s_k^{(l)}]$, provided by the channel estimators, and the *extrinsic* information $\lambda_r^{21}[s_k^{(l)}]$ delivered to the channel estimators in the next iteration. This extrinsic information is the information of the symbol $s_k^{(l)}$ in the rth block obtained from the prior information of the other symbols in the frame and the code constraints. The turbo decoder also outputs the *a posteriori* probability $\mathrm{APP}(u_i)$ of every information bit u_i, which is used to do the decision in the last iteration.

5. Discussion

5.1. Initialization of the EM Algorithm

Although the EM algorithm increases (or at least does not decrease) the likelihood function at each iteration, it may get trapped at the local maximum when the initial values are too far from the true parameters. So we need a more robust method to give good initial estimates of the channel and noise covariance matrices, which are used to initialize our EM algorithm.

For this purpose, we choose the iterative weighted least-squares with projections (ILSP) method for space-time coding systems proposed in Reference [15]. For completeness, we summarize below our implementation of this method:

Step 1: Fix $\mathbf{H} = \hat{\mathbf{H}}$ and compute

$$\hat{\mathbf{X}}_l = \mathrm{proj}\left[\mathbf{H}^{\mathrm{H}}\mathbf{R}_{yy}^{-1}\mathbf{Y}_l\right], \quad l = 1, \ldots, L \tag{14}$$

Step 2: Fix $\mathbf{X}_1 = \hat{\mathbf{X}}_1, \ldots, \mathbf{X}_L = \hat{\mathbf{X}}_L$ and compute

$$\hat{\mathbf{H}} = \frac{1}{(L+L_{\mathrm{p}})K'} \cdot \left[\sum_{l=1}^{L}\mathbf{Y}_l\mathbf{X}_l^{\mathrm{H}} + \sum_{\ell=1}^{L_{\mathrm{p}}}\mathbf{Y}_{\mathrm{p},\ell}\mathbf{X}_{\mathrm{p},\ell}^{\mathrm{H}}\right] \tag{15}$$

Go to step 1 and repeat.

where $\mathrm{proj}[\cdot]$ denotes projection onto the nearest (in the Frobenius norm) space-time codeword. This method is initialized with the least-square estimate using the pilot codewords

$$\hat{\mathbf{H}}_{\mathrm{LS}} = \frac{1}{L_{\mathrm{p}}K'}\sum_{\ell=1}^{L_{\mathrm{p}}}\mathbf{Y}_{\mathrm{p},\ell}\mathbf{X}_{\mathrm{p},\ell}^{\mathrm{H}} \tag{16}$$

After several iterations, we obtain a rough estimate of the channel, then the estimate of the noise covariance matrix is computed using Equation (6), both of which will be used to initialize the EM algorithm.

5.2. Modified Cramér–Rao Bound

The exact Cramér–Rao bound (CRB) for the unknown parameters under the data model in Section 2 is difficult to compute. Here, we derive the modified CRB (MCRB) [9], which is a lower bound on the exact CRB. First, we rewrite Equation (1) by stacking all K time samples from the lth received space-time data matrix into a single vector:

$$\boldsymbol{y}_l = \mathbf{Z}_l\boldsymbol{h}_l + \boldsymbol{e}_l \tag{17}$$

$$\mathbf{Z}_l = \mathbf{X}_l^{\mathrm{T}} \otimes \mathbf{I}_{n_{\mathrm{R}}} \tag{18}$$

where $\boldsymbol{y}_l = \mathrm{vec}(\mathbf{Y}_l)$, $\boldsymbol{h}_l = \mathrm{vec}\{\mathbf{H}_l\}$, $\boldsymbol{e}_k = \mathrm{vec}(\mathbf{E}_l)$, $\otimes$ denotes the Kronecker product, and the vec operator stacks the columns of a matrix one below another into a single column vector. Then, Equation (17) holds for the pilot data as well, with $\mathbf{Y}_l$ and $\mathbf{X}_l$ replaced by $\mathbf{Y}_{\mathrm{p},\ell}$ and $\mathbf{X}_{\mathrm{p},\ell}$, respectively. Define also $\mathbf{Z}_{\mathrm{p},\ell} = \mathbf{X}_{\mathrm{p},\ell}^{\mathrm{T}} \otimes \mathbf{I}_{n_{\mathrm{R}}}$ and the vector of the unknown channel and noise parameters $\boldsymbol{\rho} = [\boldsymbol{\eta}^{\mathrm{T}}, \boldsymbol{\psi}^{\mathrm{T}}]^{\mathrm{T}}$, where $\boldsymbol{\eta} = [\mathrm{Re}(\boldsymbol{h})^{\mathrm{T}}, \mathrm{Im}(\boldsymbol{h})^{\mathrm{T}}]^{\mathrm{T}}$ and $\boldsymbol{\psi} = [\mathrm{Re}\{\mathrm{vech}(\boldsymbol{\Sigma})\}^{\mathrm{T}}, \mathrm{Im}\{\underline{\mathrm{vech}}(\boldsymbol{\Sigma})\}^{\mathrm{T}}]^{\mathrm{T}}$. (The vech and $\underline{\mathrm{vech}}$ operators create a single column vector by stacking elements below the main diagonal columnwise; vech includes the main diagonal, whereas $\underline{\mathrm{vech}}$ omits it.) The MCRB for the unknown parameters $\boldsymbol{\rho}$ is identical to the exact CRB for these parameters when the space-time codewords $\mathbf{X}_l$ are *known*, and is equal to:

$$\mathrm{MCRB}_{\boldsymbol{\rho}} = \begin{bmatrix} \mathrm{MCRB}_{\boldsymbol{\eta}} & 0 \\ 0 & \mathrm{MCRB}_{\boldsymbol{\psi}} \end{bmatrix} \tag{19}$$

where

$$\mathrm{MCRB}_{\boldsymbol{\eta}} = \frac{1}{2K'(L+L_{\mathrm{p}})} \times \begin{bmatrix} \mathrm{Re}\{\mathbf{I}_{n_{\mathrm{T}}} \otimes \boldsymbol{\Sigma}\} & -\mathrm{Im}\{\mathbf{I}_{n_{\mathrm{T}}} \otimes \boldsymbol{\Sigma}\} \\ \mathrm{Im}\{\mathbf{I}_{n_{\mathrm{T}}} \otimes \boldsymbol{\Sigma}\} & \mathrm{Re}\{\mathbf{I}_{n_{\mathrm{T}}} \otimes \boldsymbol{\Sigma}\} \end{bmatrix} \tag{20}$$

$$\mathrm{MCRB}_{\boldsymbol{\psi}} = \frac{1}{K'(L+L_{\mathrm{p}})} \cdot \mathcal{I}_{\boldsymbol{\psi}}^{-1} \tag{21}$$

and the (i,k)th element of $\mathcal{I}_{\boldsymbol{\psi}}$ is

$$[\mathcal{I}_{\boldsymbol{\psi}}]_{i,k} = \mathrm{tr}\left\{\boldsymbol{\Sigma}^{-1}\frac{\partial\boldsymbol{\Sigma}}{\partial\psi_i}\boldsymbol{\Sigma}^{-1}\frac{\partial\boldsymbol{\Sigma}}{\partial\psi_k}\right\} \triangleq \mathcal{I}(\psi_i, \psi_k) \tag{22}$$

for $i, k = 1, 2, \ldots, n_{\mathrm{R}}^2$. Denote by $\boldsymbol{\Sigma}_{p,q}$ the (p,q) element of $\boldsymbol{\Sigma}$, for $p, q = 1, 2, \ldots, n_{\mathrm{R}}$. Using this notation, we further simplify Equation (22): for $p_1 > q_1$ and $p_2 > q_2$, we have

$$\mathcal{I}(\mathrm{Re}\{\boldsymbol{\Sigma}_{p_1,q_1}\}, \mathrm{Re}\{\boldsymbol{\Sigma}_{p_2,q_2}\}) = \mathcal{I}(\mathrm{Re}\{\boldsymbol{\Sigma}_{p_2,q_2}\}, \mathrm{Re}\{\boldsymbol{\Sigma}_{p_1,q_1}\}) = 2 \cdot \mathrm{Re}\{[\boldsymbol{\Sigma}^{-1}]_{q_2,p_1} \cdot [\boldsymbol{\Sigma}^{-1}]_{q_1,p_2} + [\boldsymbol{\Sigma}^{-1}]_{q_2,q_1} \cdot [\boldsymbol{\Sigma}^{-1}]_{p_1,p_2}\} \quad (23)$$

$$\mathcal{I}(\mathrm{Re}\{\boldsymbol{\Sigma}_{p_1,q_1}\}, \mathrm{Im}\{\boldsymbol{\Sigma}_{p_2,q_2}\}) = \mathcal{I}(\mathrm{Im}\{\boldsymbol{\Sigma}_{p_2,q_2}\}, \mathrm{Re}\{\boldsymbol{\Sigma}_{p_1,q_1}\}) = -2 \cdot \mathrm{Im}\{[\boldsymbol{\Sigma}^{-1}]_{q_2,p_1} \cdot [\boldsymbol{\Sigma}^{-1}]_{q_1,p_2} + [\boldsymbol{\Sigma}^{-1}]_{q_2,q_1} \cdot [\boldsymbol{\Sigma}^{-1}]_{p_1,p_2}\} \quad (24)$$

$$\mathcal{I}(\mathrm{Im}\{\boldsymbol{\Sigma}_{p_1,q_1}\}, \mathrm{Im}\{\boldsymbol{\Sigma}_{p_2,q_2}\}) = \mathcal{I}(\mathrm{Im}\{\boldsymbol{\Sigma}_{p_2,q_2}\}, \mathrm{Im}\{\boldsymbol{\Sigma}_{p_1,q_1}\}) = 2 \cdot \mathrm{Re}\{-[\boldsymbol{\Sigma}^{-1}]_{q_2,p_1} \cdot [\boldsymbol{\Sigma}^{-1}]_{q_1,p_2} + [\boldsymbol{\Sigma}^{-1}]_{q_2,q_1} \cdot [\boldsymbol{\Sigma}^{-1}]_{p_1,p_2}\} \quad (25)$$

for $p_1 = q_1$ and $p_2 > q_2$

$$\mathcal{I}(\boldsymbol{\Sigma}_{p_1,p_1}, \mathrm{Re}\{\boldsymbol{\Sigma}_{p_2,q_2}\}) = \mathcal{I}(\mathrm{Re}\{\boldsymbol{\Sigma}_{p_2,q_2}\}, \boldsymbol{\Sigma}_{p_1,p_1}) = 2 \cdot \mathrm{Re}\{[\boldsymbol{\Sigma}^{-1}]_{q_2,p_1} \cdot [\boldsymbol{\Sigma}^{-1}]_{p_1,p_2}\} \quad (26)$$

$$\mathcal{I}(\boldsymbol{\Sigma}_{p_1,q_1}, \mathrm{Im}\{\boldsymbol{\Sigma}_{p_2,q_2}\}) = \mathcal{I}(\mathrm{Im}\{\boldsymbol{\Sigma}_{p_2,q_2}\}, \boldsymbol{\Sigma}_{p_1,q_1}) = -2 \cdot \mathrm{Im}\{[\boldsymbol{\Sigma}^{-1}]_{q_2,p_1} \cdot [\boldsymbol{\Sigma}^{-1}]_{p_1,p_2}\} \quad (27)$$

and, for $p_1 = q_1$ and $p_2 = q_2$

$$\mathcal{I}(\boldsymbol{\Sigma}_{p_1,p_1}, \boldsymbol{\Sigma}_{p_2,p_2}) = |[\boldsymbol{\Sigma}^{-1}]_{p_1,p_2}|^2 \quad (28)$$

6. Simulation Results

We use numerical simulations to evaluate performance of the proposed iterative channel estimation and decoding scheme for a turbo coded MIMO system in a frequency-flat correlated Rayleigh fading environment with $n_\mathrm{T} = 2$ transmit and $n_\mathrm{R} = 2$ receive antennas. Our performance metrics are the average mean-square error (MSE), bit error rate (BER), and frame error rate (FER), averaged over random channel realizations generated using an independent identically distributed Rayleigh fading model with unit-variance channel coefficients. The Alamouti transmission scheme [1] was used to generate the space-time codewords $\mathbf{X}_l$, implying $K' = K = 2$. The transmitted symbols $\{s_k^{(l)}\}$ were generated from a 4-PSK constellation (i.e., $M = 4$) with normalized energy. The space-time codewords were transmitted in R blocks as one frame, and each block consisted of $L_\mathrm{p} = 2$ pilot codewords followed by $L = 32$ data codewords. The signal was corrupted by additive complex Gaussian noise with spatial noise covariance matrix $\boldsymbol{\Sigma}$ whose (p, q)th element is

$$\boldsymbol{\Sigma}_{p,q} = \sigma^2 \cdot 0.9^{|p-q|} \cdot \exp[j(\pi/2)(p-q)] \quad (29)$$

which is the noise covariance model used in Reference [18] (see also references therein). For simplicity, we assume that $\boldsymbol{\Sigma}$ does not change within the data frame, but this knowledge is not used in the channel and noise covariance estimation. The turbo code consisted of two parallel concatenated (37, 21) recursive systematic convolutional codes connected with a random interleaver. Puncturing was employed to achieve the code rate $R_\mathrm{c} = 1/2$. The bit signal-to-noise ratio (SNR) per receive antenna was defined as

$$\mathrm{SNR} = 10\log_{10}\left[\frac{L + L_\mathrm{p}}{L} \cdot \frac{n_\mathrm{T} K}{K' \log_2(M) \cdot R_\mathrm{c} \cdot \sigma^2}\right] = 10\log_{10}\left[\frac{L + L_\mathrm{p}}{L} \cdot \frac{n_\mathrm{T}}{\sigma^2}\right] \ (\mathrm{dB}) \quad (30)$$

To initialize the EM algorithm, four iterations of the iterative weighted ILSP method were carried out.

We compare the proposed EM-based scheme with an iterative receiver using deterministic ML channel estimation with soft decision feedback, which is similar to the iterative receiver derived in Section 4, but with the channel estimation algorithm replaced by the deterministic ML method [3,7,11]. The deterministic ML channel estimator utilizes the expectations of the coded symbols computed from the extrinsic information produced by the decoder. Both methods were initialized using the iterative weighted ILSP method.

Figure 3 shows the BER performance versus bit SNR per receive antenna with $R = 16$ blocks. Three iterations have been carried out between the EM channel estimators and the turbo decoder. Clearly, iterating between the channel estimation and turbo decoding can improve the error performance for both

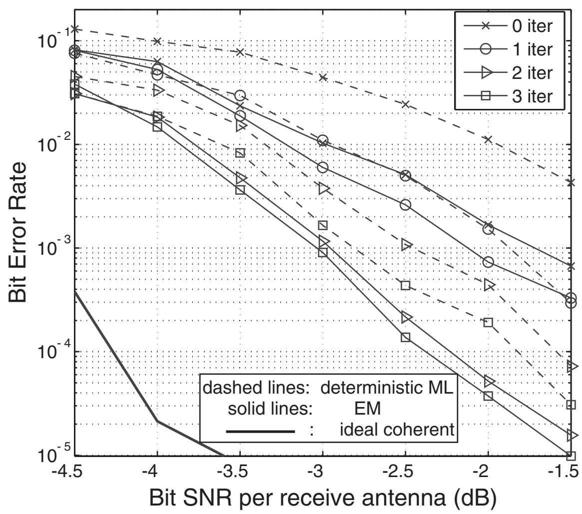


Fig. 3. BER for EM-based, deterministic ML, and ideal coherent detectors with $R = 16$.

EM and deterministic ML methods. As a comparison, we also present the performance of ideal coherent detector for exactly known $\mathbf{H}$ and $\mathbf{\Sigma}$. After three iterations, our method outperforms the deterministic ML by about 0.6 dB at $\text{BER} = 10^{-4}$, and comes within about 2 dB of the performance of the ideal coherent detector.

Next, we study the performance of the proposed EM-based scheme for long frame length, see Figures 4 and 5. Each data frame consisted of $R = 64$ blocks, and there were six iterations between the EM channel estimators and the turbo decoder.

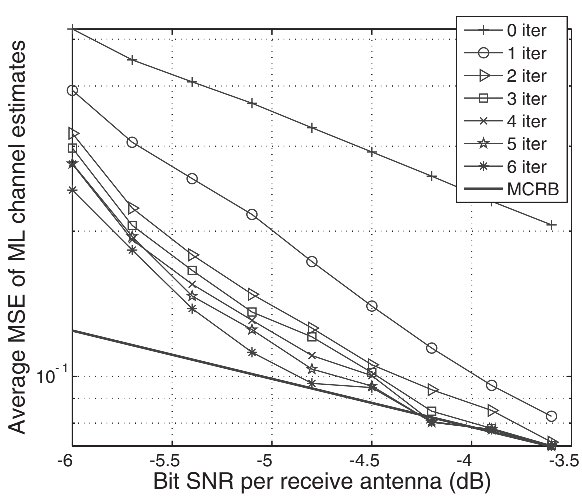


Fig. 4. MSE of the channel estimates for EM-based method with $R = 64$.

In Figure 4, the average MSE of the channel estimates improves as the number of iterations between the channel estimation and turbo decoding increases, and reaches the MCRB at $\text{SNR} \geq -4$ dB. Although the MCRB is the CRB assuming known information symbols, and hence a lower bound of the exact CRB, it can be regarded as the exact CRB when $\text{SNR} \geq -4$ dB since the BER has become very small (see Figure 5(a)) such that almost all the information symbols are correctly decoded.

The BER and FER performances of the proposed scheme are depicted in Figure 5(a), (b), respectively. After three iterations, the improvement provided by more iterations becomes smaller and smaller: an effect of diminishing returns. Such a saturation effect is more obvious in the Figure 5(b). The BER performance difference between our method and the ideal coherent detector becomes negligible at a BER level of 10^{-6}, which is achieved only after three iterations. Unlike the BER performance which depends on the specific turbo code used in the system, FER provides a performance measurement of the iterative receiver itself. It is seen from Figure 5(b) that upon convergence, our method approaches the performance of the ideal coherent detector, which is a lower bound of the performance for such MIMO systems.

The method proposed in this paper takes into account spatial correlation of the noise. To demonstrate the importance of this factor, we compare the proposed method with another EM-based iterative receiver, termed white-noise EM method, under a spatial correlated noise scenario. The method proposed in Section 4 differs from the white-noise EM method by the channel estimation algorithm. The white-noise EM algorithm assumes spatially white noise, that is $\mathbf{\Sigma} = \sigma^2 \mathbf{I}_{n_R}$, where σ^2 is the variance parameter to be estimated. A similar algorithm assuming known noise variance is proposed in Reference [10]. Figure 6 shows the BER performance of the white-noise EM method after three and six iterations between channel estimation and turbo decoding. As a comparison, the performance of the ideal coherent detector is also provided. It is seen that the proposed method has a 7 dB advantage over the white-noise EM method at a BER of 10^{-5} after both three and six iterations.

7. Conclusions

We developed an EM-based iterative channel estimation and decoding scheme for a coded system over

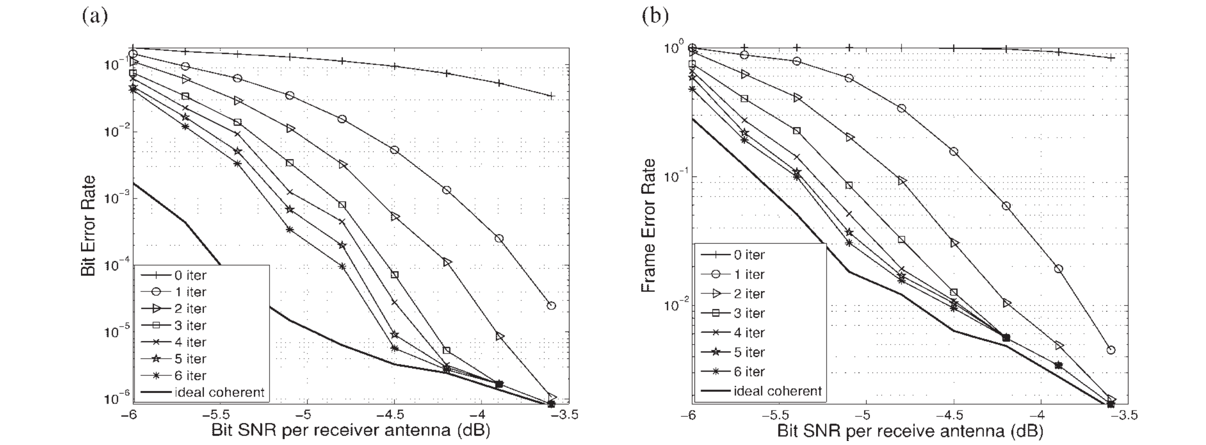


Fig. 5. BER and FER for EM-based and ideal coherent detectors with $R = 64$: (a) BER versus bit SNR per receive antenna; (b) FER versus bit SNR per receive antenna.

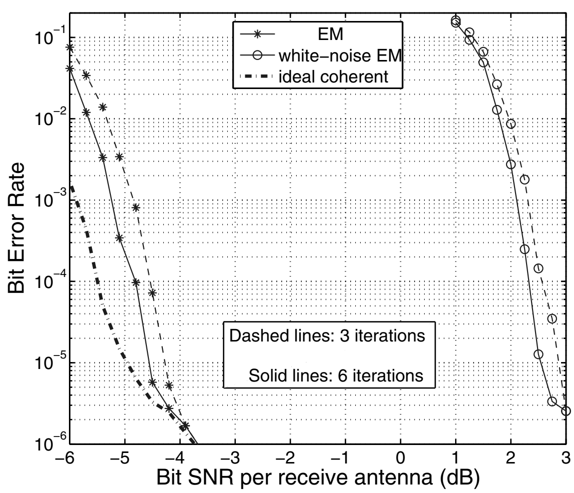


Fig. 6. BER for EM-based, white-EM based, and ideal coherent detectors with $R = 64$.

MIMO Rayleigh block fading channels in spatially correlated noise. By exchanging the extrinsic information of the transmitted symbols, both the channel estimation and the decoding can be improved. We also presented the MCRBs for the unknown parameters. Numerical simulations demonstrated the good performance of the proposed method and other competitive schemes.

One possible extension of this paper is to develop an adaptive version of the channel estimation algorithm that can account for continuously varying channels (as opposed to the block-fading scenario considered here) and also reduce the EM algorithm complexity. It is also of interest to adapt the algorithm for frequency-selective MIMO channels, possibly in combination with orthogonal frequency-division multiplexing (OFDM).

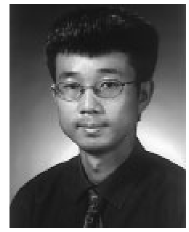
References

1. Alamouti SM. A simple transmit diversity technique for wireless communications. *IEEE Journal on Selected Areas in Communications* 1998; **16**(8): 1451–1458.
2. Baranski Z, Haimovich AM, Garcia-Frias J. EM-based iterative receiver for space-time coded modulation with noise variance estimation. In *Proceedings of GLOBECOM*, Vol. 1, 2002; pp. 355–359.
3. Bohlin P, Lundberg M. A blind and robust space-time receiver for a turbo coded system. In *Proceedings of Asilomar Conference on Signals, Systems & Computers*, Vol. 1, 2000; pp. 532–536.
4. Cozzo C, Hughes BL. Joint channel estimation and data detection in space-time communications. *IEEE Transactions on Communications* 2003; **51**(8): 1266–1270.
5. Dempster AP, Laird NM, Rubin DB. Maximum likelihood from incomplete data via the EM algorithm, *Journal of the Royal Statistical Society* 1977; **39**: 1–38.
6. Dogandžić A, Mo W, Wang Z. Semi-blind SIMO flat-fading channel estimation in unknown spatially correlated noise using the EM algorithm. *IEEE Transactions on Signal Processing* 2004; **52**(6): 1791–1797.
7. Dogandžić A, Nehorai A. Space-time fading channel estimation and symbol detection in and unknown spatially correlated noise. *IEEE Transactions on Signal Processing* 2002; **50**(3): 457–474.
8. Ganesan G, Stoica P. Space-time block codes: a maximum SNR approach. *IEEE Transactions on Information Theory* 2001; **47**(4): 1650–1656.
9. Gini F, Reggiannini R, Mengali U. The modified Cramer–Rao bound in vector parameter estimation. *IEEE Transactions on Communications* 1998; **46**(1): 52–60.
10. González-López M, Míguez J, Castedo L. Turbo aided maximum likelihood channel estimation for space-time coded systems. *IEEE International Symposium on Personal, Indoor and Mobile Radio Communications* 2002; **1**: 364–368.

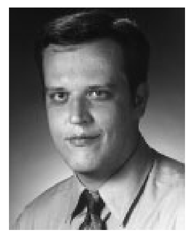
11. Larsson EG, Stoica P, Li J. Orthogonal space-time block codes: maximum likelihood detection for unknown channels and unstructured interferences. *IEEE Transactions on Signal Processing* 2003; **51**(2): 362–372.
12. Li Y, Georghiades CN, Huang G. Iterative maximum-likelihood sequence estimation for space-time and coded systems. *IEEE Transactions on Communications* 2001; **49**(6): 948–951.
13. Mo W, Wang Z, Dogandžić A. Iterative channel estimation and decoding for coded MIMO system in unknown spatially correlated noise. In *Proceedings of the 41th Allerton Conference*, 2003; pp. 324–333.
14. Muirhead RJ. *Aspects of Multivariate Statistical Theory.* Wiley: New York, USA, 1982.
15. Ranheim A, des Rosiers AP, Siegel PH. An iterative receiver algorithm for space-time encoded signals. In *Proceedings Asilomar Conference on Signals, Systems & Computers*, Vol. 1, 2000; pp. 516–520.
16. Sandell M, Luschi C, Strauch P, Yan R. Iterative channel estimation using soft decision feedback. In *Proceedings of GLOBECOM*, Vol. 6, 1998; pp. 3728–3733.
17. Tarokh V, Jafarkhani H, Calderbank AR. Space-time block codes from orthogonal designs. *IEEE Transactions on Information Theory* 1999; **45**(5): 1456–1467.
18. Viberg M, Stoica P, Ottersten B. Maximum likelihood array processing in spatially correlated noise and fields using parameterized signals. *IEEE Transactions on Signal Processing* 1997; **45**(4): 996–1004.

Authors' Biographies

Wei Mo was born in Shenyang, China, in 1976. He received his B.S. and M.Sc. degrees in Engineering Physics from the Tsinghua University, China, in 1998 and 2001, respectively. He is currently pursuing his Ph.D. with the Department of Electrical and Computer Engineering at Iowa State University. His general interests include wireless communication, statistical signal processing, and information theory. His current research is mainly focused on MIMO systems and wireless sensor networks.



Zhengdao Wang was born in Dalian, China, in 1973. He received his B.S. degree in Electrical Engineering and Information Science from the University of Science and Technology of China (USTC), 1996, the M.Sc. degree in Electrical and Computer Engineering from the University of Virginia, 1999, and Ph.D. in Electrical and Computer Engineering from the University of Minnesota, 2002. He is now with the Department of Electrical and Computer Engineering at the Iowa State University. His interests are in the areas of signal processing, communications, and information theory. He was the co-recipient of the IEEE Signal Processing Magazine Best Paper Award in 2003 and the IEEE Communications Society Marconi Paper Prize Award in 2004.



Aleksandar Dogandzic received his Dipl. Ing. degree (summa cum laude) in Electrical Engineering from the University of Belgrade, Yugoslavia, in 1995, and his M.S. degree and Ph.D. in Electrical Engineering and Computer science from the University of Illinois at Chicago (UIC) in 1997 and 2001, respectively, under the guidance of professor Arye Nehorai. In August 2001, he joined the Department of Electrical and Computer Engineering, Iowa State University, Ames, IA, as an assistant professor. His research interests are in statistical signal processing theory and applications. In 1996, Dr Dogandzic received the Distinguished Electrical Engineering M.S. Student Award by the Chicago Chapter of the IEEE Communications Society. He was awarded the Aileen S. Andrew Foundation Graduate Fellowship in 1997, the UIC University Fellowship in 2000, and the 2001 Outstanding Thesis Award in the Division of Engineering, Mathematics, and Physical Sciences, UIC. He is the recipient of the 2003 Young Author Best Paper Award and 2004 Signal Processing Magazine Award by the IEEE Signal Processing Society.