

MIMO Channel Estimation Using the Variational Expectation-Maximization Method

Zhengwei Jiang*, Teng Joon Lim*, Roya Doostnejad[†] and Taiwen Tang*

Abstract—In this paper, the variational expectation-maximization (VEM) algorithm is used to provide channel estimates in a space-time decoder. The proposed estimator works for many types of space-time codes (STC), including full-rate full-diversity (FRFD) codes, Bell Laboratories layered space-time architecture (BLAST) and orthogonal STC's. The principle idea is to treat the channel coefficients as the unknown parameters, and the transmitted symbols as the unobserved variables in the EM algorithm. The posterior distribution of the symbols is approximated by a factorized distribution, whose Kullback-Liebler divergence with the true distribution is minimized, thereby explaining the “variational” aspect of the technique. The channel estimates may then be used for coherent decoding – here we use the K -best detector for a 2×2 system. Simulation results show that the new channel estimator can reduce system complexity greatly compared with traditional schemes, while achieving near-ideal performance.

I. INTRODUCTION

Coherent detection of space-time codes requires accurate channel estimates, usually acquired through a training period in which known pilot symbols are transmitted. However, even with maximum likelihood channel estimation, a non-negligible estimation error exists, whose variance is no smaller than the Cramer-Rao lower bound. The existence of this residual estimation error is often accepted as unavoidable, and with some types of space-time codes have only a small detrimental effect on performance.

However, with sophisticated STC's that provide both full rate and full diversity, such as threaded algebraic space time (TAST) codes [1], [2] and perfect space time block (PSTB) codes [3], [4], channel estimation accuracy is much more critical, as we will show in simulations later in the paper. It is then sensible to explore the use of joint detection and estimation (or decision-directed estimation), during the data transmission phase, to further refine the estimates.

Given imperfect CSI with a known Gaussian error distribution, ML symbol detection can be performed [5]. However this is not a practical solution when the number of bits to be detected jointly is even moderately large. An alternative method is the expectation-maximization (EM) approach, which iterates between an E-step and an M-step, in order to obtain the ML estimate of the unknown channel. The EM algorithm has two sets of unknowns – hidden variables and parameters. By treating the channel as the parameter and the

transmitted symbols as the hidden variables [6], [7], the E-step requires the *a posteriori* probability (APP) distribution of the symbols, assuming the most recent channel estimate is exact. Since the symbol APP is difficult to obtain exactly, it has been approximated in [6], [7] using the outputs of a soft-in soft-out error control decoder. The resulting scheme can be interpreted as either an iterative (or turbo) detector, or an EM-based receiver. Since decoding is performed multiple times, complexity is very high; performance can also be poor if the initial channel estimates (obtained from a training period) are weak, resulting in poor reliability information delivered to the decoder.

If instead the channel is the hidden variable and the symbols are the parameters in the EM algorithm, then only hard estimates of the symbols are produced, as in [8]. Such an algorithm cannot produce the soft inputs essential in the decoding of powerful codes like LDPC (low-density parity check) and Turbo codes. The APP of the channel is obtained but this is of little obvious use.

We propose to use the structure of Figure I, where channel estimation is performed using the variational EM algorithm [9] to be described later. Decoding is performed only once, and given that much of the complexity in a receiver is due to the decoder, the proposed structure is less complex than those of [6], [7]. Channel estimation performance is enhanced by the VEM method, which after a few inner iterations converges to a value close to the ML channel estimate. Detection performance is therefore excellent, as will be demonstrated in simulations in Section IV.

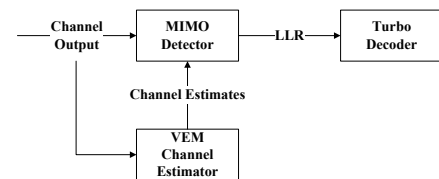


Fig. 1. The receiver using VEM channel estimator

II. SYSTEM DESCRIPTION

Consider an $N_R \times N_T$ quasi-static flat fading MIMO channel. The baseband-equivalent signal transmitted over T channel uses and N_T antennas is written in a column vector

$$\mathbf{x}(l) = [x_{1,1}(l), x_{1,2}(l), \dots, x_{1,N_T}(l), \dots, x_{T,N_T}(l)]^T.$$

Assuming that an error-control codeword spans LT channel uses, the transmitted signal can be written in a $TN_T \times L$ matrix

$$\mathbf{X} = [\mathbf{x}(1), \dots, \mathbf{x}(L)].$$

*Zhengwei Jiang, Teng Joon Lim and Taiwen Tang are with the Department of Electrical & Computer Engineering, University of Toronto, Toronto, Ontario, Canada.

[†]Roya Doostnejad is with Redline Communications Inc., Markham, Ontario, Canada

Defining the channel output as $\mathbf{Y} \in \mathbb{C}^{TN_R \times L}$, then

$$\mathbf{Y} = (\mathbf{I}_T \otimes \mathbf{H}_c) \mathbf{X} + \mathbf{W} = (\mathbf{I}_T \otimes \mathbf{H}_c) \mathbf{G} \mathbf{U} + \mathbf{W}, \quad (1)$$

where \otimes is the Kronecker product, \mathbf{I}_T is the $T \times T$ identity matrix, \mathbf{H}_c is the $N_R \times N_T$ channel matrix, $\mathbf{U} \in \mathbb{Q}^{TN_T \times L}$ are the modulated symbols drawn from the symbol constellation \mathbb{Q} , and \mathbf{W} is a matrix of i.i.d. Gaussian random variables. The decomposition of \mathbf{X} into $\mathbf{G} \mathbf{U}$ unifies various space-time coding schemes, e.g. for V-BLAST (Vertical BLAST) \mathbf{G} is the $N_T \times N_T$ identity matrix and $T = 1$. We also identify \mathbf{G} as an $N_T^2 \times N_T^2$ matrix and $T = N_T$ for PSTB and TAST codes. The definition of \mathbf{G} can be found in [2] [3] [4], e.g. for PSTB in a 2×2 channel, \mathbf{G}_{PSTB} can be expressed as

$$\mathbf{G}_{PSTB} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & i \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix} \begin{pmatrix} \mathbf{M} & \mathbf{0} \\ \mathbf{0} & \mathbf{M} \end{pmatrix}, \quad (2)$$

$$\mathbf{M} \doteq \begin{pmatrix} 0.63 - 0.39i & 1.02 - 0.63i \\ 0.63 + 1.02i & -0.39 - 0.63i \end{pmatrix}.$$

For pilot-based channel estimation, training symbols are transmitted without space-time modulation/coding. The signal received over the pilot interval can be written as

$$\mathbf{Y}_p = \mathbf{H}_c \mathbf{U}_p + \mathbf{W}, \quad (3)$$

where \mathbf{U}_p is the matrix of pilot symbols. Since \mathbf{U}_p is known at the receiver, the ML channel estimate, denoted as $\hat{\mathbf{H}}_c^{(0)}$, is

$$\hat{\mathbf{H}}_c^{(0)} = (\mathbf{Y}_p \mathbf{U}_p^T) (\mathbf{U}_p \mathbf{U}_p^T)^{-1}. \quad (4)$$

The estimation error is therefore

$$\mathbf{E}_c = \mathbf{H}_c - \hat{\mathbf{H}}_c^{(0)} = (\mathbf{W} \mathbf{U}_p^T) (\mathbf{U}_p \mathbf{U}_p^T)^{-1}. \quad (5)$$

Assuming $E[\mathbf{W}^T \mathbf{W}] = \sigma_n^2 \mathbf{I}$, then we have

$$E[\mathbf{E}_c^T \mathbf{E}_c] = \sigma_n^2 (\mathbf{U}_p \mathbf{U}_p^T)^{-1}. \quad (6)$$

From this equation, we see that the variance of the channel estimation error is proportional to the channel noise and inversely proportional to the length of the pilot sequence. This is the model we will use throughout the paper.

III. VARIATIONAL EM CHANNEL ESTIMATOR

In [10], an important link was found between the classes of variational inference (VI) algorithms and expectation-maximization (EM) algorithms. In brief, a VI algorithm postulates a simple, parameterized probability distribution in place of the true *a posteriori* distribution of the unknown variables in a problem. It then finds the parameters that minimize the Kullback-Liebler (KL) divergence between the two, and uses the postulated distribution for inference. The EM algorithm in its traditional description is for ML parameter estimation – it proceeds iteratively through the so-called E- and M-steps, and is guaranteed to converge to a local maximum of the likelihood function of the unknown parameters.

The link between the two is that, if in a VI algorithm one postulates a factorized distribution with two sets of variables, say \mathbf{x} and Θ , and uses a point distribution for Θ while keeping the distribution of \mathbf{x} optimal, then EM and VI are identical.

To be precise, the variational view of the EM algorithm is that $p(\mathbf{x}, \Theta | \mathbf{y})$, where \mathbf{y} is the vector of observations, is approximated as

$$p(\mathbf{x}, \Theta | \mathbf{y}) \approx Q(\mathbf{x}) \delta(\Theta - \Theta^{(t)}) \quad (7)$$

at the t -th iteration, where $\Theta^{(t)}$ denotes the estimate of Θ to be obtained in the current iteration. Then we form the variational free energy (FE), or KL divergence between the postulated and true distributions:

$$\mathcal{F}(Q(\mathbf{x}), \Theta^{(t)}) = \int Q(\mathbf{x}) \delta(\Theta - \Theta^{(t)}) \log \frac{Q(\mathbf{x}) \delta(\Theta - \Theta^{(t)})}{p(\mathbf{x}, \Theta | \mathbf{y})} d\mathbf{x} d\Theta. \quad (8)$$

Finally, we find the parameters of $Q(\mathbf{x})$ and the value of $\Theta^{(t)}$ that minimize \mathcal{F} , and use these for the next iteration. Convergence to a local maximum of $p(\mathbf{y} | \Theta)$ is assured, if

$$Q(\mathbf{x}) = p(\mathbf{x} | \mathbf{y}, \Theta^{(t-1)}).$$

However, if the above form of $Q(\mathbf{x})$ is hard to work with, we can replace it with another approximation, and hence arrive at the variational EM (VEM) method.

In the following, we will detail the application of VEM as a channel estimator in an $N_R \times N_T$ MIMO channel. In order to simplify the development, we rewrite the system in (1) using real-valued signals and with 4-QAM as the modulation scheme. Other modulation formats can also be used but the mapping from symbols to bits is more involved. Let us denote

$$\begin{aligned} \mathbf{R} &= [\text{real}(\mathbf{Y})^T, \text{imag}(\mathbf{Y})^T]^T \\ \mathbf{B} &= [\text{real}(\mathbf{U})^T, \text{imag}(\mathbf{U})^T]^T \\ \mathbf{N} &= [\text{real}(\mathbf{W})^T, \text{imag}(\mathbf{W})^T]^T \end{aligned} \quad (9)$$

$$\begin{aligned} \tilde{\mathbf{H}} &= (\mathbf{I}_T \otimes \mathbf{H}_c) \mathbf{G} \\ \mathbf{H} &= \begin{pmatrix} \text{real}(\tilde{\mathbf{H}}) & -\text{imag}(\tilde{\mathbf{H}}) \\ \text{imag}(\tilde{\mathbf{H}}) & \text{real}(\tilde{\mathbf{H}}) \end{pmatrix}, \end{aligned}$$

then we have

$$\mathbf{R} = \mathbf{H} \mathbf{B} + \mathbf{N}. \quad (10)$$

In (10), \mathbf{B} is a matrix of binary (Bernoulli) random variables. We will treat \mathbf{B} as the hidden variables in VEM. We also assume bit interleaved coded modulation (BICM) is implemented, thus \mathbf{b}_l , \mathbf{r}_l , $l = 1, 2, \dots, L$ are independent, where \mathbf{b}_l , \mathbf{r}_l are the l^{th} columns of \mathbf{B} and \mathbf{R} respectively. Besides, we will use the mean field approximation

$$Q_l(\mathbf{b}_l) = \prod_{k=1}^{2N_T T} Q_{lk}(\mathbf{b}_l(k)),$$

where $\mathbf{b}_l(k) \in \pm 1$ is the k^{th} value of \mathbf{b}_l . Therefore, we define the variational free energy using the following distributions

$$\begin{cases} Q_l(\mathbf{b}_l) = \prod_{k=1}^{2N_T T} \gamma_l(k)^{1+\mathbf{b}_l(k)/2} (1 - \gamma_l(k))^{1-\mathbf{b}_l(k)/2} \\ p(\mathbf{b}_l) = \prod_{k=1}^{2N_T T} \xi_l(k)^{1+\mathbf{b}_l(k)/2} (1 - \xi_l(k))^{1-\mathbf{b}_l(k)/2} \\ p(\mathbf{r}_l | \mathbf{b}_l) = \mathcal{N}(\mathbf{H} \mathbf{b}_l, \sigma^2 \mathbf{I}) \end{cases}, \quad (11)$$

where $\xi_l(k)$, $\gamma_l(k)$ are the prior and posterior probabilities of $\mathbf{b}_l(k) = 1$ respectively, and $\sigma^2 = \sigma_n^2/2$. Note that $\gamma_l(k)$ parameterizes the distribution of $\mathbf{b}_l(k)$, and hence the parameters to be optimized in $Q_l(\mathbf{b}_l)$ are $\{\gamma_l(k)\}_{k=1}^{2N_T T}$.

With \mathbf{H}_c as the parameters, we can now formulate the FE function as follows

$$\mathcal{F} = \int \delta(\mathbf{H}_c - \hat{\mathbf{H}}_c) \prod_{l=1}^L Q_l(\mathbf{b}_l) \log \frac{\delta(\mathbf{H}_c - \hat{\mathbf{H}}_c) \prod_{l=1}^L Q_l(\mathbf{b}_l)}{p(\mathbf{B}, \mathbf{H}_c | \mathbf{R})} d\mathbf{H}_c d\mathbf{b}_1 d\mathbf{b}_2 \dots d\mathbf{b}_L. \quad (12)$$

Following the mathematical manipulations in [9], the FE can be re-written as

$$\begin{aligned} \mathcal{F} &= -\log p(\hat{\mathbf{H}}_c) + \sum_{l=1}^L \left(\int_{\mathbf{b}_l} Q_l(\mathbf{b}_l) \log \frac{Q_l(\mathbf{b}_l)}{p(\mathbf{b}_l, \mathbf{r}_l | \hat{\mathbf{H}}_c)} d\mathbf{b}_l \right) \\ &= \sum_{l=1}^L \left(\int_{\mathbf{b}_l} Q_l(\mathbf{b}_l) \log \frac{Q_l(\mathbf{b}_l)}{p(\mathbf{b}_l) p(\mathbf{r}_l | \mathbf{b}_l, \hat{\mathbf{H}}_c)} d\mathbf{b}_l \right) \\ &\quad - \log p(\hat{\mathbf{H}}_c). \end{aligned} \quad (13)$$

Minimizing the FE function above iteratively, we arrive at the VEM algorithm, consisting of two main steps as follows.

Initialization Use the pilot sequence to get the initial channel estimation $\hat{\mathbf{H}}_c^{(0)}$.

E-Step Minimize \mathcal{F} over the distribution of \mathbf{b}_l , parameterized by $\mathbf{m}_l = [2\gamma_l(1) - 1, 2\gamma_l(2) - 1, \dots, 2\gamma_l(2N_T T) - 1]^T$, i.e. in the j^{th} iteration,

$$\mathbf{m}_l^{(j)} = \arg \max_{\mathbf{m}_l} \int_{\mathbf{b}_l} Q_l(\mathbf{b}_l) \frac{Q_l(\mathbf{b}_l)}{p(\mathbf{b}_l) p(\mathbf{r}_l | \mathbf{b}_l, \hat{\mathbf{H}}_c^{(j-1)})} d\mathbf{b}_l, \quad (14)$$

for $l = 1, 2, \dots, L$.

M-Step Minimize \mathcal{F} over $\hat{\mathbf{H}}_c$, i.e.

$$\hat{\mathbf{H}}_c^{(j)} = \arg \min_{\hat{\mathbf{H}}_c} \sum_{l=1}^L \left(\int_{\mathbf{b}_l} Q_l^{(j)}(\mathbf{b}_l) \frac{Q_l^{(j)}(\mathbf{b}_l)}{p(\mathbf{b}_l) p(\mathbf{r}_l | \mathbf{b}_l, \hat{\mathbf{H}}_c)} d\mathbf{b}_l \right) - \log p(\hat{\mathbf{H}}_c). \quad (15)$$

The E-step is identical to what was presented in [9], and so we will provide only the final result. The k^{th} element of \mathbf{m}_l should be updated iteratively according to

$$\log \frac{1 + \mathbf{m}_l(k)}{1 - \mathbf{m}_l(k)} \leftarrow \log \frac{1 + \mathbf{p}_l(k)}{1 - \mathbf{p}_l(k)} + \frac{2}{\sigma^2} (\eta_k^T \mathbf{r}_l - \beta_k^T \mathbf{m}_l) \quad (16)$$

where $\mathbf{p}_l = [2\xi_l(1) - 1, 2\xi_l(2) - 1, \dots, 2\xi_l(2N_T T) - 1]^T$, η_k is the k^{th} column of \mathbf{H} , and β_k is the k^{th} column of $\mathbf{H}^T \mathbf{H} - \text{diag}(\mathbf{H}^T \mathbf{H})$.

The M-Step requires some new derivations, details of which are in the Appendix. Replacing the new estimate of $\mathbf{m}_l^{(j)}$, $l = 1, 2, \dots, L$ obtained from (16) into (15), omitting the terms independent of \mathbf{H}_c and assuming \mathbf{H}_c to be an i.i.d. circularly symmetric complex Gaussian random matrix with a mean of

$\hat{\mathbf{H}}_c^{(0)}$ and a covariance of σ_e^2 for each element of the matrix, we have

$$\mathcal{F} = \text{const} + \frac{1}{2\sigma_e^2} \|\mathbf{H}_c\|_F^2 + \sum_{l=1}^L \left\{ \frac{1}{2\sigma^2} [\text{tr}(\mathbf{H}^T \mathbf{H}) - 2\mathbf{r}_l^T \mathbf{H} \mathbf{m}_l + \mathbf{m}_l^T (\mathbf{H}^T \mathbf{H} - \text{diag}(\mathbf{H}^T \mathbf{H})) \mathbf{m}_l] \right\}. \quad (17)$$

From the derivations in the Appendix, we can get the solution which minimizes the function above, i.e. the update of channel estimate $\hat{\mathbf{H}}_c^{(j)} = \hat{\mathbf{H}}_R^{(j)} + i\hat{\mathbf{H}}_I^{(j)}$,

$$\begin{cases} \hat{\mathbf{H}}_R^{(j)} = (\mathbf{P}_6^T \mathbf{P}_1^{-1} - \mathbf{P}_5^T \mathbf{P}_2^{-1})(\mathbf{P}_1 \mathbf{P}_2^{-1} - \mathbf{P}_2^T \mathbf{P}_1^{-1})^{-1} / 2 \\ \hat{\mathbf{H}}_I^{(j)} = (\mathbf{P}_6^T \mathbf{P}_2^{-T} - \mathbf{P}_5^T \mathbf{P}_1^{-1})(\mathbf{P}_2 \mathbf{P}_1^{-1} - \mathbf{P}_1 \mathbf{P}_2^{-T})^{-1} / 2 \end{cases} \quad (18)$$

Based on the discussion above, we can summarize the VEM channel estimator in Table I. Two points should be noticed in the proposed algorithms. The initial estimate of the channel is obtained from the training sequence. We use MMSE detector to set the initial value of \mathbf{m}_l to avoid converging to a local minimum.

TABLE I
VEM CHANNEL ESTIMATOR

Initialization: $\hat{\mathbf{H}}_c^{(0)}$ is the pilot assisted channel estimation, $\mathbf{LLR}_{pri} = \mathbf{0}$.
Iterative Search for optimized estimation:

```

for j = 1 : J do
  MMSE LLR calculator  $\Rightarrow \mathbf{m}_l, l = 1, 2, \dots, L$ 
  for i = 1 : I (Iteration according to (16)) do
    for k = 1 : 2NTT do
       $\mathbf{LLR}_{pos}(\mathbf{b}_l(k)) = \mathbf{LLR}_{pri}(\mathbf{b}_l(k)) + \frac{2}{\sigma^2} (\eta_k^T \mathbf{r}_l - \beta_k^T \mathbf{m}_l)$ 
       $\mathbf{m}_l(k) = \tanh[\mathbf{LLR}_{pos}(\mathbf{b}_l(k))]$ 
    end for
  end for
  Update the estimation of channel following (19) ~ (28)
end for
OUTPUT: output  $\hat{\mathbf{H}}_c^{(J)}$  into following MIMO detector and decoding.

```

IV. SIMULATIONS

In this section, we simulate the algorithm proposed in section III. The proposed algorithm using VEM is compared with the conventional EM based iterative detection and decoding (IDD). Frame error rate (FER) and mean square error (MSE) of the estimated channel are presented for both algorithms. Different space time codes are used for the simulations. We also analyze the complexity of the proposed algorithm.

A. Performance Analysis of VEM Channel Estimator

The 2×2 MIMO channel with i.i.d. Rayleigh fading with zero mean and unit variance is studied here. In our simulations, rate 1/2 Turbo code is used, whose component encoders use (7,5) RSC (recursive systematic convolution) code. The channel is assumed to be constant over 512 time slots. The pilot sequence is of 8 time slots. Thus we have $\sigma_e^2 = \sigma_n^2/4$ according to (6). Also, K-best detection is employed with $K = 15$ based on the VEM channel estimate [11].

In Fig. 2, we compare three groups of curves. The dotted lines present K-best detection and decoding in the presence of perfect CSI. The dashed lines present iterative joint K-best detection and decoding in the presence of imperfect CSI.

In this algorithm, the channel estimation is updated using the soft-outputs of the decoder in each iteration of detection and decoding [6] [7]. The solid line present our proposed algorithm. Illustrated by this figure, our algorithm using VEM is much better than the EM-based IDD with an around 3dB improvement and less than 1 dB loss from those of perfect CSI. Besides, we observe that PSTB codes are better than other STC's in the FER performance and are robust to the channel estimation error.

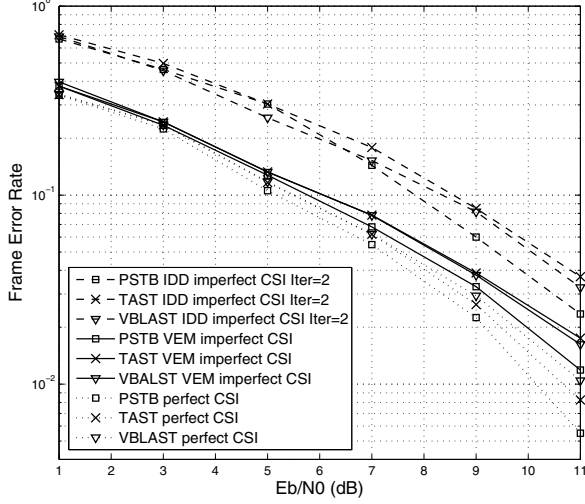


Fig. 2. The FER performances of PSTB codes, TAST codes, VBLAST using K -best detection and decoding in the presence of perfect CSI, using Iterative K -best Detection and Decoding in the presence of imperfect CSI (Iter=2) and using our proposed algorithm.

Fig. 3 shows the relations between the channel estimation error averaged over fading channels and the number of iterations for our proposed algorithm at different E_b/N_0 's. In this figure, we observe that MSE of our proposed algorithm using VEM decreases significantly at all different E_b/N_0 's.

We also compare the MSE versus E_b/N_0 performance using VEM with the performance using EM based IDD as shown in Fig. 4. The lower bound in this figure is calculated by assuming that all the symbols are recovered correctly. Obviously from this figure, our algorithm outperforms EM based IDD, especially when E_b/N_0 is low. From these groups of curves, we also conclude that PSTB codes are more robust to the channel estimation error.

B. Complexity Analysis

Let us consider the complexity of the Turbo decoder. Total complexity = $2 \times \text{Iteration times} \times \text{BCJR complexity}$, and BCJR complexity = The length of bits \times The number of operations per bit. Thus, the complexity of Turbo decoder can be derived as $2N_{iter}N_TTLN_{oper}$, where N_{iter} denotes the iteration number and N_{oper} denotes the number of operations per bit. Also we have a factor of two because there are two BCJR decoders for the Turbo decoding at the receiver. Then let us consider the complexity of VEM. From Table III, it is easy to find that the complexity of the iterative detector is $O(IJN_TTL)$, where $O(\cdot)$ denotes the order of magnitude. Clearly, when the terms I and J are not large (we choose $I = 5$ and $J = 5$ in

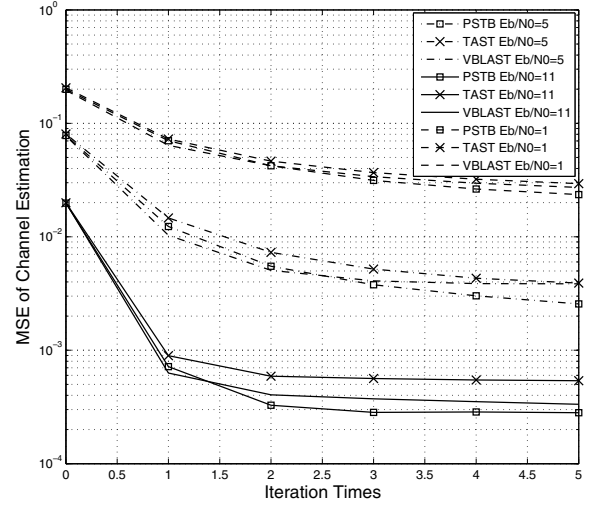


Fig. 3. The MSE performance of the channel estimation $\|\mathbf{H}_c - \hat{\mathbf{H}}_c^{(j)}\|_F^2$ when $E_b/N_0 = 1, 5, 11$ for different kinds of Space Time Codes using VEM.

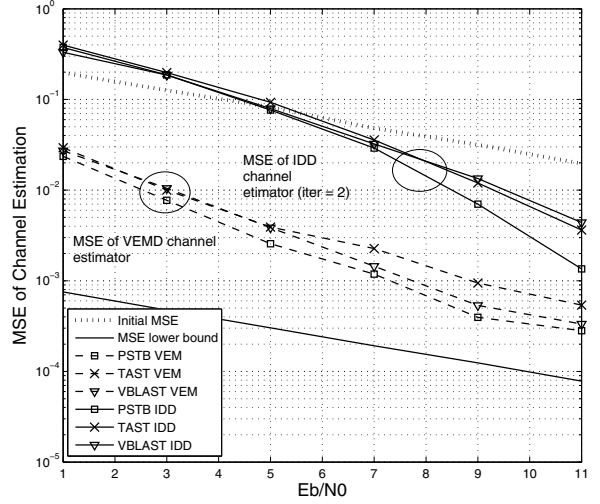


Fig. 4. The MSE comparison of the channel estimation $\|\mathbf{H}_c - \hat{\mathbf{H}}_c\|_F^2$ between our proposed algorithm and EM based IDD receiver, with 2 iterations of the latter.

the simulations), the complexity of the receiver mainly comes from the Turbo decoder.

In [6] [7], the complexity of the receivers is still mainly due to the Turbo decoding operations. Iterations between the detector and the decoder are required, which at least double the complexity for the decoder as compared with our proposed receiver.

V. CONCLUSION

In this paper, we proposed a novel low complexity channel estimator using VEM. We adopted the soft K -best detection algorithm to detect the signals based on the VEM channel estimates. The proposed receiver with joint K -best detection and VEM channel estimation works with different kinds of space time codes. We compared this proposed algorithm with the conventional EM based iterative detection and decoding algorithm in terms of FER and MSE of the estimated channels. The proposed algorithm improves the FER performance by around 3 dB in a 2×2 MIMO channel. The VEM channel

estimator also shows an order of magnitude lower of MSE of the channel estimates compared with the conventional EM based iterative detection and decoding algorithm in this MIMO channel.

REFERENCES

- [1] H. E. Gamal and M. O. Damen, "Universal Space-Time Coding," *IEEE Trans. Inf. Th.*, vol. 49, no. 5, pp. 1097–1119, May 2003.
- [2] M. O. Damen, H. E. Gamal, and N. C. Beaulieu, "Linear Threaded Algebraic Space-Time Constellations," *IEEE Trans. Inf. Th.*, vol. 49, no. 10, pp. 2372–2388, Oct 2003.
- [3] F. Oggier, G. Rekaya, J. Belfiore, and E. Viterbo, "Perfect Space-Time Block Codes," *IEEE Trans. Inf. Th.*, vol. 52, no. 9, pp. 3885–3902, Sep 2006.
- [4] P. Elia, K. R. Kumar, S. A. Pawar, P. V. Kumar, and H. Lu, "Explicit Space-Time Codes Achieving the Diversity-Multiplexing Gain Trade-off," *IEEE Trans. Inf. Th.*, vol. 52, no. 9, pp. 3869–3884, Sep 2006.
- [5] G. Taricco and E. Biglieri, "Space-Time Decoding With Imperfect Channel Estimation," *IEEE Trans. Wireless Commun.*, vol. 4, no. 4, pp. 1874–1888, Jul 2005.
- [6] M. A. Khalighi and J. J. Boutros, "Semi-Blind Channel Estimation Using the EM Algorithm in Iterative MIMO APP Detectors," *IEEE Trans. Wireless Commun.*, vol. 5, no. 11, pp. 3165–3173, Nov 2006.
- [7] J. Choi, "MIMO-BICM Iterative Receiver With the EM Based Channel Estimation and Simplified MMSE Combining With Soft Cancellation," *IEEE Trans. on Signal Processing*, vol. 54, no. 8, pp. 3247–3251, Aug 2006.
- [8] J. Choi, "An EM based Joint Data Detection and Channel Estimation Incorporating with Initial Channel Estimate," *IEEE Commun. Lett.*, vol. 12, no. 9, pp. 654–656, Sep 2008.
- [9] D. D. Lin and T. J. Lim, "A Variational Inference Framework for Soft-In Soft-Out Detection in Multiple-Access Channels," *IEEE Trans. Inf. Th.*, vol. 55, no. 5, pp. 2345–2364, May 2009.
- [10] R. M. Neal and G. E. Hinton, "A view of the EM algorithm that justifies incremental, sparse, and other variants," in *Learning in graphical models*, M. I. Jordan, Ed. 1999, pp. 355 – 368, MIT Press.
- [11] Z. Guo and M. Dong, "Algorithm and Implementation of the K-Best Sphere Decoding for MIMO Detection," *IEEE Jour. Select. Areas in Commun.*, vol. 24, no. 3, pp. 491 – 503, Mar. 2006.

APPENDIX

We will investigate the optimization of (17) over \mathbf{H}_c . Let $\mathbf{A} = \sum_{l=1}^L \mathbf{m}_l \mathbf{m}_l^T$ and $\mathbf{C} = 2 \sum_{l=1}^L \mathbf{m}_l \mathbf{r}_l^T$, the free energy function can be rewritten as

$$\mathcal{F} = \text{const} + \frac{1}{2\sigma_e^2} \|\mathbf{H}_c - \hat{\mathbf{H}}_c^{(0)}\|_F^2 + \frac{1}{2\sigma^2} \left\{ L \text{tr}(\mathbf{H}^T \mathbf{H}) + (\mathbf{H}^T \mathbf{H} - \text{diag}(\mathbf{H}^T \mathbf{H})) \mathbf{A} - \text{tr}(\mathbf{H} \mathbf{C}) \right\}. \quad (19)$$

Let $\mathbf{D} = \mathbf{A} + L\mathbf{I} - \text{diag}(\mathbf{A})$, after some mathematics manipulations

$$\mathcal{F} = \text{const} + \frac{1}{2\sigma_e^2} \|\mathbf{H}_c - \hat{\mathbf{H}}_c^{(0)}\|_F^2 + \frac{1}{2\sigma^2} \text{tr}(\mathbf{H}^T \mathbf{H} \mathbf{D} - \mathbf{H} \mathbf{C}). \quad (20)$$

Denotes $\text{real}(\tilde{\mathbf{H}})$ with $\tilde{\mathbf{H}}_R$, and $\text{imag}(\tilde{\mathbf{H}})$ with $\tilde{\mathbf{H}}_I$, and $\tilde{\mathbf{H}}$ is defined in (9). Since \mathbf{D} is symmetric, Let

$$\mathbf{D} = \begin{pmatrix} \mathbf{D}_1 & \mathbf{D}_2^T \\ \mathbf{D}_2 & \mathbf{D}_3 \end{pmatrix}, \quad \mathbf{C} = \begin{pmatrix} \mathbf{C}_1 & \mathbf{C}_3 \\ \mathbf{C}_2 & \mathbf{C}_4 \end{pmatrix}, \quad (21)$$

and define $\mathbf{E}_1 = \mathbf{D}_1 + \mathbf{D}_3$, $\mathbf{E}_2 = \mathbf{D}_2^T - \mathbf{D}_2$, $\mathbf{E}_3 = -(\mathbf{C}_1 + \mathbf{C}_4)$, $\mathbf{E}_4 = -(\mathbf{C}_3 - \mathbf{C}_2)$, the term $\text{tr}(\mathbf{H}^T \mathbf{H} \mathbf{D} - \mathbf{H} \mathbf{C})$ can be put

into

$$\begin{aligned} \text{tr}(\mathbf{H}^T \mathbf{H} \mathbf{D} - \mathbf{H} \mathbf{C}) &= \text{tr} \left((\tilde{\mathbf{H}}_R^T \tilde{\mathbf{H}}_R + \tilde{\mathbf{H}}_I^T \tilde{\mathbf{H}}_I) \mathbf{E}_1 + \right. \\ &\quad \left. + (\tilde{\mathbf{H}}_R^T \tilde{\mathbf{H}}_I - \tilde{\mathbf{H}}_I^T \tilde{\mathbf{H}}_R) \mathbf{E}_2 \right. \\ &\quad \left. + \tilde{\mathbf{H}}_R \mathbf{E}_3 + \tilde{\mathbf{H}}_I \mathbf{E}_4 \right). \end{aligned} \quad (22)$$

Let $\mathbf{H}_c = \mathbf{H}_R + i\mathbf{H}_I$, $\mathbf{G} = \mathbf{G}_R + i\mathbf{G}_I$, so we can denote $\tilde{\mathbf{H}}_R$ and $\tilde{\mathbf{H}}_I$ respectively as follows

$$\begin{aligned} \tilde{\mathbf{H}}_R &= (\mathbf{I}_T \otimes \mathbf{H}_R) \mathbf{G}_R - (\mathbf{I}_T \otimes \mathbf{H}_I) \mathbf{G}_I \\ \tilde{\mathbf{H}}_I &= (\mathbf{I}_T \otimes \mathbf{H}_R) \mathbf{G}_I + (\mathbf{I}_T \otimes \mathbf{H}_I) \mathbf{G}_R. \end{aligned} \quad (23)$$

Define F_1, F_2, \dots, F_6 as follows

$$\begin{aligned} \mathbf{F}_1 &= \mathbf{G}_R \mathbf{E}_1 \mathbf{G}_R^T + \mathbf{M}_I \mathbf{E}_1 \mathbf{G}_I^T + \mathbf{G}_I \mathbf{E}_2 \mathbf{G}_R^T - \mathbf{G}_R \mathbf{E}_2 \mathbf{G}_I^T \\ \mathbf{F}_2 &= \mathbf{G}_R \mathbf{E}_1 \mathbf{G}_I^T - \mathbf{M}_I \mathbf{E}_1 \mathbf{G}_R^T + \mathbf{G}_R \mathbf{E}_2 \mathbf{G}_R^T + \mathbf{G}_I \mathbf{E}_2 \mathbf{G}_I^T \\ \mathbf{F}_3 &= \mathbf{G}_I \mathbf{E}_1 \mathbf{G}_R^T - \mathbf{M}_R \mathbf{E}_1 \mathbf{G}_I^T - \mathbf{G}_I \mathbf{E}_2 \mathbf{G}_I^T - \mathbf{G}_R \mathbf{E}_2 \mathbf{G}_R^T \\ \mathbf{F}_4 &= \mathbf{G}_I \mathbf{E}_1 \mathbf{G}_I^T + \mathbf{M}_R \mathbf{E}_1 \mathbf{G}_R^T - \mathbf{G}_R \mathbf{E}_2 \mathbf{G}_I^T + \mathbf{G}_I \mathbf{E}_2 \mathbf{G}_R^T \\ \mathbf{F}_5 &= \mathbf{G}_R \mathbf{E}_3 + \mathbf{G}_I \mathbf{E}_4 \\ \mathbf{F}_6 &= \mathbf{G}_R \mathbf{E}_4 - \mathbf{G}_I \mathbf{E}_3, \end{aligned} \quad (24)$$

and let $\mathbf{Q}_i = \mathbf{F}_i(1:2, 1:2) + \mathbf{F}_i(3:4, 3:4) + \dots + \mathbf{F}_i(2n^2-1:2n^2, 2n^2-1:2n^2)$, putting (24) and (23) into (22), we have

$$\begin{aligned} \text{tr}(\mathbf{H}^T \mathbf{H} \mathbf{D} - \mathbf{H} \mathbf{C}) &= \text{tr} \left(\mathbf{H}_R^T \mathbf{H}_R \mathbf{Q}_1 + \mathbf{H}_I^T \mathbf{H}_I \mathbf{Q}_4 \right. \\ &\quad \left. + \mathbf{H}_R^T \mathbf{H}_I \mathbf{Q}_2 + \mathbf{H}_I^T \mathbf{H}_R \mathbf{Q}_2^T \right. \\ &\quad \left. + \mathbf{H}_R \mathbf{Q}_5 + \mathbf{H}_I \mathbf{Q}_6 \right). \end{aligned} \quad (25)$$

Put (25) into (20)

$$\begin{aligned} \mathcal{F} &= \text{const} + \frac{1}{2\sigma_e^2} \|\mathbf{H}_c - \hat{\mathbf{H}}_c^{(0)}\|_F^2 \\ &\quad + \frac{1}{2\sigma_n^2} \text{tr} \left(\mathbf{H}_R^T \mathbf{H}_R \mathbf{Q}_1 + \mathbf{H}_I^T \mathbf{H}_I \mathbf{Q}_4 + \mathbf{H}_R^T \mathbf{H}_I \mathbf{Q}_2 \right. \\ &\quad \left. + \mathbf{H}_I^T \mathbf{H}_R \mathbf{Q}_2^T + \mathbf{H}_R \mathbf{Q}_5 + \mathbf{H}_I \mathbf{Q}_6 \right) \\ &= \text{const} + \frac{1}{2\sigma_n^2} \left(\mathbf{H}_R^T \mathbf{H}_R \mathbf{P}_1 + \mathbf{H}_R^T \mathbf{H}_I \mathbf{P}_2 \right. \\ &\quad \left. + \mathbf{H}_I^T \mathbf{H}_R \mathbf{P}_2^T + \mathbf{H}_I^T \mathbf{H}_I \mathbf{P}_4 + \mathbf{H}_R \mathbf{P}_5 + \mathbf{H}_I \mathbf{P}_6 \right). \end{aligned} \quad (26)$$

where $\mathbf{P}_1 = \mathbf{Q}_1 + \frac{\sigma_n^2}{\sigma_e^2} \mathbf{I}$, $\mathbf{P}_2 = \mathbf{Q}_2$, $\mathbf{P}_4 = \mathbf{Q}_4 + \frac{\sigma_n^2}{\sigma_e^2} \mathbf{I}$, $\mathbf{P}_5^T = \mathbf{Q}_5^T - 2 \frac{\sigma_n^2}{\sigma_e^2} \hat{\mathbf{H}}_R^{(0)}$, $\mathbf{P}_6^T = \mathbf{Q}_6^T - 2 \frac{\sigma_n^2}{\sigma_e^2} \hat{\mathbf{H}}_I^{(0)}$. Arrange $\frac{\partial \mathcal{F}}{\partial \mathbf{H}_R} = 0$, $\frac{\partial \mathcal{F}}{\partial \mathbf{H}_I} = 0$, and notice \mathbf{P}_1 and \mathbf{P}_4 are symmetric and $\mathbf{P}_1 = \mathbf{P}_4$, we have

$$\begin{cases} \mathbf{H}_R \mathbf{P}_1 + \mathbf{H}_I \mathbf{P}_2 + \mathbf{P}_5^T / 2 = 0 \\ \mathbf{H}_I \mathbf{P}_1 + \mathbf{H}_R \mathbf{P}_2^T + \mathbf{P}_6^T / 2 = 0, \end{cases} \quad (27)$$

whose solution is

$$\begin{cases} \hat{\mathbf{H}}_R = (\mathbf{P}_6^T \mathbf{P}_1^{-1} - \mathbf{P}_5^T \mathbf{P}_2^{-1}) (\mathbf{P}_1 \mathbf{P}_2^{-1} - \mathbf{P}_2^T \mathbf{P}_1^{-1})^{-1} / 2 \\ \hat{\mathbf{H}}_I = (\mathbf{P}_6^T \mathbf{P}_2^{-T} - \mathbf{P}_5^T \mathbf{P}_1^{-1}) (\mathbf{P}_2 \mathbf{P}_1^{-1} - \mathbf{P}_1 \mathbf{P}_2^{-T})^{-1} / 2. \end{cases} \quad (28)$$