Transactions Letters

A Computationally Efficient Joint Channel Estimation and Data Detection for SIMO Systems

The-Hanh Pham, A. Nallanathan, *Senior Member, IEEE*, and Ying-Chang Liang, *Senior Member, IEEE*

***Abstract*—In this letter, we consider the problem of joint channel estimation and data detection for single-input multi-output (SIMO) systems under fast-fading environment with unknown spatially correlated noise. We present a computationally efficient iterative receiver based on the expectation conditional maximization (ECM) algorithm. Bit error rate (BER) performance of the proposed iterative receiver is simulated and compared with that of the maximum likelihood (ML) receiver with perfect channel state information (CSI). Simulation results show that the proposed iterative receiver achieves nearly the performance of ML receiver with perfect CSI and requires only a few iterations.**

***Index Terms*—Channel estimation, data detection, Expectation Maximization (EM) algorithm, Expectation Conditional Maximization (ECM) algorithm, single input multiple output (SIMO) system.**

I. Introduction

THE single-input multi-output (SIMO) channel model has been widely used in data transmission systems with antenna diversity. In practice, the channel state information (CSI) must be estimated so that coherent data detection can be carried out. One general approach to obtain the CSI is to use pilot symbols. Another approach is, besides a portion of pilot symbols, to iteratively estimate the channel and detect the data where detected data is used to refine the CSI estimate. For example, in [1, 2], authors use the expectation maximization (EM) [3] algorithm to follow this approach. All these methods assume that the additive noise is white in both spatial and temporal domains. However, in practice, the noise could be spatially correlated when co-channel interference (CCI) exists.

Recently, the EM algorithm has been used in [4] to estimate the channel coefficients and noise covariance matrix iteratively in a SIMO system under the framework of generalized multivariate analysis [5], considering the unknown symbols as the *missing* data. This method is derived for the case that the channel coefficients are static for some symbol periods (quasi-static fading). In this method, the prior probability density function (pdf) of the channel is not taken into account.

In [6], the author considers a SIMO system under quasi-static fading channels with spatially correlated noise. In that work, to tackle the complex problem of estimating more than one parameter, the author takes the advantage of the space-alternating generalized expectation-maximization (SAGE) algorithm [7], in which the parameters (unknown symbols and noise covariance matrix) are divided into two groups and are updated sequentially. In the process of updating each group, the channel coefficients are considered as missing data as opposed to [4]. Basically, the SAGE algorithm is designed to improve the convergence rate and reduce the complexity of the maximization step of the EM algorithm by dividing the parameters to be estimated into *smaller* groups and associating each smaller group with a *smaller* hidden data space [7]. However, in [6], the author only takes the advantage of the idea of the SAGE algorithm but does not reduce the hidden data spaces for each parameter. Hence, the hidden data space for each smaller group is still *big*; thus, the computational complexity is still very high.

In this letter, we study the problem of joint channel estimation and data detection for SIMO systems under the cases that the CSI is time varying and the additive noises are spatially correlated. The noise covariance matrix is also unknown. We treat the channel coefficients as the missing data and propose an expectation conditional maximization (ECM) [8] based algorithm for maximum likelihood (ML) estimation of both the unknown symbols and the spatial noise covariance matrix in a SIMO system under fast fading channels. The ECM algorithm is designed to directly overcome the complicated M-step of the EM algorithm while having the advantage of sequential updating approach as the SAGE algorithm. We also sketch the approach that considers signal as the missing data [4] for the fast fading channels using the pdf of channel. This is Bayesian in a sense that it takes into account the knowledge of the prior pdf of the channel coefficients. We show that the proposed algorithm maintains the same performance as the SAGE method and the approach extended from [4]. However, the proposed algorithm saves around 50% of computational resources compared with the SAGE algorithm and 9% of

Manuscript received November 13 2006; revised June 11 2007, December 18 2007, and June 1 2008; accepted July 14 2008. The associate editor coordinating the review of this manuscript and approving it for publication was A. Stefanov.

T.-H. Pham is with the Department of Electrical & Computer Engineering, National University of Singapore, Singapore 119260 (e-mail: elepth@nus.edu.sg).

A. Nallanathan is with the Division of Engineering, King's College London, UK (e-mail: nallanathan@ieee.org).

Y.-C. Liang is with the Institute for Infocomm Research (I^{2}R), 1 Fusionopolis Way, #21-01 Connexis, Singapore 138632 (e-mail: ycliang@i2r.a-star.edu.sg).

Digital Object Identifier 10.1109/T-WC.2008.060934

computational resources compared with the approach extended from [4].

The rest of the letter is organized as follows. Section II presents the system model. Section III describes the proposed iterative receiver. Computational complexity is compared in Section IV. Simulation results are presented in Section V. Finally, conclusions are drawn in Section VI.

Notations: The capital bold italic letters denote matrices and the small bold italic letters denote row/column vectors; $\delta(t)$ is the Kronecker delta function; $\otimes$ is the Kronecker product; $\boldsymbol{I}_T$ is the identity matrix of size T; $\mathbf{0}_{n\times m}$ is the $n \times m$ zero matrix; transpose, Hermitian transpose, inverse and determinant of matrix $\boldsymbol{A}$ are denoted by $\boldsymbol{A}^{\mathcal{T}}$, $\boldsymbol{A}^{\mathcal{H}}$, $\boldsymbol{A}^{-1}$ and $|\boldsymbol{A}|$, respectively; tr$[\boldsymbol{A}]$ stands for the trace operation; diag$(\boldsymbol{a})$ denotes the diagonal matrix with the diagonal element constructed from vector $\boldsymbol{a}$; diag$_M(a)$ denotes the diagonal matrix of size M where all elements of the main diagonal is a; $\boldsymbol{A}(n:m, p:q)$ is a matrix formed by rows n to m and columns p to q of $\boldsymbol{A}$; sign$[\cdot]$ denotes the sign function; $\Re(\cdot)$ is the real part of a complex vector/matrix; a^* is the conjugate of a complex number a.

II. System Model

Consider a SIMO system comprising a single transmit and M receive antennas operating in a Rayleigh fading environment. The additive noise is spatially correlated among M receive antennas but are independent from one symbol to another. The transmitted symbol at time t, s_t, belongs to an M-ary constant modulus constellation $\mathcal{C} = \{c_1, c_2, \cdots, c_{|\mathcal{C}|}\}$ with $|c_u|^2 = 1$, $u = 1, 2, \cdots, |\mathcal{C}|$. Let $\boldsymbol{s} = [s_1\ s_2\ \cdots\ s_T] \in \mathcal{C}^{1\times T}$ be a block of transmitted symbols. To facilitate the derivations that follow, we also define $\underline{\boldsymbol{S}} = \text{diag}(\boldsymbol{s})$ and $\boldsymbol{S}_t = \text{diag}_M(s_t)$ of size $M \times M$. Let the $T \times 1$ vector $\boldsymbol{h}^{(m)} = \left[h_1^{(m)}\ \cdots\ h_t^{(m)} \cdots\ h_T^{(m)}\right]^{\mathcal{T}} \in \mathbb{C}^{T\times 1}$ be the channel coefficients between the transmit antenna and the m^{th} receive antenna, where $m = 1, 2, \cdots, M$ and the subscript t denotes time index. We assume that M channels from the transmit antenna to M receive antennas are independent. We also define the $M \times 1$ vector of channel coefficients at time t from the transmit antenna to all M receive antennas as $\boldsymbol{h}_t = \left[h_t^{(1)}\ h_t^{(2)}\ \cdots\ h_t^{(M)}\right]^{\mathcal{T}} \in \mathbb{C}^{M\times 1}$.

At any time t, the received signal at the m^{th} receive antenna, $y_t^{(m)}$, $m = 1, 2, \cdots, M$, can be expressed as

$$y_t^{(m)} = h_t^{(m)} s_t + n_t^{(m)}, \tag{1}$$

where $n_t^{(m)}$ is the noise at the m^{th} receive antenna at time t. If we collect M received signals at time t from all M receive antennas to form a vector, $\boldsymbol{y}_t = \left[y_t^{(1)}\ y_t^{(2)} \cdots\ y_t^{(M)}\right]^{\mathcal{T}} \in \mathbb{C}^{M\times 1}$, it can be written as

$$\boldsymbol{y}_t = \boldsymbol{h}_t s_t + \boldsymbol{n}_t,\ t = 1, 2, \cdots, T, \tag{2}$$

where $\boldsymbol{n}_t = [n_t^{(1)}\ n_t^{(2)}\ \cdots\ n_t^{(M)}]^{\mathcal{T}}$. The noise is assumed to be temporally white, i.e., $E\left\{\boldsymbol{n}_{t_1}\boldsymbol{n}_{t_2}^{\mathcal{H}}\right\} = \boldsymbol{\Sigma}\delta(t_1 - t_2)$ where $\boldsymbol{n}_{t_1}$ and $\boldsymbol{n}_{t_2}$ are the noise vectors at time t_1 and t_2, respectively. If we collect T received signal vectors to form a matrix $\boldsymbol{Y} = \left[\boldsymbol{y}_1\ \boldsymbol{y}_2\ \cdots \boldsymbol{y}_T\right] \in \mathbb{C}^{M\times T}$, it can be expressed as

$$\boldsymbol{Y} = \boldsymbol{H}\underline{\boldsymbol{S}} + \boldsymbol{N} \tag{3}$$

where $\boldsymbol{H} = \left[\boldsymbol{h}_1\ \boldsymbol{h}_2\ \cdots \boldsymbol{h}_T\right] \in \mathbb{C}^{M\times T}$, $\boldsymbol{N} = [\boldsymbol{n}_1\ \boldsymbol{n}_2\ \cdots\ \boldsymbol{n}_T] \in \mathbb{C}^{M\times T}$ is an additive noise matrix.

We assume that the vector $\boldsymbol{h}^{(m)}$ can be modeled as a zero-mean complex Gaussian random vector with covariance matrix $\boldsymbol{R} = E\{\boldsymbol{h}^{(m)}(\boldsymbol{h}^{(m)})^{\mathcal{H}}\}$ of size $T \times T$. Here, we adopt the Jakes' model [9], i.e., the $(i,j)^{th}$ element of $\boldsymbol{R}$ is given by

$$\boldsymbol{R}(i,j) = J_0(2\pi f_d T_s(i-j)) \equiv r_{i-j}, \tag{4}$$

where $J_0(\cdot)$ is the first kind Bessel function of zero order, f_d is the maximum Doppler shift and T_s is the symbol period. Note that $r_{i-j} = r_{j-i}$, thus $\boldsymbol{R}$ is a symmetric Toeplitz matrix with the first row of $\boldsymbol{r} = [r_0\ r_1\ \cdots\ r_{T-1}]$. We assume that the value of $f_d T_s$ is known at the receiver.

III. Proposed Iterative Receiver

In this section, we present the joint estimation of channel coefficients $\boldsymbol{h}_t$, $t = 1, 2, \cdots, T$, noise covariance $\boldsymbol{\Sigma}$ and data symbols $\boldsymbol{s}$ based on the received signal matrix $\boldsymbol{Y}$ and some pilot symbols to kick start the algorithm.

We define $\boldsymbol{y} = [\boldsymbol{y}_1^{\mathcal{T}}\ \boldsymbol{y}_2^{\mathcal{T}}\ \cdots\ \boldsymbol{y}_T^{\mathcal{T}}]^{\mathcal{T}} \in \mathbb{C}^{MT\times 1}$, $\boldsymbol{S} = \text{diag}(\boldsymbol{S}_1, \boldsymbol{S}_2, \cdots, \boldsymbol{S}_T)$, $\boldsymbol{h} = [\boldsymbol{h}_1^{\mathcal{T}}\ \boldsymbol{h}_2^{\mathcal{T}}\ \cdots\ \boldsymbol{h}_T^{\mathcal{T}}]^{\mathcal{T}} \in \mathbb{C}^{MT\times 1}$ and $\boldsymbol{n} = [\boldsymbol{n}_1^{\mathcal{T}}\ \boldsymbol{n}_2^{\mathcal{T}}\ \cdots\ \boldsymbol{n}_T^{\mathcal{T}}]^{\mathcal{T}} \in \mathbb{C}^{MT\times 1}$. Without loss of information, (3) can be written as

$$\boldsymbol{y} = \boldsymbol{S}\boldsymbol{h} + \boldsymbol{n}. \tag{5}$$

The covariance matrix of $\boldsymbol{h}$, $\boldsymbol{K}_{\boldsymbol{h}} = E\{\boldsymbol{h}\boldsymbol{h}^{\mathcal{H}}\}$, is a $MT \times MT$ symmetric Toeplitz matrix with the first row being $[r_0\ \mathbf{0}_{1\times(M-1)}\ r_1\ \mathbf{0}_{1\times(M-1)}\ \cdots\ r_{T-1}\ \mathbf{0}_{1\times(M-1)}]$ and the covariance matrix of $\boldsymbol{n}$ is $\boldsymbol{K}_{\boldsymbol{n}} = E\{\boldsymbol{n}\boldsymbol{n}^{\mathcal{H}}\} = \boldsymbol{I}_T \otimes \boldsymbol{\Sigma}$.

We denote $\boldsymbol{\theta} \triangleq (\boldsymbol{S}, \boldsymbol{\Sigma})$ as the parameters to be estimated and call $\hat{\boldsymbol{\theta}}^{[k]} = (\hat{\boldsymbol{S}}^{[k]}, \hat{\boldsymbol{\Sigma}}^{[k]})$ be the estimate of $\boldsymbol{\theta}$ after the k^{th} iteration of the proposed algorithm. In EM terminology, $\boldsymbol{y}$ is the incomplete data, and we define $\boldsymbol{X} = (\boldsymbol{y}, \boldsymbol{h})$ as the complete data for the parameter we want to estimate. The proposed ECM based algorithm contains E-step and CM-step as follows.

A. E-step

In this step, we compute

$$Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}^{[k]}) = E\{\log f(\boldsymbol{X}|\boldsymbol{S}, \boldsymbol{\Sigma})|\boldsymbol{y}, \hat{\boldsymbol{S}}^{[k]}, \hat{\boldsymbol{\Sigma}}^{[k]}\}. \tag{6}$$

The pdf of $\boldsymbol{X}$ as a function of $\boldsymbol{S}$ and $\boldsymbol{\Sigma}$ is

$$f(\boldsymbol{X}|\boldsymbol{S}, \boldsymbol{\Sigma}) = f(\boldsymbol{y}|\boldsymbol{h}, \boldsymbol{S}, \boldsymbol{\Sigma}) f(\boldsymbol{h}|\boldsymbol{S}, \boldsymbol{\Sigma}), \tag{7}$$

in which

$$f(\boldsymbol{y}|\boldsymbol{h}, \boldsymbol{S}, \boldsymbol{\Sigma}) = \frac{1}{|\pi \boldsymbol{K}_{\boldsymbol{n}}|} \times \exp\{-(\boldsymbol{y} - \boldsymbol{S}\boldsymbol{h})^{\mathcal{H}} \boldsymbol{K}_{\boldsymbol{n}}^{-1} (\boldsymbol{y} - \boldsymbol{S}\boldsymbol{h})\}, \tag{8}$$

and $f(\boldsymbol{h}|\boldsymbol{S}, \boldsymbol{\Sigma}) = f(\boldsymbol{h}) = \frac{1}{|\pi \boldsymbol{K}_{\boldsymbol{h}}|} \exp\left\{-\boldsymbol{h}^{\mathcal{H}} \boldsymbol{K}_{\boldsymbol{h}}^{-1} \boldsymbol{h}\right\}$ because of the independence among $\boldsymbol{h}$, $\boldsymbol{S}$ and $\boldsymbol{\Sigma}$.

Substituting (7) into (6), after dropping some constant terms that do not relate to $\boldsymbol{\theta}$, we have

$$\begin{aligned} Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}^{[k]}) &= -T\log|\boldsymbol{\Sigma}| \\ &\quad - E\{(\boldsymbol{y}-\boldsymbol{S}\boldsymbol{h})^{\mathcal{H}}\boldsymbol{K}_{\boldsymbol{n}}^{-1}(\boldsymbol{y}-\boldsymbol{S}\boldsymbol{h})|\boldsymbol{y},\hat{\boldsymbol{S}}^{[k]},\hat{\boldsymbol{\Sigma}}^{[k]}\} \\ &= -T\log|\boldsymbol{\Sigma}| - \mathrm{tr}\big[\boldsymbol{K}_{\boldsymbol{n}}^{-1}\boldsymbol{S}\hat{\boldsymbol{K}}_{\boldsymbol{h}}^{[k]}\boldsymbol{S}^{\mathcal{H}}\big] \\ &\quad - \big(\boldsymbol{y}-\boldsymbol{S}\hat{\boldsymbol{h}}^{[k]}\big)^{\mathcal{H}}\boldsymbol{K}_{\boldsymbol{n}}^{-1}\big(\boldsymbol{y}-\boldsymbol{S}\hat{\boldsymbol{h}}^{[k]}\big) \\ &= -T\log|\boldsymbol{\Sigma}| - \mathrm{tr}\big[\boldsymbol{S}^{\mathcal{H}}\boldsymbol{K}_{\boldsymbol{n}}^{-1}\boldsymbol{S}\hat{\boldsymbol{K}}_{\boldsymbol{h}}^{[k]}\big] \\ &\quad - \big(\boldsymbol{y}-\boldsymbol{S}\hat{\boldsymbol{h}}^{[k]}\big)^{\mathcal{H}}\boldsymbol{K}_{\boldsymbol{n}}^{-1}\big(\boldsymbol{y}-\boldsymbol{S}\hat{\boldsymbol{h}}^{[k]}\big) \\ &= -T\log|\boldsymbol{\Sigma}| - \mathrm{tr}\big[\boldsymbol{K}_{\boldsymbol{n}}^{-1}\hat{\boldsymbol{K}}_{\boldsymbol{h}}^{[k]}\big] \\ &\quad - \big(\boldsymbol{y}-\boldsymbol{S}\hat{\boldsymbol{h}}^{[k]}\big)^{\mathcal{H}}\boldsymbol{K}_{\boldsymbol{n}}^{-1}\big(\boldsymbol{y}-\boldsymbol{S}\hat{\boldsymbol{h}}^{[k]}\big), \end{aligned} \tag{9}$$

where

$$\begin{aligned} \hat{\boldsymbol{h}}^{[k]} &\triangleq E\{\boldsymbol{h}|\boldsymbol{y},\hat{\boldsymbol{S}}^{[k]},\hat{\boldsymbol{\Sigma}}^{[k]}\} \\ &= \boldsymbol{K}_{\boldsymbol{h}}\big(\hat{\boldsymbol{S}}^{[k]}\big)^{\mathcal{H}}\big(\hat{\boldsymbol{S}}^{[k]}\boldsymbol{K}_{\boldsymbol{h}}(\hat{\boldsymbol{S}}^{[k]})^{\mathcal{H}}+\hat{\boldsymbol{K}}_{\boldsymbol{n}}^{[k]}\big)^{-1}\boldsymbol{y} && \text{(10a)} \\ &= \big(\boldsymbol{K}_{\boldsymbol{h}}^{-1}+(\hat{\boldsymbol{S}}^{[k]})^{\mathcal{H}}\big(\hat{\boldsymbol{K}}_{\boldsymbol{n}}^{[k]}\big)^{-1}\hat{\boldsymbol{S}}^{[k]}\big)^{-1}\big(\hat{\boldsymbol{S}}^{[k]}\big)^{\mathcal{H}}\big(\hat{\boldsymbol{K}}_{\boldsymbol{n}}^{[k]}\big)^{-1}\boldsymbol{y} && \text{(10b)} \\ &= \big(\boldsymbol{K}_{\boldsymbol{h}}^{-1}+\big(\hat{\boldsymbol{K}}_{\boldsymbol{n}}^{[k]}\big)^{-1}\big)^{-1}\big(\hat{\boldsymbol{S}}^{[k]}\big)^{\mathcal{H}}\big(\hat{\boldsymbol{K}}_{\boldsymbol{n}}^{[k]}\big)^{-1}\boldsymbol{y} && \text{(10c)} \\ &= \big(\boldsymbol{K}_{\boldsymbol{h}}-\boldsymbol{K}_{\boldsymbol{h}}\big(\boldsymbol{K}_{\boldsymbol{h}}+\hat{\boldsymbol{K}}_{\boldsymbol{n}}^{[k]}\big)^{-1}\boldsymbol{K}_{\boldsymbol{h}}\big)\big(\hat{\boldsymbol{S}}^{[k]}\big)^{\mathcal{H}}\big(\hat{\boldsymbol{K}}_{\boldsymbol{n}}^{[k]}\big)^{-1}\boldsymbol{y}, && \text{(10d)} \end{aligned}$$

and

$$\begin{aligned} \hat{\boldsymbol{K}}_{\boldsymbol{h}}^{[k]} &\triangleq E\{\big(\boldsymbol{h}-\hat{\boldsymbol{h}}^{[k]}\big)\big(\boldsymbol{h}-\hat{\boldsymbol{h}}^{[k]}\big)^{\mathcal{H}}|\boldsymbol{y},\hat{\boldsymbol{S}}^{[k]},\hat{\boldsymbol{\Sigma}}^{[k]}\} \\ &= \boldsymbol{K}_{\boldsymbol{h}}-\boldsymbol{K}_{\boldsymbol{h}}\big(\hat{\boldsymbol{S}}^{[k]}\big)^{\mathcal{H}}\big(\hat{\boldsymbol{S}}^{[k]}\boldsymbol{K}_{\boldsymbol{h}}(\hat{\boldsymbol{S}}^{[k]})^{\mathcal{H}}+\hat{\boldsymbol{K}}_{\boldsymbol{n}}^{[k]}\big)^{-1}\hat{\boldsymbol{S}}^{[k]}\boldsymbol{K}_{\boldsymbol{h}} \\ &= \big(\boldsymbol{K}_{\boldsymbol{h}}^{-1}+(\hat{\boldsymbol{S}}^{[k]})^{\mathcal{H}}\big(\hat{\boldsymbol{K}}_{\boldsymbol{n}}^{[k]}\big)^{-1}\hat{\boldsymbol{S}}^{[k]}\big)^{-1} \\ &= \boldsymbol{K}_{\boldsymbol{h}}-\boldsymbol{K}_{\boldsymbol{h}}\big(\boldsymbol{K}_{\boldsymbol{h}}+\hat{\boldsymbol{K}}_{\boldsymbol{n}}^{[k]}\big)^{-1}\boldsymbol{K}_{\boldsymbol{h}}, \end{aligned} \tag{11}$$

and $\hat{\boldsymbol{K}}_{\boldsymbol{n}}^{[k]} = \boldsymbol{I}_T \otimes \hat{\boldsymbol{\Sigma}}^{[k]}$. To have (10b) from (10a), we resort to the Matrix Inversion Lemma. The modulus constellation property is used to simplify (10b) to (10c). The reason why we expand (10c) to (10d) is that the matrix $\boldsymbol{K}_{\boldsymbol{h}}$ is ill-conditioned so its inversion suffers from the numerical error which produces unreliable results.

B. CM-step

In this step, the updated value of $\boldsymbol{\theta}$, $\hat{\boldsymbol{\theta}}^{[k+1]}$, is determined as

$$\hat{\boldsymbol{\theta}}^{[k+1]} = \arg\max_{\boldsymbol{\theta}}\ Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}^{[k]}). \tag{12}$$

In contrast to the conventional EM algorithm, where the updating process of parameters are taken place simultaneously, the ECM algorithm maximizes the function $Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}^{[k]})$ in two steps. In the first step, $Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}^{[k]})$ is maximized with respect to (w.r.t.) one of $(\boldsymbol{S},\boldsymbol{\Sigma})$ while the other is kept at its most updated value. We denote $\hat{\boldsymbol{\theta}}^{[k+c/2]}$ be the estimate of $\boldsymbol{\theta}$ at c^{th} step of k^{th} iteration of the ECM algorithm, $c = 1, 2$. Then, the CM-step of the ECM algorithm consists of two following steps.

1) Step 1: In this step, we have $\hat{\boldsymbol{\theta}}^{[k+1/2]} = (\hat{\boldsymbol{S}}^{[k+1]}, \hat{\boldsymbol{\Sigma}}^{[k]})$ where

$$\begin{aligned} \hat{\boldsymbol{S}}^{[k+1]} &= \arg\max_{\boldsymbol{S}}\ Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}^{[k]})\big|_{\boldsymbol{\Sigma}=\hat{\boldsymbol{\Sigma}}^{[k]}} \\ &= \arg\min_{\boldsymbol{S}}\ \sum_{t=1}^{T}(\boldsymbol{y}_t-\hat{\boldsymbol{h}}_t^{[k]}s_t)^{\mathcal{H}}(\hat{\boldsymbol{\Sigma}}^{[k]})^{-1}(\boldsymbol{y}_t-\hat{\boldsymbol{h}}_t^{[k]}s_t). \end{aligned} \tag{13}$$

We can see that the process of updating $\hat{\boldsymbol{S}}^{[k+1]}$ can be transformed to that of $\hat{\boldsymbol{s}}^{[k+1]}$. More explicitly, the updated value of $\hat{s}_t^{[k+1]}$, $t = 1, 2, \cdots, T$, is

$$\hat{s}_t^{[k+1]} = \arg\max_{s_t}\ \Re\big(\boldsymbol{y}_t^{\mathcal{H}}(\hat{\boldsymbol{\Sigma}}^{[k]})^{-1}\hat{\boldsymbol{h}}_t^{[k]}s_t\big). \tag{14}$$

If we use BPSK symbols, the decision on $\hat{s}_t^{[k+1]}$ reduces to

$$\hat{s}_t^{[k+1]} = \mathrm{sign}\Big[\Re\big(\boldsymbol{y}_t^{\mathcal{H}}(\hat{\boldsymbol{\Sigma}}^{[k]})^{-1}\hat{\boldsymbol{h}}_t^{[k]}\big)\Big]. \tag{15}$$

2) Step 2: In this step, we have $\hat{\boldsymbol{\theta}}^{[k+1]} = (\hat{\boldsymbol{S}}^{[k+1]}, \hat{\boldsymbol{\Sigma}}^{[k+1]})$ where

$$\begin{aligned} \hat{\boldsymbol{\Sigma}}^{[k+1]} &= \arg\max_{\boldsymbol{\Sigma}}\ Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}^{[k]})\big|_{\boldsymbol{S}=\hat{\boldsymbol{S}}^{[k+1]}} \\ &= \arg\max_{\boldsymbol{\Sigma}}\ \Big\{-T\log|\boldsymbol{\Sigma}| - \mathrm{tr}\big[\boldsymbol{K}_{\boldsymbol{n}}^{-1}\hat{\boldsymbol{K}}_{\boldsymbol{h}}^{[k]}\big] \\ &\quad - \big(\boldsymbol{y}-\hat{\boldsymbol{S}}^{[k+1]}\hat{\boldsymbol{h}}^{[k]}\big)\boldsymbol{K}_{\boldsymbol{n}}^{-1}\big(\boldsymbol{y}-\hat{\boldsymbol{S}}^{[k+1]}\hat{\boldsymbol{h}}^{[k]}\big)\Big\}. \end{aligned} \tag{16}$$

If we define $\big[\hat{\boldsymbol{K}}_{\boldsymbol{h}}^{[k]}\big]_t = \hat{\boldsymbol{K}}_{\boldsymbol{h}}^{[k]}\big((t-1)M+1:tM,(t-1)M+1:tM\big)$ with $t = 1, 2, \cdots, T$ then (16) can be written as

$$\begin{aligned} \hat{\boldsymbol{\Sigma}}^{[k+1]} &= \arg\max_{\boldsymbol{\Sigma}}\ \Big\{-T\log|\boldsymbol{\Sigma}| - \mathrm{tr}\big[\boldsymbol{\Sigma}^{-1}\big(\sum_{t=1}^{T}\big[\hat{\boldsymbol{K}}_{\boldsymbol{h}}^{[k]}\big]_t\big)\big] \\ &\quad - \mathrm{tr}\big[\big(\boldsymbol{Y}-\hat{\boldsymbol{H}}^{[k]}\hat{\underline{\boldsymbol{S}}}^{[k+1]}\big)^{\mathcal{H}}\boldsymbol{\Sigma}^{-1}\big(\boldsymbol{Y}-\hat{\boldsymbol{H}}^{[k]}\hat{\underline{\boldsymbol{S}}}^{[k+1]}\big)\big]\Big\}, \end{aligned} \tag{17}$$

where $\hat{\boldsymbol{H}}^{[k]} = \big[\hat{\boldsymbol{h}}_1^{[k]}\ \hat{\boldsymbol{h}}_2^{[k]}\ \cdots\ \hat{\boldsymbol{h}}_T^{[k]}\big]$.

If we differentiate (17) w.r.t. $\boldsymbol{\Sigma}$ and equating the result to zero, we can obtain the updated value of $\boldsymbol{\Sigma}$ as

$$\begin{aligned} \hat{\boldsymbol{\Sigma}}^{[k+1]} &= \frac{\sum_{t=1}^{T}[\hat{\boldsymbol{K}}_{\boldsymbol{h}}^{[k]}]_t}{T} \\ &\quad + \frac{(\boldsymbol{Y}-\hat{\boldsymbol{H}}^{[k]}\hat{\underline{\boldsymbol{S}}}^{[k+1]})(\boldsymbol{Y}-\hat{\boldsymbol{H}}^{[k]}\hat{\underline{\boldsymbol{S}}}^{[k+1]})^{\mathcal{H}}}{T}. \end{aligned} \tag{18}$$

IV. Computational Complexity

We now look at the computational complexity of the proposed algorithm and compare it with the approach using the SAGE algorithm. Since the transmitted and received signals as well as the channel matrix are complex, all processing is conducted in complex domain. Thus, throughout this section, multiplications, divisions, and additions refer to complex operations and are denoted by CMs, CDs, and CAs, respectively.

We consider the complexity of the k^{th} iteration of our proposed algorithm. The iteration consists of 4 steps and the complexity of each step is addressed as follows:

- **Step 1**: Calculation of $\hat{\boldsymbol{h}}^{[k]} = \big(\boldsymbol{K}_{\boldsymbol{h}} - \boldsymbol{K}_{\boldsymbol{h}}\big(\boldsymbol{K}_{\boldsymbol{h}} + \hat{\boldsymbol{K}}_{\boldsymbol{n}}^{[k]}\big)^{-1}\boldsymbol{K}_{\boldsymbol{h}}\big)\big(\hat{\boldsymbol{S}}^{[k]}\big)^{\mathcal{H}}\big(\hat{\boldsymbol{K}}_{\boldsymbol{n}}^{[k]}\big)^{-1}\boldsymbol{y}$
  In order to save computation cost, we calculate $\big(\boldsymbol{K}_{\boldsymbol{h}} - \boldsymbol{K}_{\boldsymbol{h}}\big(\boldsymbol{K}_{\boldsymbol{h}} + \hat{\boldsymbol{K}}_{\boldsymbol{n}}^{[k]}\big)^{-1}\boldsymbol{K}_{\boldsymbol{h}}\big)$ which cost us $3(MT)^3$ CMs, $(MT)^3$ CDs and $3(MT)^3 - 2(MT)^2 + MT$ CAs. The remaining calculation requires $2(MT)^3 + (MT)^2$ CMs

TABLE I

ALGORITHM AND COMPLEXITY OF THE APPROACH THAT CONSIDERS SIGNAL AS THE MISSING DATA

	Algorithm (k^{th} iteration)	Complexity
Step 1:	Calculate $\bar{s}_t^{[k]} \triangleq E\{s_t \vert \{\boldsymbol{y}_t\}_{t=1}^T, \{\hat{\boldsymbol{h}}_t^{[k]}\}_{t=1}^T, \hat{\boldsymbol{\Sigma}}^{[k]}\} = \sum_{c=1}^{\vert\mathcal{C}\vert} c_u \hat{\rho}_{u,t}^{[k]}$ where $\hat{\rho}_{u,t}^{[k]} = \frac{\exp\left\{2\Re\left\{\boldsymbol{y}_t^{\mathcal{H}}(\hat{\boldsymbol{\Sigma}}^{[k]})^{-1}\hat{\boldsymbol{h}}_t^{[k]}c_u\right\}\right\}}{\sum_{l=1}^{U}\exp\left\{2\Re\left\{\boldsymbol{y}_t^{\mathcal{H}}(\hat{\boldsymbol{\Sigma}}^{[k]})^{-1}\hat{\boldsymbol{h}}_t^{[k]}c_l\right\}\right\}}$ and form $\bar{\boldsymbol{S}}^{[k]} = \mathrm{diag}(\bar{\boldsymbol{S}}_1^{[k]}, \cdots, \bar{\boldsymbol{S}}_T^{[k]})$, $\bar{\boldsymbol{S}}_t^{[k]} = \mathrm{diag}_M(\bar{s}_t^{[k]})$. Calculate $\mathrm{cov}^{[k]}(s_t) \triangleq 1 - \vert\bar{s}_t^{[k]}\vert^2$.	$M^3 + (M^2 + M + 2\vert\mathcal{C}\vert)T$ CMs M^3 CDs $M^3 + (T-2)M^2 + (\vert\mathcal{C}\vert - 1)T + M$ CAs
Step 2:	Update $\boldsymbol{h}$ as $\hat{\boldsymbol{h}}^{[k+1]} = \left(\boldsymbol{K_h} - \boldsymbol{K_h}(\boldsymbol{K_h} + \hat{\boldsymbol{K}}_{\boldsymbol{n}}^{[k]})^{-1}\boldsymbol{K_h}\right)(\bar{\boldsymbol{S}}^{[k]})^{\mathcal{H}}\left(\hat{\boldsymbol{K}}_{\boldsymbol{n}}^{[k]}\right)^{-1}\boldsymbol{y}$	$5(MT)^3 + (MT)^2$ CMs $(MT)^3$ CDs $5(MT)^3 - 3(MT)^2$ CAs
Step 3:	Update $\boldsymbol{\Sigma}$ as $\hat{\boldsymbol{\Sigma}}^{[k+1]} = \frac{\left(\boldsymbol{Y} - \hat{\boldsymbol{H}}^{[k+1]}\bar{\underline{\boldsymbol{S}}}^{[k]}\right)\left(\boldsymbol{Y} - \hat{\boldsymbol{H}}^{[k+1]}\bar{\underline{\boldsymbol{S}}}^{[k]}\right)^{\mathcal{H}}}{T}$ $+\frac{1}{T}\sum_{t=1}^{T}\mathrm{cov}^{[k]}(s_t)\hat{\boldsymbol{h}}_t^{[k+1]}(\hat{\boldsymbol{h}}_t^{[k+1]})^{\mathcal{H}}$ where $\bar{\underline{\boldsymbol{S}}}^{[k]} = \mathrm{diag}(\bar{s}_t^{[k]}, \cdots, \bar{s}_T^{[k]})$	$MT^2 + 3TM^2$ CMs M^2 CDs $MT^2 + 2M^2T - 2M^2$ CAs

and $2(MT)^3 - (MT)^2 - MT$ CAs. Hence, we need $5(MT)^3 + (MT)^2$ CMs, $(MT)^3$ CDs and $5(MT)^3 - 3(MT)^2$ CAs.

- **Step 2**: Calculation of $\hat{\boldsymbol{K}}_{\boldsymbol{h}}^{[k]} = \boldsymbol{K_h} - \boldsymbol{K_h}(\boldsymbol{K_h} + \hat{\boldsymbol{K}}_{\boldsymbol{n}}^{[k]})^{-1}\boldsymbol{K_h}$
  This quantity is computed in the previous step. Therefore, we do not need any calculation here.
- **Step 3**: Updating signal: $\hat{s}_t^{[k+1]} = \arg\max_{s_t} \Re\left(\boldsymbol{y}_t^{\mathcal{H}}(\hat{\boldsymbol{\Sigma}}^{[k]})^{-1}\hat{\boldsymbol{h}}_t^{[k]}s_t\right)$ for $t = 1, 2, \cdots, T$.
  We calculate $(\hat{\boldsymbol{\Sigma}}^{[k]})^{-1}$ which costs M^3 CMs/CDs and $(M^3 - 2M^2 + M)$ CAs and it is used for all t's. For each value of t, the determination of $\boldsymbol{y}_t^{\mathcal{H}}(\hat{\boldsymbol{\Sigma}}^{[k]})^{-1}\hat{\boldsymbol{h}}_t^{[k]}$ requires $(M^2 + M)$ CMs and $(M^2 - 1)$ CAs. If we use BPSK modulation at the transmit antenna, the above update equation of $\hat{s}_t^{[k+1]}$ reduces to (15) which requires no more calculation. Therefore, this step requires $M^3 + (M^2 + M)T$ CMs, M^3 CDs and $M^3 + (T-2)M^2 + M - T$ CAs.
- **Step 4**: Updating $\boldsymbol{\Sigma}$: $\hat{\boldsymbol{\Sigma}}^{[k+1]} = \frac{\sum_{t=1}^{T}[\hat{\boldsymbol{K}}_{\boldsymbol{h}}^{[k]}]_t}{T} + \frac{(\boldsymbol{Y} - \hat{\boldsymbol{H}}^{[k]}\hat{\underline{\boldsymbol{S}}}_T^{[k+1]})(\boldsymbol{Y} - \hat{\boldsymbol{H}}^{[k]}\hat{\underline{\boldsymbol{S}}}_T^{[k+1]})^{\mathcal{H}}}{T}$
  The calculation of $(\boldsymbol{Y} - \hat{\boldsymbol{H}}^{[k]}\hat{\underline{\boldsymbol{S}}}^{[k+1]})(\boldsymbol{Y} - \hat{\boldsymbol{H}}^{[k]}\hat{\underline{\boldsymbol{S}}}^{[k+1]})^{\mathcal{H}}$ requires $(MT^2 + M^2T)$ CMs and $(MT^2 + M^2T - M^2)$ CAs. It requires $(M^2T - M^2)$ CAs to calculate $\sum_{t=1}^{T}[\hat{\boldsymbol{K}}_{\boldsymbol{h}}^{[k]}]_t$ and another M^2 CDs to complete the determination of $\hat{\boldsymbol{\Sigma}}^{[k+1]}$.

If the problem is solved by using the SAGE algorithm, then $\boldsymbol{\theta} = (\boldsymbol{S}, \boldsymbol{\Sigma})$ should be divided into two groups of $\boldsymbol{S}$ and $\boldsymbol{\Sigma}$ and they are updated in sequential manner. For each group, the admissible hidden-data space [7] is defined as $(\boldsymbol{y}, \boldsymbol{h})$. In updating $\boldsymbol{S}$, the channel estimation in Step 1 and updating signal in Step 3 are performed. The updating of $\boldsymbol{\Sigma}$ requires the re-calculation of channel estimation in Step 1, as well as Step 2 and Step 4. Therefore, in each iteration (i.e., updating both $\boldsymbol{S}$ and $\boldsymbol{\Sigma}$), the SAGE algorithm requires one more time of calculating channel estimation in Step 1 than our proposed algorithm. This makes our proposed ECM based algorithm less complex than the SAGE based algorithm, and it will be further shown by simulations presented in the next section.

V. SIMULATION RESULTS

We consider a SIMO system with $M = 2$ receive antennas. The signal is corrupted by additive complex noise with two kinds, the first one has the noise covariance matrix of $\boldsymbol{\Sigma} = \sigma^2\boldsymbol{I}_M$ (white noise) and the other has noise covariance matrix $\boldsymbol{\Sigma}$ whose $(m,n)^{th}$ element is [4]

$$\boldsymbol{\Sigma}(m,n) = \sigma^2.(0.9)^{|m-n|}.\exp\left[j\left(\frac{\pi}{2}\right)(m-n)\right] \qquad (19)$$

i.e., correlated noise. The fading coefficients between the transmitter and each receive antenna are generated according to the Jakes' model. The data is grouped into a block of $T = 20$ symbols which are drawn from BPSK constellation, preceded by a pilot symbol. The receiver operates on frames comprising $F = 5$ successive blocks and the adjacent pilots.

Relying on the relations among channel coefficients by the Jakes model, the Wiener theory in [10] interpolates $F + 1$ successive pilot symbols to get the channel coefficients initialization. For each frame, the initialization of the noise covariance matrix for the first block is determined by $\hat{\boldsymbol{\Sigma}}^{[0]} = \frac{1}{FT+F+1}\sum_{t=1}^{FT+F+1}\boldsymbol{y}_t\boldsymbol{y}_t^{\mathcal{H}}$. The obtained $\boldsymbol{\Sigma}$ when the algorithm converges for the first block is transferred to the second block as the initialization and so on.

In this section, we also present the performance of the approach that considers signal as the missing data. Due to the space limitation, we are not able to provide its detailed derivation. However, the summary of the approach and its complexity are given in the Table I.

In Fig. 1, we plot the BER as a function of SNR under white and correlated noise environments at fading rate $f_dT_s = 0.01$. In order to compare the performance of our algorithm, we also plot the BER of ML detection with perfect CSI and the noise covariance matrix $\boldsymbol{\Sigma}$, as well as the BER of the proposed algorithm when $\boldsymbol{\Sigma}$ is known (therefore, the step of updating $\boldsymbol{\Sigma}$ is not required). We can observe that the obtained performance with unknown $\boldsymbol{\Sigma}$ is near that of ML with perfect CSI and $\boldsymbol{\Sigma}$. The gaps between the proposed algorithm and the ML with perfect CSI are 1.5dB and 2dB for white and correlated noise environments, respectively. In this figure, we also include the BER result of the approach that takes signal as the missing

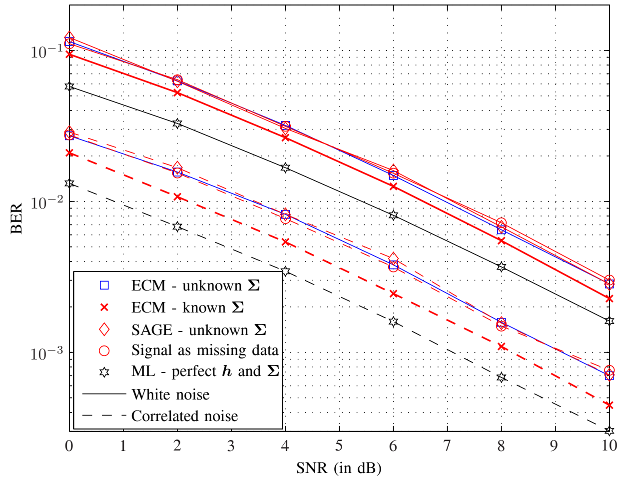

Fig. 1. BER vs. SNR in white and correlated noise environments.

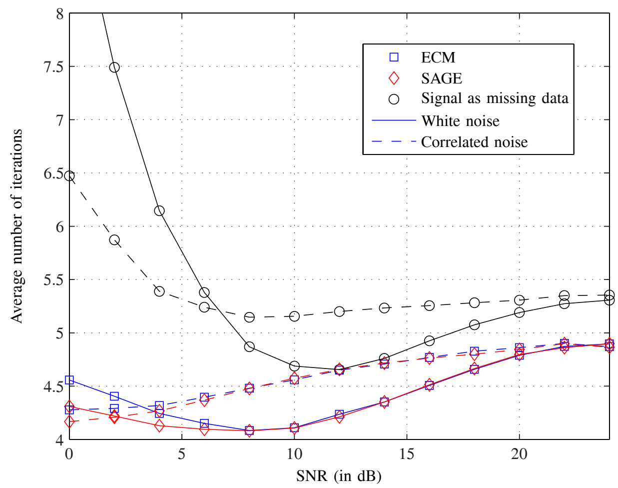

Fig. 2. Average number of iterations: proposed ECM based, SAGE based and the approach that considers signal as the missing data.

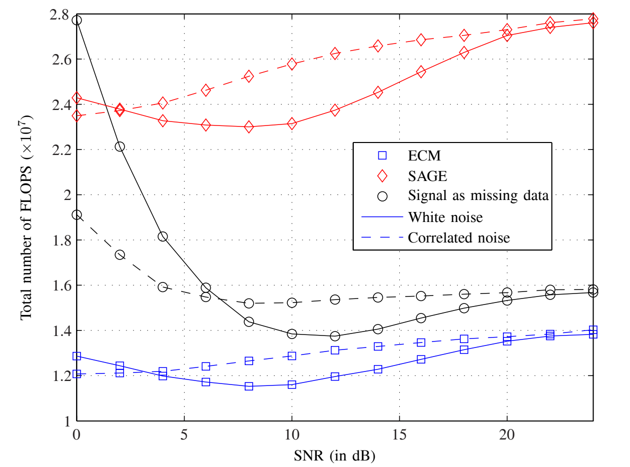

Fig. 3. Total number of FLOPS: proposed ECM based, SAGE based and the approach that considers signal as the missing data.

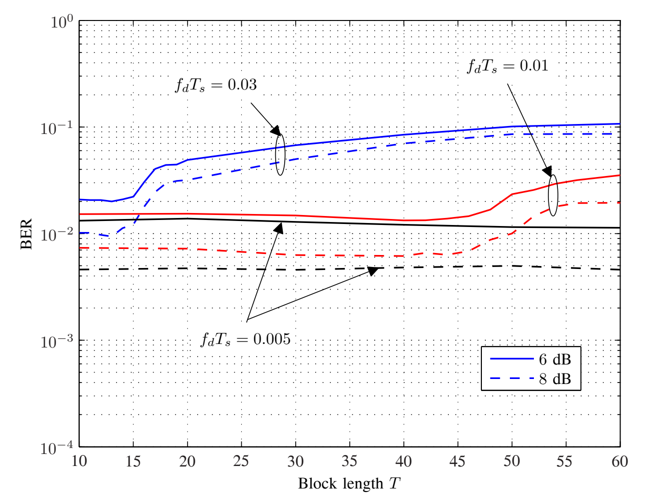

Fig. 4. Effect of block length T in white noise environment.

data and the SAGE based algorithm. Our approach has the same performance in terms of BER with the others.

In Fig. 2, the required average number of iterations for our proposed algorithm, the SAGE based algorithm and the approach extended form [4] are presented. We can see that the SAGE based algorithm has smaller average number of iterations in low-SNR region for both noise cases. This is due to the estimation of channel in each step of updating $\boldsymbol{S}$ and $\boldsymbol{\Sigma}$. This makes convergence faster. However, at high SNR, our approach and the SAGE based algorithm have the same average number of iterations. The approach extended from [4] needs higher number of iterations for the whole considered SNR region.

In Fig. 3, we show the higher computational cost (in terms of needed total number of flops) of the SAGE based algorithm compared to our proposed algorithm. Our proposed algorithm is about 2 times faster than the SAGE based algorithm (or equivalently, our proposed algorithm saves around 50% of computational resources compared with the SAGE based algorithm). Note that one complex multiplication/division takes six floating-point operation (flops) and one complex addition needs two flops. Our approach is about 9% faster than the approach extended from [4].

Finally, we consider the impact of block length T on the performance of the proposed algorithm. T represents the trade-off between consumed energy in pilot symbols and not sampling the fading process fast enough to have good channel initializations. Fig. 4 and Fig. 5 show the effect at fading rates of $f_dT_s = 0.03$, $f_dT_s = 0.01$ and $f_dT_s = 0.005$ with white and correlated noise environments, respectively. The BER goes up steeply when the block size causes the sampling rate of fading process to fall below the Nyquist rate of $1/(2f_dT_s)$ [10].

VI. Conclusions

In this letter, we propose a computationally efficient iterative receiver for SIMO systems under fast-fading environment with unknown spatially correlated noise based on the ECM algorithm. The obtained performance for the fading rate of 10^{-2} is around 1.5dB and 2dB away from the performance of ML receiver with perfect CSI and noise covariance matrix in the white and correlated noise environments, respectively.

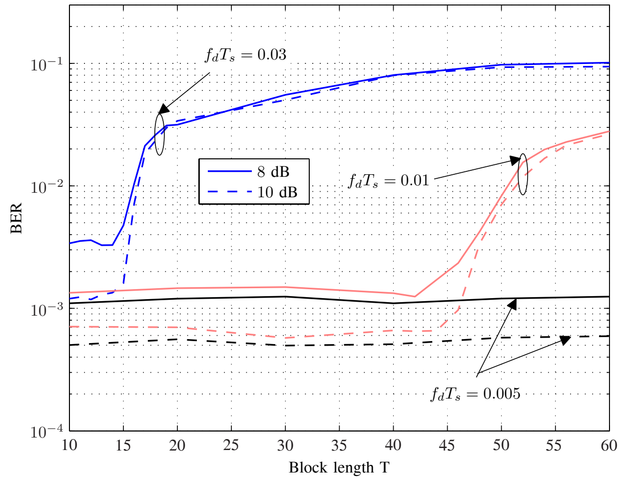

Fig. 5. Effect of block length T in correlated noise environment.

Furthermore, we also make a comparison in terms of BER and computational complexity with the approach based on the SAGE algorithm which is proposed in [6], the approach considering signal as the missing data in [4] for quasi-static fading channels. We outline these approaches applied for fast fading channels and show that our proposed approach is computationally efficient yet maintains the same BER performance.

ACKNOWLEDGEMENT

The authors would like to thank the anonymous reviewers for their critical comments that greatly improved this paper.

REFERENCES

[1] C. Cozzo and B. L. Hughes, "Joint channel estimation and data detection in space-time communications," *IEEE Trans. Commun.*, no. 8, pp. 1266–1270, 2003.

[2] J. Choi, "MIMO-BICM doubly-iterative receiver with the EM based channel estimation," in *Proc. IEEE Vehicular Technology Conference*, vol. 5, Sep. 2005, pp. 952–956.

[3] G. J. McLachlan and T. Krishnan, *The EM Algorithm and Extension*. New York: Wiley, 1997.

[4] A. Dogandžić, W. Mo, and Z. Wang, "Semi-blind SIMO flat-fading channel estimation in unknown spatially correlated noise using the EM algorithm," *IEEE Trans. Signal Processing*, vol. 52, no. 6, pp. 1791–1797, Jun. e2004.

[5] R. F. Potthoff and S. N. Roy, "A generalized multivariate analysis of variance model useful especially for growth curve problems," *Biometrika*, vol. 51, no. 3/4, pp. 313–326, 1964.

[6] P. Bohlin, "Iterative least squares detection using EM-algorithm," in *Proc. 27th General Assembly of URSI*, Maastricht, The Netherland, Aug. 2002.

[7] T. A. Fesler and A. O. Hero, "Space-alternating generalized expectation-maximization algorithm," *IEEE Trans. Signal Processing*, vol. 42, no. 10, pp. 2664–2677, Oct. 1994.

[8] X. L. Meng and D. B. Rubin, "Maximum likelihood estimation via the ECM algorithm: A general framework," *Biometrika*, vol. 80, no. 2, pp. 267–278, June 1993.

[9] W. C. Jakes, *Microwave Mobile Communications*. New York: IEEE press, 1993.

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