

RESEARCH

Open Access

# Iterative channel estimation and data detection for MIMO-OFDM systems operating in time-frequency dispersive channels under unknown background noise

Ke Zhong\*, Xia Lei and Shaoqian Li

## Abstract

In this paper, the challenging problem of joint channel estimation and data detection for multiple-input multiple-output orthogonal frequency division multiplexing systems operating in time-frequency dispersive channels under unknown background noise is investigated. Based on two different but equivalent signal models, two expectation-maximization algorithm-based iterative schemes for joint data detection and channel and noise variance estimation are proposed. The first scheme jointly detects data and estimates the channel and noise variance, but the computational complexity is high, owing to the simultaneous detection and estimation for all antennas. To reduce the computational complexity, a complexity-reduced scheme that is detecting data and estimating channel for only one antenna during each iteration and holding the unknown quantities of other antennas to their last values is proposed, whose performance only slightly degrades compared to the first scheme. Moreover, both schemes are derived as closed-form expressions, and therefore, our proposed schemes are free of exhaustive search. Simulation results demonstrate quick convergence of the proposed algorithm, and after convergence, the performance of the proposed algorithm is close to that of the optimal channel estimation and data detection case, which assumes full training and perfect channel state information.

**Keywords:** Multiple-input multiple-output (MIMO); Orthogonal frequency division multiplexing (OFDM); Time-frequency (TF) dispersive channels; Unknown noise variance; Expectation-maximization (EM)

## 1 Introduction

Multiple-input multiple-output (MIMO) communication [1] can significantly increase the throughput without increasing the transmit power and additional bandwidth. Orthogonal frequency division multiplexing (OFDM) [2] can provide high data rate transmission capability and is robust against multipath (time-dispersive) fading channels. MIMO combined with OFDM (MIMO-OFDM) [3] has been adopted in various international standards such as 3GPP-LTE, WiMAX, and IMT-Advanced.

Meanwhile, vehicles with increased speeds, such as high-speed cars, subways, and trains which exceed 350 km/h, play an increasingly important role in peoples' lives.

Consequently, mobility support is widely regarded as one of the key features in current and future wireless communication systems. High mobility causes the transmission channel to change rapidly in time, which results in frequency dispersion of the channel. For coherent detection in MIMO-OFDM systems, channel state information (CSI) is indispensable [3].

CSI acquisition is particularly challenging in time-frequency (TF) dispersive channels because channel responses vary sample by sample, and therefore, the number of unknown channel parameters in an OFDM

\*Correspondence: zhongke@uestc.edu.cn

<sup>1</sup>National Key Laboratory of Science and Technology on Communications, University of Electronic Science and Technology of China, Chengdu 611731, People's Republic of China

symbol period increases significantly (much greater than in frequency-nondispersive channels). Furthermore, in practical communication scenarios, the knowledge of the power of background noise is required to perform many signal processing algorithms, such as channel estimation [4] and decoding [5] in MIMO-OFDM systems.

In this paper, joint data detection and channel and noise variance estimation for MIMO-OFDM systems operating in TF dispersive channels under unknown background noise are investigated. We employ the expectation-maximization (EM) algorithm [6,7], which is an iterative numerical method employed to compute the maximum likelihood (ML) estimates, to develop an iterative algorithm to solve this challenging problem.

For MIMO systems, the literature along these lines can be categorized as follows:

*EM for channel estimation and data detection assuming the noise variance is known:* EM-based joint channel estimation and data detection algorithms in time-nondispersive and frequency-nondispersive channels (TnDFnD channels) are proposed in [8-10], and in time-dispersive and frequency-nondispersive channels (TDFnD channels) are proposed in [11-13], respectively. However, the maximization step (M-step) for data detection proposed in these papers is not obtained as a closed-form solution, and therefore, a brute-force searching over all of the possibilities is required.

*EM for channel and noise variance estimation:* In TnDFnD channels, EM-based joint channel and noise variance estimation algorithms are proposed in [14-16]. However, data detection is obtained by an extra ML estimator and a maximizing *a posteriori* probabilities (APP) detector in [14,15], respectively.

In [16], a full training sequence is adopted to perform the proposed EM algorithm, and therefore, no data detection is addressed. In TDFnD channels, EM-based joint channel and noise variance estimation algorithms are proposed in [17-19]. However, data detection is not addressed in these papers.

*EM for data detection and noise variance estimation:* In TnDFnD channels, an EM-based joint data detection and noise variance estimation algorithm is proposed in [20]. However, the channel estimate is only obtained by pilot symbols and is not included in the EM updating process.

*EM only for data detection assuming the noise variance is known:* In TnDFnD channels, EM-based data detection algorithms are proposed in [21,22]. However, channel estimation is not addressed in [21], and the channel knowledge is assumed ideally known at

the receiver in [22]. In TF channels, an EM-based data detection algorithm is proposed in [23] to solve a maximum a posteriori probability (MAP) detection problem. However, the data estimate is not given by a closed form, and therefore, the exhaustive search is required.

*EM only for channel estimation assuming the noise variance is known:* In TnDFnD channels, EM-based channel estimation algorithms are proposed in [24-28]. However, the data estimates are obtained by extra MAP estimators in [24-26] and APP estimators in [27,28], respectively. In TDFnD channels, EM-based channel estimation algorithms are proposed in [29-32]. However, the data estimates are obtained by an extra BI-GDFE detector in [29], a minimum mean-squared error (MMSE) detector in [30], a trellises approach in [31], respectively, and data detection is not addressed in [32].

In this paper, based on two different but equivalent signal models, two EM algorithm-based iterative schemes which integrate data detection and channel and noise variance estimation are proposed in a consistent way so as to iteratively improve the system performance.

The first scheme jointly detects data and estimates the channel and noise variance, but the computational complexity is high, owing to the simultaneous detection and estimation for all antennas. To reduce the computational complexity of the first scheme, another scheme that performs data detection and channel estimation for only one antenna during each iteration and holding the unknown quantities of other antennas to their last values is proposed, whose performance only slightly degrades compared to the first scheme. Furthermore, the estimates of data, channel, and noise variance are all obtained as closed-form results, and therefore, the proposed schemes are free of exhaustive search. Simulation results demonstrate quick convergence of the proposed algorithm, and after convergence, the performance of the proposed iterative algorithm is close to that of the optimal channel estimation and data detection case, which assumes full training and perfect CSI.

The remainder of this paper is organized as follows. The system model for MIMO-OFDM systems operating in TF dispersive channels under unknown background noise is introduced in Section 2.

In Section 3, an EM-based scheme for joint data detection and channel and noise variance estimation is proposed. In Section 4, a reduced complexity EM-based scheme is proposed. Section 5 gives some simulation results that demonstrate the effectiveness of the proposed schemes. Finally, conclusions are drawn in Section 6.

*Notation:* Matrices and vectors are represented by boldface uppercase and lowercase letters, respectively.

A hat over a variable (e.g.,  $\hat{\mathbf{x}}$ ) indicates an estimate of the variable.  $\mathbb{E}\{\cdot\}$  denotes the expectation. Superscripts  $[\cdot]^T$ ,  $[\cdot]^{-1}$ , and  $[\cdot]^H$  denote the transpose, the matrix inversion, and the Hermitian operations, respectively.  $\mathbf{I}_N$  is an identity matrix with dimension  $N$ .  $\text{diag}\{\mathbf{x}\}$  and  $\text{Blkdiag}\{\cdot\}$  stand for the diagonal matrix with vector  $\mathbf{x}$  on its diagonal and the block diagonal concatenation of input arguments, respectively. The symbol  $\otimes$  denotes convolution, and  $\otimes$  stands for the Kronecker product.  $\text{Tr}\{\mathbf{X}\}$  and  $|\mathbf{X}|$  are the trace and the determinant of a square matrix  $\mathbf{X}$ , respectively.  $\Re\{\cdot\}$  is the real part of the element in the bracket.  $\langle \cdot \rangle_K$  denotes the mode  $K$  operation. The matrix  $\mathbf{F}$  is the normalized fast Fourier transform (FFT) matrix with  $[\mathbf{F}]_{m,n} = \frac{1}{\sqrt{N}} e^{-j2\pi mn/N}$ .

## 2 System model

### 2.1 Transmitted MIMO-OFDM systems with scattered pilots

We consider a MIMO-OFDM system with  $N_T$  transmit and  $N_R$  receive antennas. For the  $i$ th transmit antenna, the time domain signal  $\mathbf{s}^i = [s^i(0), s^i(1), \dots, s^i(N-1)]^T$  is generated by taking the  $N$ -point inverse FFT of the source signal in the frequency domain  $\mathbf{x}^i = [x^i(0), x^i(1), \dots, x^i(N-1)]^T$  as  $\mathbf{s}^i = \mathbf{F}^H \mathbf{x}^i$ .

In general, the elements of  $\mathbf{x}^i$  can be categorized into:

$$\mathbf{x}^i(m) = \begin{cases} x_d^i(m) & \forall m \in I_d^i \\ x_p^i(m) & \forall m \in I_p^i \end{cases} \quad (1)$$

where  $I_d^i$  is the index set of subcarriers allocated for data symbols (with  $N_d$  elements), and  $I_p^i$  is the index set of subcarriers allocated for pilot symbols (with  $N_p$  elements), respectively. Notice that  $N = N_d + N_p$ . From (1), we have  $\mathbf{x}^i = \mathbf{E}_d^i \mathbf{x}_d^i + \mathbf{E}_p^i \mathbf{x}_p^i$ , where  $\mathbf{E}_d^i$  and  $\mathbf{E}_p^i$  denote the matrices collecting columns of  $\mathbf{I}_N$  corresponding to  $I_d^i$  and  $I_p^i$ , respectively, and  $\mathbf{x}_d^i = [x_d^i(0), x_d^i(1), \dots, x_d^i(N_d-1)]^T$  and

$\mathbf{x}_p^i = [x_p^i(0), x_p^i(1), \dots, x_p^i(N_p-1)]^T$  denote the data and pilot vectors, respectively.

A cyclic prefix (CP) with length  $N_{cp}$  larger than that of the longest channel response is inserted at the beginning of each OFDM symbol to prevent intersymbol interference.

### 2.2 TF dispersive channels under unknown background noise model

At the receive antenna  $j$ , assuming perfect timing and frequency synchronization are achieved, the  $n$ th sample of the received signal is given by:

$$y^j(n) = \sum_{l=0}^{N_T-1} h^{ji}(n, l) \otimes s^i(n) + w^j(n), \quad (2)$$

where  $h^{ji}(n, l)$  is the TF dispersive channel of the  $l$ th path with length  $L$  at time  $n$ , associated with the  $i$ th transmit antenna and the  $j$ th receive antenna, and  $w^j(n)$  denotes the unknown background noise and is assumed to obey complex Gaussian distribution with zero mean and unknown variance  $\sigma^2$ , which is assumed to be the same across all receive antennas.

After discarding the CP and stacking all  $N$  samples, the received signal for a whole OFDM symbol at the receive antenna  $j$  can be expressed in a vector form as:

$$\mathbf{y}^j = \sum_{i=0}^{N_T-1} \mathbf{H}^{ji} \mathbf{s}^i + \mathbf{w}^j, \quad (3)$$

where  $\mathbf{y}^j = [y^j(0), y^j(1), \dots, y^j(N-1)]^T$  and  $\mathbf{w}^j = [w^j(0), w^j(1), \dots, w^j(N-1)]^T$  denote the received signal at the receive antenna  $j$  and the corresponding noise, respectively.

$\mathbf{H}^{ji}$  represents the corresponding TF dispersive channel matrix and is expressed as:

$$\mathbf{H}^{ji} = \begin{bmatrix} h^{ji}(0,0) & \mathbf{0} \dots & h^{ji}(0,L-1) \dots & h^{ji}(0,1) \\ \vdots & \vdots & \ddots & \vdots \\ h^{ji}(L-1,L-1) & h^{ji}(L-1,L-2) \dots & h^{ji}(L-1,0) & \mathbf{0} \dots \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} \dots & h^{ji}(N-1,L-1) & h^{ji}(N-1,L-2) \dots & h^{ji}(N-1,0) \end{bmatrix} \quad (4)$$

It is observed from (4) that the number of unknowns in  $\mathbf{h}^{ji}$  is  $NL$ , which is much larger than the number of received samples. Therefore, direct estimation of  $\mathbf{h}^{ji}$  is almost impossible (i.e., this will give rise to the identifiability problem).

To overcome this problem, in this paper, a parsimonious (low-dimensional) representation of  $h^{ji}(n, l)$  using the basis expansion model (BEM) [33,34] is adopted, i.e., using an expansion with respect to time  $n$  of each path  $l$  of  $h^{ji}(n, l)$  into a basis  $\{b_{n,q}\}_{q=0}^Q$  as:

$$h^{ji}(n, l) = \sum_{q=0}^Q \beta_{q,l}^{ji} b_{n,q} \quad (5)$$

where  $\beta_{q,l}^{ji}$  is the  $q$ th BEM coefficient of the  $l$ th path associated with the channel between the  $i$ th transmit antenna and the  $j$ th receive antenna;  $b_{n,q}$  is the basis that captures channel time variations, and  $Q + 1$  is the number of the basis. BEM is motivated by the observation that the temporal ( $n$ ) variation of  $h(n, l)$  is usually rather smooth due to the channel's limited Doppler spread and therefore  $\{b_{n,q}\}_{q=0}^Q$  can be chosen as a small set (i.e.,  $Q \ll N$ ) of smooth functions.

Below, two equivalent expressions for the received signal will be derived, from which closed-form solution for data detection and channel estimation can be obtained, as will be shown in the following sections.

Notice that (3) can be rewritten as:

$$\mathbf{y}^j = \sum_{i=0}^{N_T-1} \mathbf{G}[\mathbf{s}^i] \mathbf{h}^{ji} + \mathbf{w}^j, \quad (6)$$

where  $\mathbf{G}[\mathbf{s}^i] = [\text{diag}\{\mathbf{s}_{cs,0}^i\}, \text{diag}\{\mathbf{s}_{cs,1}^i\}, \dots, \text{diag}\{\mathbf{s}_{cs,L-1}^i\}]$  with  $\mathbf{s}_{cs,l}^i$  representing cyclically shifts (cs)  $\mathbf{s}^i$  by  $l$  positions and  $\mathbf{h}^{ji} = [(\mathbf{h}_0^j)^T, (\mathbf{h}_1^j)^T, \dots, (\mathbf{h}_{L-1}^j)^T]^T$  with  $\mathbf{h}_l^j = [h^{ji}(0, l), h^{ji}(1, l), \dots, h^{ji}(N-1, l)]^T$ . (6) can be put into a more compact form as:

$$\mathbf{y}^j = \mathbf{G}[\mathbf{s}] \mathbf{h}^j + \mathbf{w}^j, \quad (7)$$

where  $\mathbf{G}[\mathbf{s}] = [\mathbf{G}[\mathbf{s}^0], \mathbf{G}[\mathbf{s}^1], \dots, \mathbf{G}[\mathbf{s}^{N_T-1}]]$  and  $\mathbf{h}^j = [(\mathbf{h}^{j0})^T, (\mathbf{h}^{j1})^T, \dots, (\mathbf{h}^{j(N_T-1)})^T]^T$ . Using (5),  $\mathbf{h}_l^j$  can be expressed in a vector form as

$$\mathbf{h}_l^j = \mathbf{B} \boldsymbol{\beta}_l^{ji}, \quad (8)$$

where  $\mathbf{B} = [\mathbf{b}_0, \mathbf{b}_1, \dots, \mathbf{b}_Q]$  with  $\mathbf{b}_q = [b_{0,q}, b_{1,q}, \dots, b_{N-1,q}]^T$  and  $\boldsymbol{\beta}_l^{ji} = [\beta_{0,l}^{ji}, \beta_{1,l}^{ji}, \dots, \beta_{Q,l}^{ji}]^T$ . Substituting (8) into (7), we obtain:

$$\mathbf{y}^j = \mathbf{G}[\mathbf{s}] (\mathbf{I}_{N_T L} \otimes \mathbf{B}) \boldsymbol{\beta}^j + \mathbf{w}^j, \quad (9)$$

where

$\boldsymbol{\beta}^j = [(\boldsymbol{\beta}^{j0})^T, (\boldsymbol{\beta}^{j1})^T, \dots, (\boldsymbol{\beta}^{j(N_T-1)})^T]^T$  with  $\boldsymbol{\beta}^{ji} = [(\beta_0^{ji})^T, (\beta_1^{ji})^T, \dots, (\beta_{L-1}^{ji})^T]^T$ . By stacking the received signals from all  $N_R$  receive antennas into a single vector using (9) and (3), two equivalent expressions of the received signal which explicitly show the dependence of the unknown BEM coefficient and unknown signal can be obtained, respectively, as:

$$\mathbf{y} = \boldsymbol{\Xi}[\mathbf{s}] \boldsymbol{\beta} + \mathbf{w} \quad (10a)$$

$$= \boldsymbol{\Theta}[\boldsymbol{\beta}] \mathbf{s} + \mathbf{w}, \quad (10b)$$

where  $\boldsymbol{\Xi}[\mathbf{s}] = \mathbf{I}_{N_R} \otimes (\mathbf{G}[\mathbf{s}] (\mathbf{I}_{N_T L} \otimes \mathbf{B}))$ ,  $\boldsymbol{\beta} = [(\boldsymbol{\beta}^0})^T, (\boldsymbol{\beta}^1})^T, \dots, (\boldsymbol{\beta}^{N_R-1})^T]^T$ ,  $\boldsymbol{\Theta}[\boldsymbol{\beta}] = [\mathbf{H}^0, \mathbf{H}^1, \dots, \mathbf{H}^{N_T-1}]$  with  $\mathbf{H}^i = [(\mathbf{H}^{0i})^H, (\mathbf{H}^{1i})^H, \dots, (\mathbf{H}^{(N_R-1)i})^H]^H$ ,  $\mathbf{y} = [(\mathbf{y}^0})^T, (\mathbf{y}^1})^T, \dots, (\mathbf{y}^{N_R-1})^T]^T$ ,  $\mathbf{s} = [(\mathbf{s}^0})^T, (\mathbf{s}^1})^T, \dots, (\mathbf{s}^{N_T-1})^T]^T$  and  $\mathbf{w} = [(\mathbf{w}^0})^T, (\mathbf{w}^1})^T, \dots, (\mathbf{w}^{N_R-1})^T]^T$ . Notice that  $\boldsymbol{\Xi}[\mathbf{s}]$  represents a function of  $\mathbf{s}$  and can be reconstructed by  $\mathbf{s}$  through (6), (7), (8) and (9). Similarly,  $\boldsymbol{\Theta}[\boldsymbol{\beta}]$  represents a function of  $\boldsymbol{\beta}$  and can be reconstructed by  $\boldsymbol{\beta}$  through (3), (4) and (5).

### 3 Iterative data detection and channel and noise variance estimation

The ML solution of all unknown quantities in (10), i.e.,  $\mathbf{s}$ ,  $\boldsymbol{\beta}$ , and  $\sigma^2$  of  $\mathbf{w}$ , involves multidimensional searches that pose prohibitively high computational complexity. In this and the next sections, the EM algorithm is employed to iteratively compute the ML estimates, with the different accuracy versus complexity trade-offs, respectively. As will be seen, our proposed schemes provide not only computationally affordable but also closed-form solutions that are free of exhaustive search.

Using the EM terminology, we take  $\mathbf{y}$  as the incomplete data,  $\boldsymbol{\beta}$  as the unobservable or missing data, and  $(\sigma^2, \mathbf{s})$  as parameters of interest. The iterative algorithm includes two steps (the E-step and the M-step) at each iteration. In the E-step, an expectation is taken with respect to  $\boldsymbol{\beta}$  conditional on the observed data  $\mathbf{y}$  and the previous estimates of  $(\sigma^2, \mathbf{s})$ , and an objective function depending only on  $(\sigma^2, \mathbf{s})$  is obtained. In the M-step, through maximizing the function obtained in the E-step, the effect of channel can be compensated, and the current updated estimates of  $(\sigma^2, \mathbf{s})$  can be obtained.

The two steps at the  $k$ th iteration are detailed as follows:

*E-step:* compute  $\mathbb{Q}(\sigma^2, \mathbf{s} | \hat{\sigma}_{k-1}^2, \hat{\mathbf{s}}_{k-1}) = \mathbb{E}\{\log f(\mathbf{y}, \boldsymbol{\beta} | \sigma^2, \mathbf{s}) | \mathbf{y}, \hat{\sigma}_{k-1}^2, \hat{\mathbf{s}}_{k-1}\}$ .

*M-step:* solve  $(\hat{\sigma}_k^2, \hat{\mathbf{s}}_k) = \arg \max_{\sigma^2, \mathbf{s}} \mathbb{Q}(\sigma^2, \mathbf{s} | \hat{\sigma}_{k-1}^2, \hat{\mathbf{s}}_{k-1})$ .

Note that conditioned upon  $\mathbf{y}$ , the only unknown or random component in the complete data  $(\mathbf{y}, \boldsymbol{\beta})$  is  $\boldsymbol{\beta}$ ; the expectation is taken with respect to the conditional probability density function  $f(\boldsymbol{\beta} | \mathbf{y}, \hat{\sigma}_{k-1}^2, \hat{\mathbf{s}}_{k-1})$ , while  $(\hat{\sigma}_k^2, \hat{\mathbf{s}}_k)$  are the estimates of  $\sigma^2$  and  $\mathbf{s}$  at the  $k$ th iteration.

More specifically, for the *E-step*: using Bayes's rule, we have:

$$f(\mathbf{y}, \boldsymbol{\beta} | \sigma^2, \mathbf{s}) = f(\mathbf{y} | \boldsymbol{\beta}, \sigma^2, \mathbf{s}) f(\boldsymbol{\beta}), \quad (11)$$

where the fact that  $\boldsymbol{\beta}$  is independent of  $\mathbf{s}$  and  $\sigma^2$  has been used. From (11), the function  $\mathbb{Q}(\sigma^2, \mathbf{s} | \hat{\sigma}_{k-1}^2, \hat{\mathbf{s}}_{k-1})$  in the E-step can be expressed as:

$$\begin{aligned} \mathbb{Q}(\sigma^2, \mathbf{s} | \hat{\sigma}_{k-1}^2, \hat{\mathbf{s}}_{k-1}) &= \mathbb{E}\{\log f(\mathbf{y} | \boldsymbol{\beta}, \sigma^2, \mathbf{s}) | \mathbf{y}, \hat{\sigma}_{k-1}^2, \hat{\mathbf{s}}_{k-1}\} \\ &\quad + \mathbb{E}\{\log f(\boldsymbol{\beta}) | \mathbf{y}, \hat{\sigma}_{k-1}^2, \hat{\mathbf{s}}_{k-1}\}, \end{aligned} \quad (12)$$

where the second term can be ignored in the following derivations, since it is not a function of parameters of interest, i.e., not a function of  $(\sigma^2, \mathbf{s})$  and therefore will not affect the following M-step. Using (10a), the likelihood function  $f(\mathbf{y} | \boldsymbol{\beta}, \sigma^2, \mathbf{s})$  is obtained as:

$$\begin{aligned} f(\mathbf{y} | \boldsymbol{\beta}, \sigma^2, \mathbf{s}) &= \frac{1}{(\pi \sigma^2)^{N_R N}} \\ &\quad \times \exp\left(-\frac{1}{\sigma^2} (\mathbf{y} - \Xi[\mathbf{s}]\boldsymbol{\beta})^H (\mathbf{y} - \Xi[\mathbf{s}]\boldsymbol{\beta})\right). \end{aligned} \quad (13)$$

Substituting (13) into (12), we have:

$$\begin{aligned} \mathbb{Q}(\sigma^2, \mathbf{s} | \hat{\sigma}_{k-1}^2, \hat{\mathbf{s}}_{k-1}) &\propto -NN_R \log(\pi \sigma^2) \\ &\quad - \frac{1}{\sigma^2} (\mathbf{y}^H \mathbf{y} - 2\Re\{\mathbf{y}^H \Xi[\mathbf{s}] \mathbb{E}\{\boldsymbol{\beta} | \mathbf{y}, \hat{\sigma}_{k-1}^2, \hat{\mathbf{s}}_{k-1}\}\} \\ &\quad + \mathbb{E}\{\boldsymbol{\beta}^H \Xi^H[\mathbf{s}] \Xi[\mathbf{s}] \boldsymbol{\beta} | \mathbf{y}, \hat{\sigma}_{k-1}^2, \hat{\mathbf{s}}_{k-1}\}). \end{aligned} \quad (14)$$

Notice that the following equation holds true for any matrix  $\mathbf{A}$  and vector  $\mathbf{a}$  with compatible dimension:

$$\mathbf{a}^H \mathbf{A}^H \mathbf{A} \mathbf{a} = \text{Tr}\{\mathbf{A}^H \mathbf{A} \mathbf{a} \mathbf{a}^H\}. \quad (15)$$

Define the conditional mean of  $\boldsymbol{\beta}$  in (14) as  $\hat{\boldsymbol{\beta}}_k = \mathbb{E}\{\boldsymbol{\beta} | \mathbf{y}, \hat{\sigma}_{k-1}^2, \hat{\mathbf{s}}_{k-1}\}$ , and using (15), we obtain:

$$\begin{aligned} \mathbb{Q}(\sigma^2, \mathbf{s} | \hat{\sigma}_{k-1}^2, \hat{\mathbf{s}}_{k-1}) &\propto -NN_R \log(\pi \sigma^2) \\ &\quad - \frac{1}{\sigma^2} \left( \mathbf{y}^H \mathbf{y} - 2\Re\{\mathbf{y}^H \Xi[\mathbf{s}] \hat{\boldsymbol{\beta}}_k\} \right. \\ &\quad \left. + \text{Tr}\{\Xi^H[\mathbf{s}] \Xi[\mathbf{s}] (\hat{\mathbf{Y}}_k + \hat{\boldsymbol{\beta}}_k \hat{\boldsymbol{\beta}}_k^H)\} \right), \end{aligned} \quad (16)$$

where  $\hat{\mathbf{Y}}_k = \mathbb{E}\{(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}_k)(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}_k)^H | \mathbf{y}, \hat{\sigma}_{k-1}^2, \hat{\mathbf{s}}_{k-1}\}$  represents the corresponding conditional covariance matrix of  $\boldsymbol{\beta}$ . It

is shown in Appendix 1 that the conditional mean and covariance matrix are approximately given by:

$$\hat{\boldsymbol{\beta}}_k = (\Xi^H[\hat{\mathbf{s}}_{k-1}] \Xi[\hat{\mathbf{s}}_{k-1}])^{-1} \Xi^H[\hat{\mathbf{s}}_{k-1}] \mathbf{y}, \quad (17)$$

$$\hat{\mathbf{Y}}_k = \hat{\sigma}_{k-1}^2 (\Xi^H[\hat{\mathbf{s}}_{k-1}] \Xi[\hat{\mathbf{s}}_{k-1}])^{-1}. \quad (18)$$

*M-step*: in this step, we aim to maximize  $\mathbb{Q}(\sigma^2, \mathbf{s} | \hat{\sigma}_{k-1}^2, \hat{\mathbf{s}}_{k-1})$  with respect to  $\sigma^2$  and  $\mathbf{s}$ . Differentiating (16) with respect to  $\mathbf{s}$  and setting the result to zero, neglecting those irrelevant terms we have:

$$\begin{aligned} &\frac{\partial \mathbb{Q}(\sigma^2, \mathbf{s} | \hat{\sigma}_{k-1}^2, \hat{\mathbf{s}}_{k-1})}{\partial \mathbf{s}} \\ &\quad \times \propto \frac{\partial}{\partial \mathbf{s}} \left\{ 2\Re\{\mathbf{y}^H \Xi[\mathbf{s}] \hat{\boldsymbol{\beta}}_k\} - \text{Tr}\{\Xi^H[\mathbf{s}] \Xi[\mathbf{s}] (\hat{\mathbf{Y}}_k + \hat{\boldsymbol{\beta}}_k \hat{\boldsymbol{\beta}}_k^H)\} \right\}. \end{aligned} \quad (19)$$

It is noted that since (19) depends on  $\mathbf{s}$  in an implicit way, direct maximization of (19) with respect to  $\mathbf{s}$  is difficult since multidimensional search is required. In what follows, an alternative expression for  $\mathbb{Q}(\sigma^2, \mathbf{s} | \hat{\sigma}_{k-1}^2, \hat{\mathbf{s}}_{k-1})$  will be derived from which a closed-form solution for the maximizing value of  $\mathbf{s}$  can be obtained. Since  $\hat{\mathbf{Y}}_k$  is a  $N_T N_R(Q+1)L \times N_T N_R(Q+1)L$  Hermitian matrix, based on eigen-decomposition, we have  $\hat{\mathbf{Y}}_k = \sum_{m=0}^{N_T N_R(Q+1)L-1} \lambda_{m,k} \boldsymbol{\mu}_{m,k} \boldsymbol{\mu}_{m,k}^H$ , where  $\lambda_{m,k}$  is the  $m$ th eigenvalue of  $\hat{\mathbf{Y}}_k$ , and  $\boldsymbol{\mu}_{m,k}$  is the  $m$ th eigenvector, associated with  $\lambda_{m,k}$ . Substituting the eigendecomposition on  $\hat{\mathbf{Y}}_k$  into (19) and using the two equivalent equations derived in (10a) and (10b), we have:

$$\begin{aligned} &2\Re\{\mathbf{y}^H \Xi[\mathbf{s}] \hat{\boldsymbol{\beta}}_k\} - \text{Tr}\{\Xi^H[\mathbf{s}] \Xi[\mathbf{s}] (\hat{\mathbf{Y}}_k + \hat{\boldsymbol{\beta}}_k \hat{\boldsymbol{\beta}}_k^H)\} \\ &= \mathbf{y}^H \Theta[\hat{\boldsymbol{\beta}}_k] \mathbf{s} + \mathbf{s}^H \Theta^H[\hat{\boldsymbol{\beta}}_k] \mathbf{y} - \sum_{m=0}^{N_T N_R(Q+1)L-1} \lambda_{m,k} \mathbf{s}^H \Theta^H[\boldsymbol{\mu}_{m,k}] \\ &\quad \times \Theta[\boldsymbol{\mu}_{m,k}] \mathbf{s} - \mathbf{s}^H \Theta^H[\hat{\boldsymbol{\beta}}_k] \Theta[\hat{\boldsymbol{\beta}}_k] \mathbf{s}. \end{aligned} \quad (20)$$

Notice that  $\Xi[\cdot]$  and  $\Theta[\cdot]$  defined in (10a) and (10b) are not only applicable to  $\mathbf{s}$  and  $\boldsymbol{\beta}$  but also applicable to any vectors with compatible dimension.

Since (20) is a quadratic form of  $\mathbf{s}$ , by setting the first derivative of (20) with respect to  $\mathbf{s}$  to zero, the  $k$ th signal estimate is then given by:

$$\begin{aligned} \tilde{\mathbf{s}}_k &= \left( \sum_{m=0}^{N_T N_R(Q+1)L-1} \lambda_{m,k} \Theta^H[\boldsymbol{\mu}_{m,k}] \Theta[\boldsymbol{\mu}_{m,k}] \right. \\ &\quad \left. + \Theta^H[\hat{\boldsymbol{\beta}}_k] \Theta[\hat{\boldsymbol{\beta}}_k] \right)^{-1} \Theta^H[\hat{\boldsymbol{\beta}}_k] \mathbf{y}. \end{aligned} \quad (21)$$

Note that  $\tilde{\mathbf{s}}_k = [(\tilde{\mathbf{s}}_k^0)^T, (\tilde{\mathbf{s}}_k^1)^T, \dots, (\tilde{\mathbf{s}}_k^{N_T-1})^T]^T$ . After OFDM demodulation, the symbol from the  $i$ th transmit antenna can be obtained as:

$$\tilde{\mathbf{x}}_k^i = \mathbf{F} \tilde{\mathbf{s}}_k^i. \quad (22)$$

Since  $\tilde{\mathbf{x}}_k^i$  is discrete, belonging to a symbol constellation point, it must be quantized to its nearest constellation point in each iteration. Consequently, constellation mapping is carried out to obtain the discrete symbol estimate as:  $\hat{\mathbf{x}}_k^i = \text{Quant}\{\tilde{\mathbf{x}}_k^i\}$ , where  $\text{Quant}\{\cdot\}$  operation denotes quantization on the element in the bracket. The data symbol estimate is thus obtained by collecting the elements of  $\hat{\mathbf{x}}_k^i$  corresponding to  $I_d^i$ .

Finally, putting  $\hat{\mathbf{s}}_k^i = \mathbf{F}^H \hat{\mathbf{x}}_k^i$ ,  $i = 0, 1, \dots, N_T - 1$  into (16) and setting the first derivative of  $\mathcal{Q}(\sigma^2, \mathbf{s} | \hat{\sigma}_{k-1}^2, \hat{\mathbf{s}}_{k-1})$  with respect to  $\sigma^2$  to zero, the  $k$ th estimate of the unknown noise variance can be obtained as

$$\hat{\sigma}_k^2 = \frac{1}{NN_R} \left( \mathbf{y}^H \mathbf{y} - 2\Re\{\mathbf{y}^H \Xi[\hat{\mathbf{s}}_k] \hat{\boldsymbol{\beta}}_k\} + \text{Tr}\{\Xi^H[\hat{\mathbf{s}}_k] \Xi[\hat{\mathbf{s}}_k] (\hat{\mathbf{Y}}_k + \hat{\boldsymbol{\beta}}_k \hat{\boldsymbol{\beta}}_k^H)\} \right). \quad (23)$$

In summary, starting from a suitable initial value, the proposed iterative EM-based scheme alternates among the explicitly closed-form results (17), (18), (21), (22), and (23) until convergence, i.e., until no significant changes are observed in the updates.

#### 4 A reduced computational complexity scheme

The computational complexity of the EM-based iterative scheme proposed in Section 3 is summarized in Table 1. Notice that the computational burden mainly comes from the joint detection and estimation simultaneous for all transmit antennas. If in each iteration, detection, and estimation can be completed one antenna by one antenna, the computational burden will be significantly reduced.

**Table 1 Computational complexity of the proposed scheme in Section 3**

Computation	Complexity
Matrix inversion in (17)	$O((N_T N_R (Q+1)L)^3)$
Eigendecomposition on (18)	$O((N_T N_R (Q+1)L)^3)$
Matrix inversion in (21)	$O((N_T N)^3)$

Recalling (6) and (3), two alternating but equivalent expressions for  $\mathbf{y}$  can be derived as:

$$\mathbf{y} = \Xi[\mathbf{s}] \boldsymbol{\beta} + \mathbf{w} = \sum_{i=0}^{N_T-1} \Phi[\mathbf{s}^i] (\mathbf{I}_{N_R L} \otimes \mathbf{B}) \boldsymbol{\beta}_{rc}^i + \mathbf{w} \quad (24a)$$

$$= \Theta[\boldsymbol{\beta}] \mathbf{s} + \mathbf{w} = \sum_{i=0}^{N_T-1} \mathbf{H}[\boldsymbol{\beta}_{rc}^i] \mathbf{s}^i + \mathbf{w}, \quad (24b)$$

where  $\Phi[\mathbf{s}^i] = \mathbf{I}_{N_R} \otimes \mathbf{G}[\mathbf{s}^i]$ ,  $\mathbf{H}[\boldsymbol{\beta}_{rc}^i] = [(\mathbf{H}^{0i})^H, (\mathbf{H}^{1i})^H, \dots, (\mathbf{H}^{(N_R-1)i})^H]^H$ , the BEM coefficients associated with the channel from the transmit antenna  $i$  to  $N_R$  receive antennas is represented by  $\boldsymbol{\beta}_{rc}^i = [(\boldsymbol{\beta}^{0i})^T, (\boldsymbol{\beta}^{1i})^T, \dots, (\boldsymbol{\beta}^{(N_R-1)i})^T]^T$  with  $\boldsymbol{\beta}^{ji} = [(\boldsymbol{\beta}_0^{ji})^T, (\boldsymbol{\beta}_1^{ji})^T, \dots, (\boldsymbol{\beta}_{L-1}^{ji})^T]^T$ . The subscript 'rc' is short for 'reduced complexity' to distinguish it from the  $\boldsymbol{\beta}^j$  defined in Section 3, which represents the BEM coefficients associated with the channel from  $N_T$  transmit antennas to the receive antenna  $j$ .

From (24a) and (24b), it is observed that by applying the mathematical framework of EM, an alternative way to choose the complete data, defined as  $\boldsymbol{\psi}$  in this scheme, is by decomposing the observed data  $\mathbf{y}$  into its signal components. The complete data  $\boldsymbol{\psi}$  is obtained as:

$$\boldsymbol{\psi} = \begin{bmatrix} \boldsymbol{\psi}^0 \\ \boldsymbol{\psi}^1 \\ \vdots \\ \boldsymbol{\psi}^{N_T-1} \end{bmatrix} = \begin{bmatrix} \Phi[\mathbf{s}^0] (\mathbf{I}_{N_R L} \otimes \mathbf{B}) \boldsymbol{\beta}_{rc}^0 \\ \Phi[\mathbf{s}^1] (\mathbf{I}_{N_R L} \otimes \mathbf{B}) \boldsymbol{\beta}_{rc}^1 \\ \vdots \\ \Phi[\mathbf{s}^{N_T-1}] (\mathbf{I}_{N_R L} \otimes \mathbf{B}) \boldsymbol{\beta}_{rc}^{N_T-1} \end{bmatrix} + \begin{bmatrix} \mathbf{w}^0 \\ \mathbf{w}^1 \\ \vdots \\ \mathbf{w}^{N_T-1} \end{bmatrix} \quad (25a)$$

$$= \begin{bmatrix} \mathbf{H}[\boldsymbol{\beta}_{rc}^0] \mathbf{s}^0 \\ \mathbf{H}[\boldsymbol{\beta}_{rc}^1] \mathbf{s}^1 \\ \vdots \\ \mathbf{H}[\boldsymbol{\beta}_{rc}^{N_T-1}] \mathbf{s}^{N_T-1} \end{bmatrix} + \begin{bmatrix} \mathbf{w}^0 \\ \mathbf{w}^1 \\ \vdots \\ \mathbf{w}^{N_T-1} \end{bmatrix}, \quad (25b)$$

where  $\boldsymbol{\psi}^i = \Phi[\mathbf{s}^i] (\mathbf{I}_{N_R L} \otimes \mathbf{B}) \boldsymbol{\beta}_{rc}^i + \mathbf{w}^i = \mathbf{H}[\boldsymbol{\beta}_{rc}^i] \mathbf{s}^i + \mathbf{w}^i$ ,  $i = 0, 1, \dots, N_T - 1$ , and  $\mathbf{w}^i$ ,  $i = 0, 1, \dots, N_T - 1$  are circularly symmetric and statistically independent Gaussian vectors satisfying  $\mathbf{w} = \sum_{i=0}^{N_T-1} \mathbf{w}^i$ .

Similar to the E-step in Section 3, for the  $k$ th iteration, we need to compute the conditional expectation of the log-likelihood function for the complete data  $\boldsymbol{\psi}$ . More specifically, for the

*E-step*: using (25a), the likelihood function can be expressed as:

$$f(\boldsymbol{\psi}|\boldsymbol{\beta}_{rc}, \mathbf{s}) = \frac{1}{(\pi)^{NN_R N_T} |\boldsymbol{\Upsilon}_{\boldsymbol{\psi}}|} \times \exp\left(-(\boldsymbol{\psi} - \boldsymbol{e})^H \boldsymbol{\Upsilon}_{\boldsymbol{\psi}}^{-1} (\boldsymbol{\psi} - \boldsymbol{e})\right), \quad (26)$$

where  $\boldsymbol{\beta}_{rc} = [(\boldsymbol{\beta}_{rc}^0)^T, (\boldsymbol{\beta}_{rc}^1)^T, \dots, (\boldsymbol{\beta}_{rc}^{N_T-1})^T]^T$ ,  $\boldsymbol{\Upsilon}_{\boldsymbol{\psi}} = \text{Blkdiag}\{\zeta^0 \sigma^2 \mathbf{I}_{NN_R}, \zeta^1 \sigma^2 \mathbf{I}_{NN_R}, \dots, \zeta^{N_T-1} \sigma^2 \mathbf{I}_{NN_R}\}$  with  $\mathbb{E}\{\mathbf{w}^i (\mathbf{w}^i)^H\} = \zeta^i \sigma^2 \mathbf{I}_{NN_R}$  and the  $\zeta^i$ 's being arbitrary non-negative and real-valued scalars satisfying  $\sum_{i=0}^{N_T-1} \zeta^i = 1$ ,  $\boldsymbol{e} = [(\Phi[\mathbf{s}^0] (\mathbf{I}_{N_{RL}} \otimes \mathbf{B}) \boldsymbol{\beta}_{rc}^0)^T, (\Phi[\mathbf{s}^1] (\mathbf{I}_{N_{RL}} \otimes \mathbf{B}) \boldsymbol{\beta}_{rc}^1)^T, \dots, (\Phi[\mathbf{s}^{N_T-1}] (\mathbf{I}_{N_{RL}} \otimes \mathbf{B}) \boldsymbol{\beta}_{rc}^{N_T-1})^T]^T$ . Notice that in this scheme, we take  $(\boldsymbol{\beta}_{rc}, \mathbf{s})$  as parameters of interest. Using (26) and neglecting those irrelevant terms,  $\mathbb{E}\{\log f(\boldsymbol{\psi}|\boldsymbol{\beta}_{rc}, \mathbf{s})|\mathbf{y}, \hat{\boldsymbol{\beta}}_{rc,k-1}, \hat{\mathbf{s}}_{k-1}\}$  can be expressed as:

$$\mathbb{E}\{\log f(\boldsymbol{\psi}|\boldsymbol{\beta}_{rc}, \mathbf{s})|\mathbf{y}, \hat{\boldsymbol{\beta}}_{rc,k-1}, \hat{\mathbf{s}}_{k-1}\} \times \propto \boldsymbol{e}^H \boldsymbol{\Upsilon}_{\boldsymbol{\psi}}^{-1} \hat{\boldsymbol{\psi}}_k + \hat{\boldsymbol{\psi}}_k^H \boldsymbol{\Upsilon}_{\boldsymbol{\psi}}^{-1} \boldsymbol{e} - \boldsymbol{e}^H \boldsymbol{\Upsilon}_{\boldsymbol{\psi}}^{-1} \boldsymbol{e}, \quad (27)$$

where  $\hat{\boldsymbol{\psi}}_k$ , the conditional mean of  $\boldsymbol{\psi}$ , can be derived as:

$$\hat{\boldsymbol{\psi}}_k = \mathbb{E}\{\boldsymbol{\psi}|\mathbf{y}, \hat{\boldsymbol{\beta}}_{rc,k-1}, \hat{\mathbf{s}}_{k-1}\} = \hat{\boldsymbol{e}}_{k-1} + \boldsymbol{\Upsilon}_{\boldsymbol{\psi}} \boldsymbol{U}^H (\boldsymbol{U} \boldsymbol{\Upsilon}_{\boldsymbol{\psi}} \boldsymbol{U}^H)^{-1} (\mathbf{y} - \boldsymbol{U} \hat{\boldsymbol{e}}_{k-1}), \quad (28)$$

where  $\boldsymbol{U} = [\mathbf{I}_{NN_R}, \mathbf{I}_{NN_R}, \dots, \mathbf{I}_{NN_R}]$  is a matrix with dimension  $NN_R \times NN_R N_T$  that connects  $\mathbf{y}$  and  $\boldsymbol{\psi}$  as  $\mathbf{y} = \sum_{i=0}^{N_T-1} \boldsymbol{\psi}^i = \boldsymbol{U} \boldsymbol{\psi}$ . Substituting the corresponding components into the right-hand side of (28), after some manipulations we obtain:

$$\hat{\boldsymbol{\psi}}_k = \begin{bmatrix} \Phi[\hat{\mathbf{s}}_{k-1}^0] (\mathbf{I}_{N_{RL}} \otimes \mathbf{B}) \hat{\boldsymbol{\beta}}_{rc,k-1}^0 + \zeta^0 \left( \mathbf{y} - \sum_{i=0}^{N_T-1} \Phi[\hat{\mathbf{s}}_{k-1}^i] (\mathbf{I}_{N_{RL}} \otimes \mathbf{B}) \hat{\boldsymbol{\beta}}_{rc,k-1}^i \right) \\ \Phi[\hat{\mathbf{s}}_{k-1}^1] (\mathbf{I}_{N_{RL}} \otimes \mathbf{B}) \hat{\boldsymbol{\beta}}_{rc,k-1}^1 + \zeta^1 \left( \mathbf{y} - \sum_{i=0}^{N_T-1} \Phi[\hat{\mathbf{s}}_{k-1}^i] (\mathbf{I}_{N_{RL}} \otimes \mathbf{B}) \hat{\boldsymbol{\beta}}_{rc,k-1}^i \right) \\ \vdots \\ \Phi[\hat{\mathbf{s}}_{k-1}^{N_T-1}] (\mathbf{I}_{N_{RL}} \otimes \mathbf{B}) \hat{\boldsymbol{\beta}}_{rc,k-1}^{N_T-1} + \zeta^{N_T-1} \left( \mathbf{y} - \sum_{i=0}^{N_T-1} \Phi[\hat{\mathbf{s}}_{k-1}^i] (\mathbf{I}_{N_{RL}} \otimes \mathbf{B}) \hat{\boldsymbol{\beta}}_{rc,k-1}^i \right) \end{bmatrix}. \quad (29)$$

Substituting (29) into (28), finally we obtain:

$$\mathbb{E}\{\log f(\boldsymbol{\psi}|\boldsymbol{\beta}_{rc}, \mathbf{s})|\mathbf{y}, \hat{\boldsymbol{\beta}}_{rc,k-1}, \hat{\mathbf{s}}_{k-1}\} \propto - \sum_{i=0}^{N_T-1} \left( \hat{\boldsymbol{\psi}}_k^i - \Phi[\mathbf{s}^i] (\mathbf{I}_{N_{RL}} \otimes \mathbf{B}) \boldsymbol{\beta}_{rc}^i \right)^H \times \left( \hat{\boldsymbol{\psi}}_k^i - \Phi[\mathbf{s}^i] (\mathbf{I}_{N_{RL}} \otimes \mathbf{B}) \boldsymbol{\beta}_{rc}^i \right), \quad (30)$$

where

$$\hat{\boldsymbol{\psi}}_k^i = \Phi[\hat{\mathbf{s}}_{k-1}^i] (\mathbf{I}_{N_{RL}} \otimes \mathbf{B}) \hat{\boldsymbol{\beta}}_{rc,k-1}^i + \zeta^i \left( \mathbf{y} - \sum_{i=0}^{N_T-1} \Phi[\hat{\mathbf{s}}_{k-1}^i] (\mathbf{I}_{N_{RL}} \otimes \mathbf{B}) \hat{\boldsymbol{\beta}}_{rc,k-1}^i \right). \quad (31)$$

It is noted from (30) that in the following M-step, the maximization of  $\mathbb{E}\{\log f(\boldsymbol{\psi}|\boldsymbol{\beta}_{rc}, \mathbf{s})|\mathbf{y}, \hat{\boldsymbol{\beta}}_{rc,k-1}, \hat{\mathbf{s}}_{k-1}\}$  with respect to  $\boldsymbol{\beta}_{rc}$  and  $\mathbf{s}$  is equivalent to the minimization of each of the single terms in (30), i.e., minimization of  $(\hat{\boldsymbol{\psi}}_k^i - \Phi[\mathbf{s}^i] (\mathbf{I}_{N_{RL}} \otimes \mathbf{B}) \boldsymbol{\beta}_{rc}^i)^H (\hat{\boldsymbol{\psi}}_k^i - \Phi[\mathbf{s}^i] (\mathbf{I}_{N_{RL}} \otimes \mathbf{B}) \boldsymbol{\beta}_{rc}^i)$  with respect to  $\boldsymbol{\beta}_{rc}^i$  and  $\mathbf{s}^i$  for each  $i$ , separately.

Notice that the multidimensional minimization for each of the terms in (30) still remains a formidable task. To solve this problem, substituting (24a) and (24a) into (31), we obtain:

$$\hat{\boldsymbol{\psi}}_k^i = \zeta^i \Phi[\mathbf{s}^i] (\mathbf{I}_{N_{RL}} \otimes \mathbf{B}) \boldsymbol{\beta}_{rc}^i + \boldsymbol{\chi}_k^i \quad (32a)$$

$$= \zeta^i \mathbf{H}[\boldsymbol{\beta}_{rc}^i] \mathbf{s}^i + \boldsymbol{\chi}_k^i, \quad (32b)$$

where

$$\boldsymbol{\chi}_k^i = \Phi[\hat{\mathbf{s}}_{k-1}^i] (\mathbf{I}_{N_{RL}} \otimes \mathbf{B}) \hat{\boldsymbol{\beta}}_{rc,k-1}^i + \zeta^i \left( \sum_{g=0, g \neq i}^{N_T-1} \Phi[\mathbf{s}^g] (\mathbf{I}_{N_{RL}} \otimes \mathbf{B}) \boldsymbol{\beta}_{rc}^g + \mathbf{w} - \sum_{i=0}^{N_T-1} \Phi[\hat{\mathbf{s}}_{k-1}^i] (\mathbf{I}_{N_{RL}} \otimes \mathbf{B}) \hat{\boldsymbol{\beta}}_{rc,k-1}^i \right). \quad (33)$$

Set  $\zeta^i = 1$ , for the  $i$ th transmit antenna we have:

$$\hat{\psi}_k^i = \Phi[\mathbf{s}^i](\mathbf{I}_{N_{RL}} \otimes \mathbf{B})\beta_{rc}^i + \chi_k^i \quad (34a)$$

$$= \mathbf{H}[\beta_{rc}^i]\mathbf{s}^i + \chi_k^i. \quad (34b)$$

Recalling that  $\sum_{i=0}^{N_T-1} \zeta^i = 1$ , and by using (31), for the transmit antennas  $\{g\}_{g=0, g \neq i}^{N_T-1}$  we have:

$$\hat{\psi}_k^g = \Phi[\hat{\mathbf{s}}_{k-1}^g](\mathbf{I}_{N_{RL}} \otimes \mathbf{B})\hat{\beta}_{rc,k-1}^g. \quad (35)$$

Using (34) and (35), (30) can be decomposed into  $N_T$  terms, each of which can be solved as follows:

*M-step*: for the  $i$ th transmit antenna, we have:

$$\begin{aligned} [\hat{\beta}_{rc,k}^i, \hat{\mathbf{s}}_k^i] = \arg \min_{\beta_{rc}^i, \mathbf{s}^i} \left\{ \left( \hat{\psi}_k^i - \Phi[\mathbf{s}^i](\mathbf{I}_{N_{RL}} \otimes \mathbf{B})\beta_{rc}^i \right)^H \right. \\ \left. \times \left( \hat{\psi}_k^i - \Phi[\mathbf{s}^i](\mathbf{I}_{N_{RL}} \otimes \mathbf{B})\beta_{rc}^i \right) \right\}, \end{aligned} \quad (36)$$

and for the transmit antennas  $\{g\}_{g=0, g \neq i}^{N_T-1}$ , we have:

$$[\hat{\beta}_{rc,k}^g, \hat{\mathbf{s}}_k^g] = [\hat{\beta}_{rc,k-1}^g, \hat{\mathbf{s}}_{k-1}^g]. \quad (37)$$

Therefore, the proposed iterative scheme starts from  $k = 0, 1, 2, \dots$  and during the  $k$ th iteration,  $i$  is set as  $i = < k >_{N_T}$ . It can be seen that we have split the estimation and detection problem for the MIMO case of Section 3 into estimation and detection problem for  $N_T$  single-input and multiple-output (SIMO) cases, where, during each iteration, parameters and data from only one transmit antenna are estimated and detected. Note that  $\chi_k^i$  given in (33) is a disturbance term that accounts for the background noise and residual interference after the  $k$ th iteration, where the interference is linearly related to the signals of all transmit antennas. Then, assuming the interference is i.i.d with zero mean, from the central limit theorem [35], it can be seen that the entries of  $\chi_k^i$  are nearly Gaussian distributed with zero mean and some variance  $\sigma_{\chi^i,k}^2$ . Under the above assumption, it turns out that the minimization problem in (36) is equivalent to the ML estimation of  $\beta_{rc}^i, \mathbf{s}^i$  and the unknown variance  $\sigma_{\chi^i}^2$  starting from the observation  $\hat{\psi}_k^i$ . Comparing (34a) and (34b) with (10a) and (10b), it is easy to see that the same EM procedure proposed in Section 3 can be directly adopted to solve the optimization problem of (36), with details shown in Appendix 2.

The computationally feasible EM scheme is summarized as follows:

**Iteration until convergence:**

**for the  $k$ th iteration do**

**Set** the transmit antenna as  $i = < k >_{N_T}$

1. *Updating  $\psi^i$ :*

$$\hat{\psi}_k^i = \mathbf{y} - \sum_{g=0, g \neq i}^{N_T-1} \Phi[\hat{\mathbf{s}}_{k-1}^g](\mathbf{I}_{N_{RL}} \otimes \mathbf{B})\hat{\beta}_{rc,k-1}^g.$$

2. *Updating  $\beta_{rc}^i, \Upsilon_{rc}^i, \mathbf{s}^i$  and  $\sigma_{\chi^i}^2$  until convergence*

(assume the convergence occurs at the  $\mathcal{L}$ th iteration):

**for the  $l$ th iteration do**

*Updating  $\beta_{rc}^i$  and  $\Upsilon_{rc}^i$ :*

$$\begin{aligned} \hat{\beta}_{rc,k-1,l}^{z,i} = ((\mathbf{B}^z)^H \mathbf{G}^H [\hat{\mathbf{s}}_{k-1,l-1}^i] \mathbf{G} [\hat{\mathbf{s}}_{k-1,l-1}^i] \mathbf{B}^z)^{-1} \\ \times (\mathbf{B}^z)^H \mathbf{G}^H [\hat{\mathbf{s}}_{k-1,l-1}^i] \hat{\psi}_k^{z,i}, \end{aligned}$$

$$\begin{aligned} \hat{\Upsilon}_{rc,k-1,l}^{z,i} = \hat{\sigma}_{\chi^i,k-1,l-1}^2 ((\mathbf{B}^z)^H \mathbf{G}^H [\hat{\mathbf{s}}_{k-1,l-1}^i] \\ \mathbf{G} [\hat{\mathbf{s}}_{k-1,l-1}^i] \mathbf{B}^z)^{-1}, \end{aligned}$$

where  $\hat{\mathbf{s}}_{k-1,0}^i = \hat{\mathbf{s}}_{k-1}^i$ ,  $\hat{\sigma}_{\chi^i,k-1,0}^2 = \hat{\sigma}_{\chi^i,k-1}^2$ ,  $\hat{\psi}_k^i = ((\hat{\psi}_k^{0,i})^T, (\hat{\psi}_k^{1,i})^T, \dots, (\hat{\psi}_k^{N_R-1,i})^T)^T$  and  $\mathbf{B}^z \triangleq \mathbf{I}_L \otimes \mathbf{B}$ .

Then,  $\hat{\beta}_{rc,k-1,l}^i$  can be obtained as  $\hat{\beta}_{rc,k-1,l}^i = ((\hat{\beta}_{rc,k-1,l}^{0,i})^T, (\hat{\beta}_{rc,k-1,l}^{1,i})^T, \dots, (\hat{\beta}_{rc,k-1,l}^{N_R-1,i})^T)^T$  and

$\hat{\Upsilon}_{rc,k-1,l}^i$  can be obtained as  $\hat{\Upsilon}_{rc,k-1,l}^i = \text{Blkdiag}\{\hat{\Upsilon}_{rc,k-1,l}^{0,i}, \hat{\Upsilon}_{rc,k-1,l}^{1,i}, \dots, \hat{\Upsilon}_{rc,k-1,l}^{N_R-1,i}\}$ .

*Updating  $\mathbf{s}^i$ :*

$$\begin{aligned} \hat{\mathbf{s}}_{k-1,l}^i = \mathbf{F}^H \text{Qant} \left\{ \mathbf{F} \left( \sum_{m=0}^{N_R(Q+1)L-1} \lambda_{rc,m,k-1,l}^i \mathbf{H}^H [\mu_{rc,m,k-1,l}^i] \right. \right. \\ \left. \left. \times \mathbf{H} [\mu_{rc,m,k-1,l}^i] + \mathbf{H}^H [\hat{\beta}_{rc,k-1,l}^i] \right. \right. \\ \left. \left. \times \mathbf{H} [\hat{\beta}_{rc,k-1,l}^i] \right)^{-1} \mathbf{H}^H [\hat{\beta}_{rc,k-1,l}^i] \hat{\psi}_k^i \right\}, \end{aligned}$$

where  $\hat{\Upsilon}_{rc,k-1,l}^i = \sum_{m=0}^{N_R(Q+1)L-1} \lambda_{rc,m,k-1,l}^i \mu_{rc,m,k-1,l}^i$  ( $\mu_{rc,m,k-1,l}^i$ )<sup>H</sup> represents the eigendecomposition of  $\hat{\Upsilon}_{rc,k-1,l}^i$ .

*Updating  $\sigma_{\chi^i}^2$ :*

$$\begin{aligned} \hat{\sigma}_{\chi^i,k-1,l}^2 = \frac{1}{NN_R} \left( (\hat{\psi}_k^i)^H \hat{\psi}_k^i - 2 \left\{ (\hat{\psi}_k^i)^H \Phi[\hat{\mathbf{s}}_{k-1,l}^i] \right. \right. \\ \left. \left. \times (\mathbf{I}_{N_{RL}} \otimes \mathbf{B}) \hat{\beta}_{rc,k-1,l}^i \right\} + \text{Tr} \left\{ (\mathbf{I}_{N_{RL}} \otimes \mathbf{B})^H \right. \right. \\ \left. \left. \times \Phi^H [\hat{\mathbf{s}}_{k-1,l}^i] \Phi [\hat{\mathbf{s}}_{k-1,l}^i] (\mathbf{I}_{N_{RL}} \otimes \mathbf{B}) \right. \right. \\ \left. \left. \times \left( \hat{\Upsilon}_{rc,k-1,l}^i + \hat{\beta}_{rc,k-1,l}^i (\hat{\beta}_{rc,k-1,l}^i)^H \right) \right\} \right). \end{aligned}$$

**end for**



3. **Set**  $[\hat{\beta}_{rc,k}^i, \hat{\mathbf{s}}_k^i, \hat{\sigma}_{\chi^i,k}^2] = [\hat{\beta}_{rc,k-1,\mathcal{L}}^i, \hat{\mathbf{s}}_{k-1,\mathcal{L}}^i, \hat{\sigma}_{\chi^i,k-1,\mathcal{L}}^2]$  for the transmit antenna  $i$ .  
**Set**  $[\hat{\beta}_{rc,k}^g, \hat{\mathbf{s}}_k^g, \hat{\sigma}_{\chi^g,k}^2] = [\hat{\beta}_{rc,k-1}^g, \hat{\mathbf{s}}_{k-1}^g, \hat{\sigma}_{\chi^g,k-1}^2]$  for the transmit antennas  $\{g\}_{g=0, g \neq i}^{N_T-1}$ .  
**end for**

The computational complexity of the proposed iterative EM-based scheme with reduced complexity is summarized in Table 2.

Note that compared to Table 1, the computational complexity of the proposed iterative EM-based scheme with reduced complexity is significantly lower than that of the EM-based scheme proposed in Section 3. However, this significant computational complexity reduction is not obtained without price. As will be shown in Section 5, there is a minor performance degradation compared to the EM-based scheme proposed in Section 3. This performance degradation is due mainly to two reasons. First, the disturbance term in (34a) and (34b) contains the background noise as well as the residual interference from other transmit antennas, whereas in (10a) and (10b), only the background noise is contained. Second, the separate estimation and detection for each antenna is seen as a suboptimal estimation and detection method compared to the joint estimation and detection for all antennas, which is optimal in the sense of estimation and detection theory [36].

#### 4.1 Initialization

The EM algorithm is guaranteed to obtain at least a local maximum after convergence [6,7].

To provide an initial value, a least square (LS) algorithm based on pilot symbols is utilized to provide a good initial estimate which will be demonstrated in the simulations. Recalling (1), (10a), and (10b), we have:

$$\mathbf{y} = \Xi[\Omega_p \mathbf{x}_p] \boldsymbol{\beta} + \Theta[\boldsymbol{\beta}] \Omega_d \mathbf{x}_d + \mathbf{w}, \quad (38)$$

where  $\Omega_p = \text{Blkdiag}\{\mathbf{F}^H \mathbf{E}_p^0, \mathbf{F}^H \mathbf{E}_p^1, \dots, \mathbf{F}^H \mathbf{E}_p^{N_T-1}\}$ ,  $\Omega_d = \text{Blkdiag}\{\mathbf{F}^H \mathbf{E}_d^0, \mathbf{F}^H \mathbf{E}_d^1, \dots, \mathbf{F}^H \mathbf{E}_d^{N_T-1}\}$ ,  $\mathbf{x}_p = [(\mathbf{x}_p^0)^T, (\mathbf{x}_p^1)^T, \dots, (\mathbf{x}_p^{N_T-1})^T]^T$ , and  $\mathbf{x}_d = [(\mathbf{x}_d^0)^T, (\mathbf{x}_d^1)^T, \dots, (\mathbf{x}_d^{N_T-1})^T]^T$ . By

treating the term containing  $\mathbf{x}_d$  as interference, the LS estimate of  $\boldsymbol{\beta}$  is obtained as:

$$\hat{\boldsymbol{\beta}}_0 = (\Xi^H[\Omega_p \mathbf{x}_p] \Xi[\Omega_p \mathbf{x}_p])^{-1} \Xi^H[\Omega_p \mathbf{x}_p] \mathbf{y}. \quad (39)$$

Substituting (39) into (10b), the initial signal detection is obtained as:

$$\hat{\mathbf{s}}_0 = (\mathbf{I}_{N_T} \otimes \mathbf{F}^H) \text{Qant} \left\{ (\mathbf{I}_{N_T} \otimes \mathbf{F}) \left( \Theta^H[\hat{\boldsymbol{\beta}}_0] \Theta[\hat{\boldsymbol{\beta}}_0] \right)^{-1} \times \Theta^H[\hat{\boldsymbol{\beta}}_0] \mathbf{y} \right\}. \quad (40)$$

Finally, for the initial variances  $\hat{\sigma}_0^2$  and  $\{\hat{\sigma}_{\chi^i,0}^2\}_{i=0}^{N_T-1}$ , they are all set to 0.

## 5 Simulation results and discussions

In this section, the performance of the proposed algorithm is demonstrated by Monte Carlo simulations. In the simulations, transmit and receive antennas are set as  $N_T = N_R = 2$ , each OFDM symbol has 64 subcarriers ( $N = 64$ ) and communicates over a bandwidth of 20 MHz. The sampling interval  $T_s$  is thus 50 ns. The length of the CP is  $N_{cp} = 8$ .

The normalized maximal Doppler shift is set as  $Nf_d T_s = 0.075$  and 0.15, respectively, where  $f_d$  represents the maximum Doppler frequency.

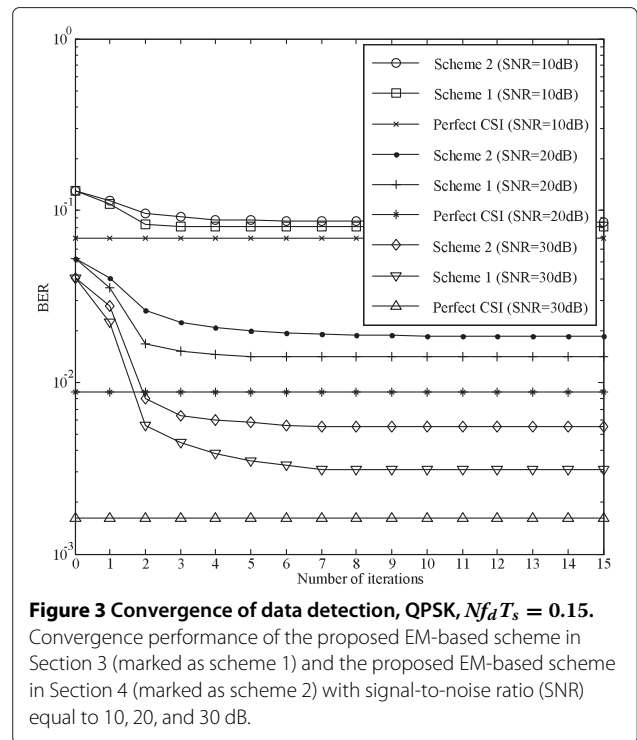
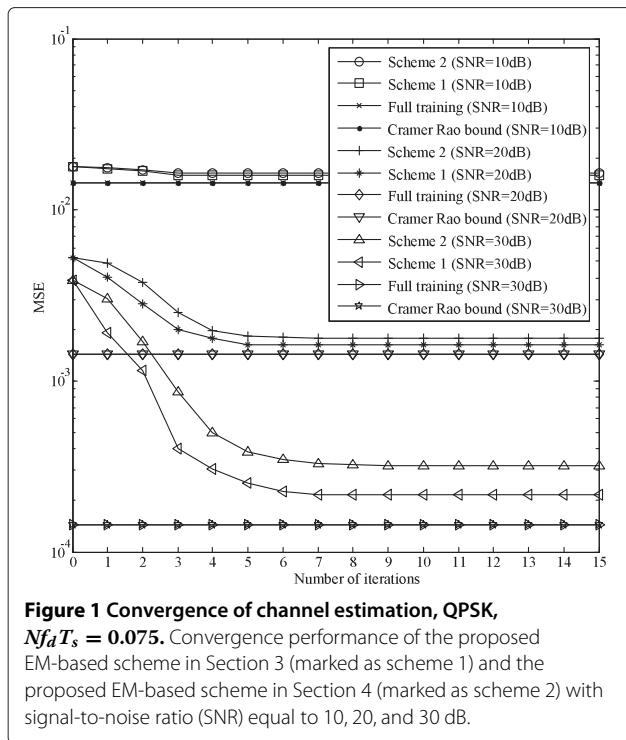
The channel has three taps ( $L = 3$ ) with an exponential power delay profile, namely  $\sigma_l^2 = \exp(-\kappa l) / (1 - \exp(-\kappa)) / (1 - \exp(-\kappa L))$ ,  $l = 0, 1, \dots, L-1$  with  $\kappa = 1/3$ . In typical communication scenarios, only a few significant paths dominate the effect of the wireless channel [4]. Therefore,  $L = 3$  is a reasonable setting. Each tap coefficient follows a complex Gaussian distribution. The data are modulated by quadrature phase shift keying (QPSK) and 16 quadrature amplitude modulation (16 QAM), respectively, with unit power. The pilot cluster follows the structure in [37], and more specifically, seven pilot clusters are used for each transmit antenna. The clusters are equal-spaced among subcarriers, and in each cluster, one nonzero pilot is guarded by one zero pilot on each side. The nonzero pilots are generated as zero-mean complex Gaussian random variables with power three times that of data symbols. Furthermore, the generalized complex exponential BEM (GCE-BEM) [34] is adopted.

### 5.1 Convergence of the proposed schemes

Figures 1, 2, and 3 present the convergence performance of the proposed EM-based scheme in Section 3 (marked as scheme 1) and the proposed EM-based scheme in Section 4 (marked as scheme 2) with signal-to-noise ratio (SNR) equal to 10, 20, and 30 dB. It can be seen that both the mean-square error (MSE) and bit error rate (BER) improve significantly in the

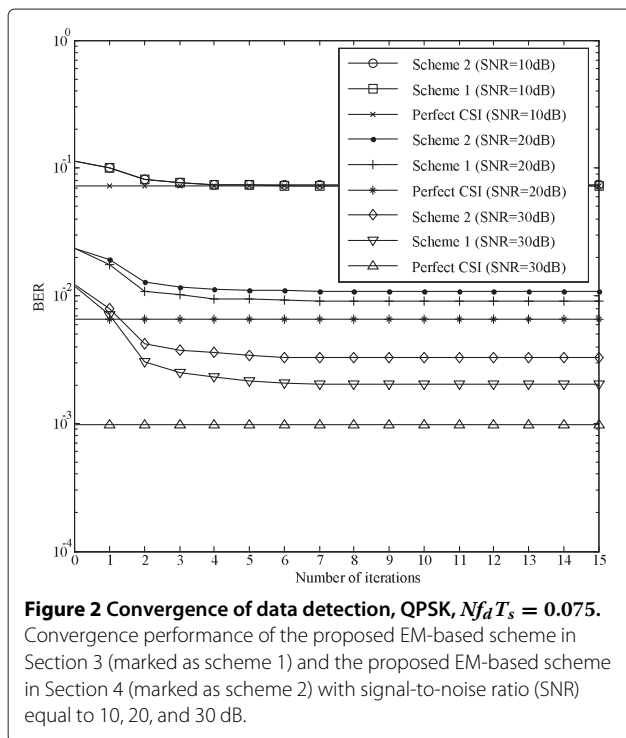
**Table 2 Computational complexity of the proposed scheme in Section 4**

Computation	Complexity
Matrix inversion in (56)	$O(((Q+1)L)^3)$
Eigendecomposition on (57)	$O(((Q+1)L)^3)$
Matrix inversion in (53)	$O(N^3)$



first few iterations and converge to stable values within eight iterations. Channel estimation with full training and data detection with perfect CSI are shown for comparison. Furthermore, according to [38], the Cramer-Rao bound is also shown for comparison. It can be seen

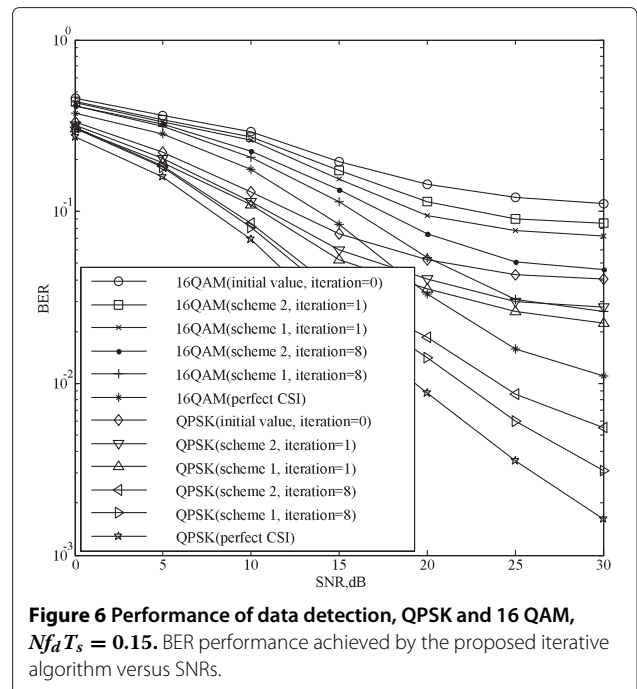
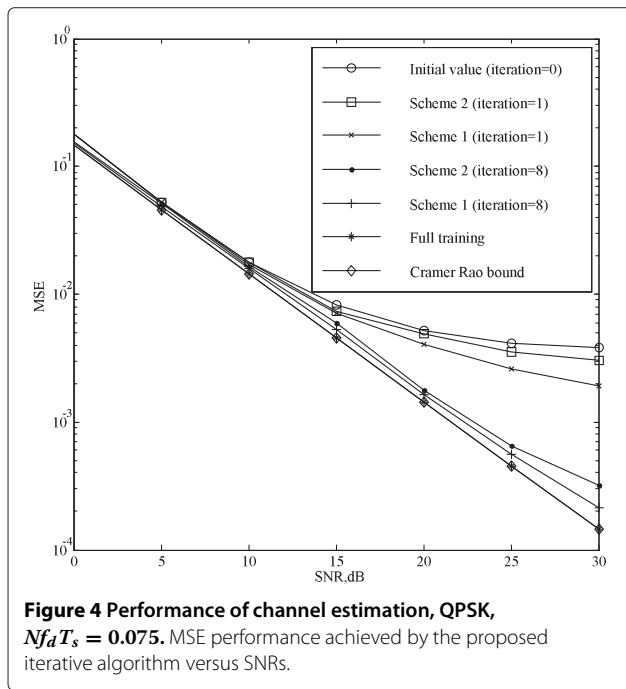
from Figure 1 that after convergence, the channel estimation performance of both schemes greatly improve that of the initial estimation (marked as iteration = 0), which indicates the ability of the proposed algorithm to cancel the interference from unknown data to channel estimation through iterations. The channel estimation performance of scheme 1 is very close to that of the Cramer-Rao bound and the full training case. The channel estimation performance of scheme 2 suffers a minor performance degradation compared to that of the scheme 1, which is the price we have to pay for the reduced computational complexity. Similar results can be observed for the performance of data detection in Figures 2 and 3, which indicates that the updated channel estimate can in turn greatly improve the data detection through iterations. Similar convergence results are also observed for the 16 QAM case, and figures are not presented here due to space limitations.



## 5.2 Performance of the proposed schemes

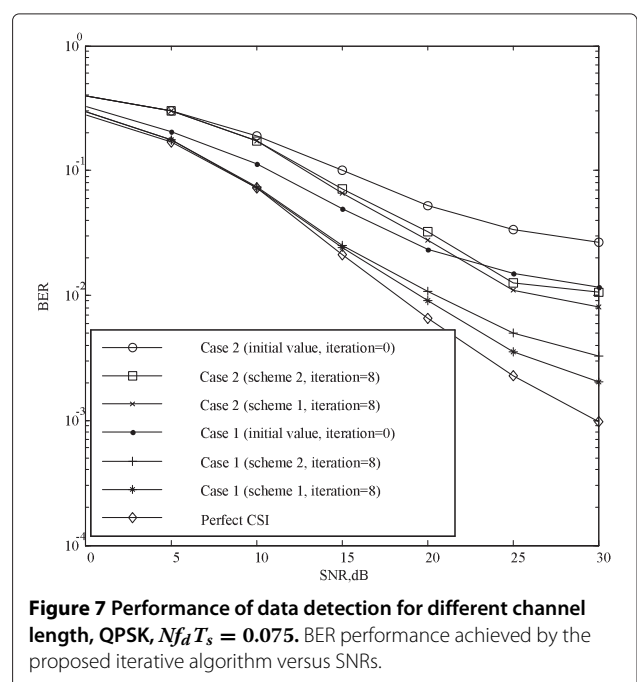
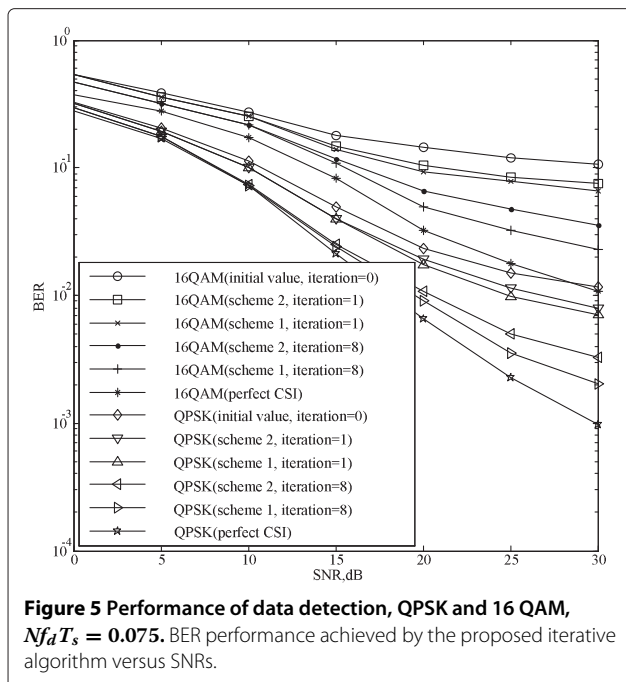
Figures 4, 5, and 6 show the MSE and BER performance achieved by the proposed iterative algorithm versus SNRs. It can be seen from Figure 4 that the performances of the proposed schemes 1 and 2 both perform much better than that of the initial value and close to that of the Cramer-Rao bound and the full training case after convergence.

Similarly, it can be seen from Figure 5 that for the case where  $Nf_d T_s = 0.075$ , the BER performance of the proposed iterative algorithm is very close to that of the ideal



case which assumes perfect CSI after convergence. For the severe case where  $Nf_d T_s = 0.15$ , it can be seen from Figure 6 that the proposed iterative algorithm can still deal with such a highly TF dispersive channel and performs well. Moreover, from Figures 5 and 6, it can be seen that for signals with both amplitude and phase variations such as 16 QAM, the proposed algorithm also performs well.

Finally, we investigate how the proposed schemes are affected by different channel lengths. A severe case where the channel length is equivalent to the number of embedded pilots (marked as case 2) is shown in Figure 7. As can be seen from the figure, compared to the originally-presented case where the channel length is 3 (marked as case 1), there is an obvious



performance degradation of the proposed schemes for the severe case 2. The reason can be explained according to the estimation theory [36] that when the channel length increases, more parameters need to be estimated, which leads to a decreased performance. On the contrary, if the channel length decreases, less parameters need to be estimated and that leads to an increased performance.

## 6 Conclusions

In this paper, two EM-based iterative data detection and channel and noise variance estimation schemes for MIMO-OFDM systems operating over TF dispersive channels under unknown background noise have been proposed. The resulting schemes achieve convergence in a few iterations and can effectively estimate TF dispersive channels and obtain reliable data detection under unknown background noise environments. The first scheme iteratively detects data and estimates the channel and noise variance simultaneously for all antennas, and moreover, the updating expressions of these estimates are all derived as closed-form results. Simulation results showed that after convergence, the performance of the first scheme is very close to that of the optimal case which assumes full training and perfect CSI. To reduce the computational complexity of the first scheme, another EM-based scheme that detecting data and estimating channel for only one antenna during each iteration and holding the unknown quantities of other antennas to their last estimates has been proposed, which is also derived as closed-form results. Simulation results showed that its performance only slightly degrades compared to the first scheme, but the computational complexity is significantly reduced.

## Appendices

### Appendix 1

#### Derivation of (17) and (18)

Using Bayes's formula, the conditional pdf of  $\beta$  is given by:

$$f(\beta | \mathbf{y}, \hat{\sigma}_{k-1}^2, \hat{\mathbf{s}}_{k-1}) = \frac{f(\mathbf{y} | \beta, \sigma^2, \mathbf{s}) f(\beta)}{f(\mathbf{y} | \sigma^2, \mathbf{s})} \Big|_{\sigma^2 = \hat{\sigma}_{k-1}^2, \mathbf{s} = \hat{\mathbf{s}}_{k-1}}, \quad (41)$$

where the fact that  $\beta$  is independent of  $\mathbf{s}$ , and  $\sigma^2$  has been used. The BEM coefficient  $\beta$  can be shown to be complex Gaussian variable [33] with zero mean and covariance matrix  $\mathbf{R}_\beta$ , that is:

$$f(\beta) = \frac{1}{\pi^{(Q+1)N_T N_R L} |\mathbf{R}_\beta|} \exp(-\beta^H \mathbf{R}_\beta^{-1} \beta). \quad (42)$$

Note that:

$$f(\mathbf{y} | \sigma^2, \mathbf{s}) = \int f(\mathbf{y} | \beta, \sigma^2, \mathbf{s}) f(\beta) d\beta. \quad (43)$$

With  $f(\mathbf{y} | \beta, \sigma^2, \mathbf{s})$  given by (13), putting (13) and (42) into (43), we have:

$$f(\mathbf{y} | \sigma^2, \mathbf{s}) = \frac{|\hat{\mathbf{Y}}_k|}{(\pi \sigma^2)^{NN_R} |\mathbf{R}_\beta|} \times \exp\left(-\frac{1}{\sigma^2} (\mathbf{y}^H \mathbf{y} - \sigma^2 \hat{\beta}_k^H \hat{\mathbf{Y}}_k^{-1} \hat{\beta}_k)\right). \quad (44)$$

Substituting (13), (42), and (44) into (41), after some manipulations we have:

$$f(\beta | \mathbf{y}, \hat{\sigma}_{k-1}^2, \hat{\mathbf{s}}_{k-1}) = \frac{1}{\pi^{(Q+1)N_T N_R L} |\hat{\mathbf{Y}}_k|} \times \exp(-(\beta - \hat{\beta}_k)^H \hat{\mathbf{Y}}_k^{-1} (\beta - \hat{\beta}_k)), \quad (45)$$

where

$$\hat{\beta}_k = (\Xi^H [\hat{\mathbf{s}}_{k-1}] \Xi [\hat{\mathbf{s}}_{k-1}] + \hat{\sigma}_{k-1}^2 \mathbf{R}_\beta^{-1})^{-1} \Xi^H [\hat{\mathbf{s}}_{k-1}] \mathbf{y}, \quad (46)$$

$$\hat{\mathbf{Y}}_k = \hat{\sigma}_{k-1}^2 (\Xi^H [\hat{\mathbf{s}}_{k-1}] \Xi [\hat{\mathbf{s}}_{k-1}] + \hat{\sigma}_{k-1}^2 \mathbf{R}_\beta^{-1})^{-1}. \quad (47)$$

Thus, the pdf  $f(\beta | \mathbf{y}, \hat{\sigma}_{k-1}^2, \hat{\mathbf{s}}_{k-1})$  is a Gaussian distribution. In addition,  $\hat{\beta}_k$  and  $\hat{\mathbf{Y}}_k$  given in (46) and (47), respectively, are in fact its conditional mean and covariance. To show that we have no prior information on  $\beta$ , we take the limit  $\|\mathbf{R}_\beta\| \rightarrow +\infty$ , which leads to (17) and (18). In this paper, we set  $\mathbf{R}_\beta^{-1}$  to zero to show we have no prior information for  $\beta$ . Indeed, there will be a performance degradation by assuming  $\mathbf{R}_\beta^{-1}$  to zero. However, this is a typical complexity versus performance trade-off. Moreover, as can be seen from simulation results in Section 5, even we set  $\mathbf{R}_\beta^{-1}$  to zero, the proposed algorithm also performs well, and its performance is acceptable.

### Appendix 2

#### Solving (36)

Comparing (34a) and (34b) with (10a) and (10b), referring to Section 3, we take  $\hat{\psi}_k^i$  as the incomplete data,  $\beta_{rc}^i$  as the unobservable or missing data, and  $(\sigma_{\chi^i}^2, \mathbf{s}^i)$  as parameters of interest. The two steps at the  $k$ th iteration are detailed as follows:

*E-step:* compute  $\mathbb{Q}(\sigma_{\chi^i}^2, \mathbf{s}^i | \hat{\sigma}_{\chi^i, k-1}^2, \hat{\mathbf{s}}_{k-1}^i) = \mathbb{E}\{\log f(\hat{\psi}_k^i, \beta_{rc}^i | \sigma_{\chi^i}^2, \mathbf{s}^i) | \hat{\psi}_k^i, \hat{\sigma}_{\chi^i, k-1}^2, \hat{\mathbf{s}}_{k-1}^i\}$ .

*M-step:* solve  $(\hat{\sigma}_{\chi^i, k}^2, \hat{\mathbf{s}}_k^i) = \arg \max_{\sigma_{\chi^i}^2, \mathbf{s}^i} \mathbb{Q}(\sigma_{\chi^i}^2, \mathbf{s}^i | \hat{\sigma}_{\chi^i, k-1}^2, \hat{\mathbf{s}}_{k-1}^i)$ .

Note that conditioned upon  $\hat{\boldsymbol{\psi}}_k^i$ , the only unknown or random component in the complete data  $(\hat{\boldsymbol{\psi}}_k^i, \boldsymbol{\beta}_{rc}^i)$  is  $\boldsymbol{\beta}_{rc}^i$ , the expectation is taken with respect to the conditional probability density function  $f(\boldsymbol{\beta}_{rc}^i | \hat{\boldsymbol{\psi}}_k^i, \hat{\sigma}_{\chi^i, k-1}^2, \hat{\mathbf{s}}_{k-1}^i)$ , while  $(\hat{\sigma}_{\chi^i, k}^2, \hat{\mathbf{s}}_k^i)$  are the estimates of  $\sigma_{\chi^i}^2$ , and  $\mathbf{s}^i$  at the  $k$ th iteration. More specifically, for the *E-step*: Using Bayes's rule, we obtain:

$$f(\hat{\boldsymbol{\psi}}_k^i, \boldsymbol{\beta}_{rc}^i | \sigma_{\chi^i}^2, \mathbf{s}^i) = f(\hat{\boldsymbol{\psi}}_k^i | \boldsymbol{\beta}_{rc}^i, \sigma_{\chi^i}^2, \mathbf{s}^i) f(\boldsymbol{\beta}_{rc}^i). \quad (48)$$

Using (48), the function  $\mathbb{Q}(\sigma_{\chi^i}^2, \mathbf{s}^i | \hat{\sigma}_{\chi^i, k-1}^2, \hat{\mathbf{s}}_{k-1}^i)$  can be expressed as:

$$\begin{aligned} \mathbb{Q}(\sigma_{\chi^i}^2, \mathbf{s}^i | \hat{\sigma}_{\chi^i, k-1}^2, \hat{\mathbf{s}}_{k-1}^i) &= \mathbb{E}\{\log f(\hat{\boldsymbol{\psi}}_k^i | \boldsymbol{\beta}_{rc}^i, \sigma_{\chi^i}^2, \mathbf{s}^i) | \hat{\boldsymbol{\psi}}_k^i, \hat{\sigma}_{\chi^i, k-1}^2, \hat{\mathbf{s}}_{k-1}^i\} \\ &\quad + \mathbb{E}\{\log f(\boldsymbol{\beta}_{rc}^i | \hat{\boldsymbol{\psi}}_k^i, \hat{\sigma}_{\chi^i, k-1}^2, \hat{\mathbf{s}}_{k-1}^i)\}, \end{aligned} \quad (49)$$

where the second term can be ignored in the following derivations, since it is not a function of parameters of interest, i.e., not a function of  $(\sigma_{\chi^i}^2, \mathbf{s}^i)$ . Using (34a), the likelihood function  $f(\hat{\boldsymbol{\psi}}_k^i | \boldsymbol{\beta}_{rc}^i, \sigma_{\chi^i}^2, \mathbf{s}^i)$  is obtained as:

$$\begin{aligned} f(\hat{\boldsymbol{\psi}}_k^i | \boldsymbol{\beta}_{rc}^i, \sigma_{\chi^i}^2, \mathbf{s}^i) &= \frac{1}{(\pi \sigma_{\chi^i}^2)^{NN_R}} \\ &\quad \times \exp\left(-\frac{1}{\sigma_{\chi^i}^2} (\hat{\boldsymbol{\psi}}_k^i - \Phi[\mathbf{s}^i] (\mathbf{I}_{N_{RL}} \otimes \mathbf{B}) \boldsymbol{\beta}_{rc}^i)^H \right. \\ &\quad \left. \times (\hat{\boldsymbol{\psi}}_k^i - \Phi[\mathbf{s}^i] (\mathbf{I}_{N_{RL}} \otimes \mathbf{B}) \boldsymbol{\beta}_{rc}^i)\right). \end{aligned} \quad (50)$$

Substituting (50) into (49) and referring to (14), (15) and (16) and Appendix 1, the conditional mean and covariance matrix are obtained as:

$$\hat{\boldsymbol{\beta}}_{rc, k}^i = ((\mathbf{I}_{N_{RL}} \otimes \mathbf{B})^H \Phi^H[\hat{\mathbf{s}}_{k-1}^i] \Phi[\hat{\mathbf{s}}_{k-1}^i] (\mathbf{I}_{N_{RL}} \otimes \mathbf{B}))^{-1} \times (\mathbf{I}_{N_{RL}} \otimes \mathbf{B})^H \Phi^H[\hat{\mathbf{s}}_{k-1}^i] \hat{\boldsymbol{\psi}}_k^i, \quad (51)$$

$$\hat{\boldsymbol{\Upsilon}}_{rc, k}^i = \hat{\sigma}_{\chi^i, k-1}^2 ((\mathbf{I}_{N_{RL}} \otimes \mathbf{B})^H \Phi^H[\hat{\mathbf{s}}_{k-1}^i] \Phi[\hat{\mathbf{s}}_{k-1}^i] (\mathbf{I}_{N_{RL}} \otimes \mathbf{B}))^{-1}. \quad (52)$$

It is noted that the matrix  $\Phi[\mathbf{s}^i]$  is of dimension  $N_R N \times N_R N L$ , and the matrix  $(\mathbf{I}_{N_{RL}} \otimes \mathbf{B})$  is of dimension  $N_R N L \times N_R (Q+1)L$ ; the  $N_R (Q+1)L \times N_R (Q+1)L$  matrix inversion required in (51) and (52) is only  $\frac{1}{N_T}$  of that needed in (17) and (18).

*M-step*: using the two equivalent expressions derived in (34a) and (34b) and similar to (19), (20), (21) and (22), the signal updating equation is obtained as:

$$\begin{aligned} \hat{\mathbf{s}}_k^i &= \left( \sum_{m=0}^{N_R(Q+1)L-1} \lambda_{rc, m, k}^i \mathbf{H}^H[\boldsymbol{\mu}_{rc, m, k}^i] \mathbf{H}[\boldsymbol{\mu}_{rc, m, k}^i] \right. \\ &\quad \left. + \mathbf{H}^H[\hat{\boldsymbol{\beta}}_{rc, k}^i] \mathbf{H}[\hat{\boldsymbol{\beta}}_{rc, k}^i] \right)^{-1} \mathbf{H}^H[\hat{\boldsymbol{\beta}}_{rc, k}^i] \hat{\boldsymbol{\psi}}_k^i, \end{aligned} \quad (53)$$

where  $\hat{\boldsymbol{\Upsilon}}_{rc, k}^i = \sum_{m=0}^{N_R(Q+1)L-1} \lambda_{rc, m, k}^i \boldsymbol{\mu}_{rc, m, k}^i (\boldsymbol{\mu}_{rc, m, k}^i)^H$  represents the eigendecomposition of  $\hat{\boldsymbol{\Upsilon}}_{rc, k}^i$ . It is noted that compared to (21) where  $N_T N \times N_T N$  matrix inversion is required, only  $N \times N$  matrix inversion is needed in (53). The symbol detection can thus be obtained after OFDM demodulation as

$$\hat{\mathbf{x}}_k^i = \text{Quant}\{\mathbf{F} \hat{\mathbf{s}}_k^i\}. \quad (54)$$

Substituting  $\hat{\mathbf{s}}_k^i = \mathbf{F}^H \hat{\mathbf{x}}_k^i$  and (50), (51) and (52) into (49) and referring to (23), the unknown noise variance for the disturbance term  $\chi_k^i$  can be obtained as:

$$\begin{aligned} \hat{\sigma}_{\chi^i, k}^2 &= \frac{1}{NN_R} \left( (\hat{\boldsymbol{\psi}}_k^i)^H \hat{\boldsymbol{\psi}}_k^i - 2 \Re\{(\hat{\boldsymbol{\psi}}_k^i)^H \Phi[\hat{\mathbf{s}}_k^i] (\mathbf{I}_{N_{RL}} \otimes \mathbf{B}) \hat{\boldsymbol{\beta}}_{rc, k}^i\} \right. \\ &\quad \left. + \text{Tr}\{(\mathbf{I}_{N_{RL}} \otimes \mathbf{B})^H \Phi^H[\hat{\mathbf{s}}_k^i] \Phi[\hat{\mathbf{s}}_k^i] (\mathbf{I}_{N_{RL}} \otimes \mathbf{B})\} \right. \\ &\quad \left. \times (\hat{\boldsymbol{\Upsilon}}_{rc, k}^i + \hat{\boldsymbol{\beta}}_{rc, k}^i (\hat{\boldsymbol{\beta}}_{rc, k}^i)^H) \right). \end{aligned} \quad (55)$$

In summary, (51), (52), (53), (54) and (55) solve the minimization problem in (36).

Notice that the computational complexity can be further reduced by observing the diagonal structure of both  $\Phi[\mathbf{s}^i]$  and  $(\mathbf{I}_{N_{RL}} \otimes \mathbf{B})$  in (24a). Therefore, (51) and (52) can be further split into  $N_R$  sub-matrices, each of which is expressed as:

$$\hat{\boldsymbol{\beta}}_{rc, k}^{z, i} = ((\mathbf{B}^z)^H \mathbf{G}^H[\hat{\mathbf{s}}_{k-1}^i] \mathbf{G}[\hat{\mathbf{s}}_{k-1}^i] \mathbf{B}^z)^{-1} (\mathbf{B}^z)^H \mathbf{G}^H[\hat{\mathbf{s}}_{k-1}^i] \hat{\boldsymbol{\psi}}_k^i, \quad (56)$$

$$\boldsymbol{\Upsilon}_{rc, k}^{z, i} = \hat{\sigma}_{\chi^i, k-1}^2 ((\mathbf{B}^z)^H \mathbf{G}^H[\hat{\mathbf{s}}_{k-1}^i] \mathbf{G}[\hat{\mathbf{s}}_{k-1}^i] \mathbf{B}^z)^{-1}, \quad (57)$$

where  $\hat{\boldsymbol{\psi}}_k^i = [(\hat{\boldsymbol{\psi}}_k^{0, i})^T, (\hat{\boldsymbol{\psi}}_k^{1, i})^T, \dots, (\hat{\boldsymbol{\psi}}_k^{N_R-1, i})^T]^T$ . Then,  $\hat{\boldsymbol{\beta}}_{rc, k}^i$  is obtained as  $\hat{\boldsymbol{\beta}}_{rc, k}^i = [(\hat{\boldsymbol{\beta}}_{rc, k}^{0, i})^T, (\hat{\boldsymbol{\beta}}_{rc, k}^{1, i})^T, \dots, (\hat{\boldsymbol{\beta}}_{rc, k}^{N_R-1, i})^T]^T$ , and  $\hat{\boldsymbol{\Upsilon}}_{rc, k}^i$  can be obtained as  $\hat{\boldsymbol{\Upsilon}}_{rc, k}^i = \text{Blkdiag}\{\boldsymbol{\Upsilon}_{rc, k}^{0, i}, \boldsymbol{\Upsilon}_{rc, k}^{1, i}, \dots, \boldsymbol{\Upsilon}_{rc, k}^{N_R-1, i}\}$ . It is noted that the matrix  $\mathbf{G}[\mathbf{s}^i]$  is of dimension  $N \times NL$ , and the matrix  $\mathbf{B}^z \triangleq \mathbf{I}_L \otimes \mathbf{B}$  is of dimension  $NL \times (Q+1)L$ ; the  $(Q+1)L \times (Q+1)L$  matrix inversion required in (56) and (57) is  $\frac{1}{N_R}$  of that needed in (51) and (52) and therefore only  $\frac{1}{N_R N_T}$  of that needed in (17) and (18).

### Competing interests

The authors declare that they have no competing interests.

### Acknowledgements

This work was supported in part by the National Science Foundation of China under grant number 61032002, 60902026, and 60972029, the Chinese Important National Science & Technology Specific Projects under grant 2011ZX03001-007-01, and the Program for New Century Excellent Talents in University, NCET-11-0058.

Received: 10 December 2012 Accepted: 15 June 2013

Published: 6 July 2013

### References

1. H Bocksteal, AJ Paulraj, *Multiple-Input Multiple-Output (MIMO) Wireless Systems* (Cambridge Univ. Press, Cambridge, 2003), pp. 1–22
2. RV Nee, R Prasad, *OFDM for Wireless Multimedia Communications* (Artech House Publishers, Norwood, 2000), pp. 1–284
3. L Hanzo, J Akhtman, M Jiang, L Wang, *MIMO-OFDM for LTE, WiFi and WiMAX: Coherent versus Non-coherent and Cooperative Turbo Transceivers* (Wiley, Hoboken, 2010), pp. 1–692
4. A Goldsmith, *Wireless Communications* (Cambridge Univ. Press, Cambridge, 2005), pp. 1–672
5. AJ Viterbi, JK Omura, *Principles of Digital Communication and Coding* (Dover Press, New York, 2009), pp. 1–576
6. AP Dempster, NM Laird, DB Rubin, Maximum likelihood from incomplete data via the EM algorithm. *J. Royal Statistical Soc., Ser. B (Methodological)*. **39**(1), 1–38 (1977)
7. T Moon, The expectation-maximization algorithm. *IEEE Signal Process. Mag.* **13**(6), 47–60 (1996)
8. A Assra, W Hamouda, A Youssef, EM-based joint channel estimation and data detection for MIMO-CDMA systems. *IEEE Trans. Veh. Technol.* **59**(3), 1205–1216 (2010)
9. J Choi, An EM based joint data detection and channel estimation incorporating with initial channel estimate. *IEEE Commun. Lett.* **12**(9), 654–656 (2008)
10. C Cozzo, BL Hughes, Joint channel estimation and data detection in space-time communications. *IEEE Trans. Commun.* **51**(8), 1266–1270 (2003)
11. X Zhang Y, DG Wang, JB Wei, in *IEEE Wireless Communications and Networking Conference*, vol. 3. Joint symbol detection and channel estimation for MIMO-OFDM systems via the variational bayesian EM algorithm (Las Vegas, 31 Mar–3 Apr 2008), pp. 13–17
12. DK So, RS Chen, Iterative EM receiver for space-time coded systems in MIMO frequency-selective fading channels with channel gain and order estimation. *IEEE Trans. Wirel. Commun.* **3**(6), 1928–1935 (2004)
13. B Lu, X Wang, Y Li, Iterative receivers for space-time block coded OFDM systems in dispersive fading channels. *IEEE Trans. Wirel. Commun.* **1**(2), 213–225 (2002)
14. A Zia, JPR Reilly, J Manton, S Shiran, An information geometry approach to ML estimation with incomplete data: application to semiblind MIMO channel identification. *IEEE Trans. Signal Process.* **55**(8), 3975–3985 (2007)
15. MA Khalighi, JJ Boutros, Semi-blind channel estimation using EM algorithm in iterative MIMO APP detectors. *IEEE Trans. Wirel. Commun.* **5**(11), 3165–3173 (2006)
16. CH Aldana, E de Cardealho, J Ciof, Channel estimation for multicarrier multiple input single output systems using the EM algorithm. *IEEE Trans. Signal Process.* **51**(12), 3280–3292 (2003)
17. J Zhang, L Hanzo, X Mu, Joint decision-directed channel and noise-variance estimation for MIMO OFDM/SDMA systems based on expectation-conditional maximization. *IEEE Trans. Veh. Technol.* **60**(5), 2139–2151 (2011)
18. J Choi, An EM-based iterative receiver for MIMO-OFDM under interference-limited environments. *IEEE Trans. Wirel. Commun.* **6**(11), 3994–4003 (2007)
19. X Wautelet, C Herzet, A Dejonghe, J Louveaux, L Vandendorpe, Comparison of EM-based algorithms for MIMO channel estimation. *IEEE Trans. Commun.* **55**(1), 216–226 (2007)
20. I Nevat, GW Peters, J Yuan, Detection of gaussian constellations in MIMO systems under imperfect CSI. *IEEE Trans. Commun.* **58**(4), 1151–1160 (2010)
21. C Georghiades, J Han, Sequence estimation in the presence of random parameters via the EM algorithm. *IEEE Trans. Commun.* **45**(3), 300–308 (1997)
22. F Chan, J Choi, Neighborhood exploring detector: an EM-based signal detector for multiple antenna systems. *IEEE Trans. Signal Process.* **55**(5), 1875–1885 (2007)
23. T Kashima, K Fukawa, H Suzuki, Adaptive MAP receiver via the EM algorithm and message passings for MIMO-OFDM mobile communications. *IEEE J. Sel. Areas Commun.* **24**(3), 437–447 (2006)
24. Y-L Ueng, Y-M Chen, J-Y Lin, A MIMO-BICM scheme using a convolutional interleaver for delay-sensitive applications. *IEEE Trans. Veh. Technol.* **59**(5), 2380–2393 (2010)
25. M Khalighi, S Bourennane, Semiblind single-carrier MIMO channel estimation using overlay pilots. *IEEE Trans. Veh. Technol.* **57**(3), 951–956 (2008)
26. J Choi, MIMO-BICM iterative receiver with the EM based channel estimation and simplified MMSE combining with soft cancellation. *IEEE Trans. Signal Process.* **54**(8), 3247–3251 (2006)
27. J Zheng, B Rao, LDPC-coded MIMO systems with unknown block fading channels: soft MIMO detector design, channel estimation, and code optimization. *IEEE Trans. Signal Process.* **54**(4), 1504–1518 (2006)
28. MA Khalighi, J Boutros, J-F H elard, Data-aided channel estimation for turbo-PIC MIMO detectors. *IEEE Commun. Lett.* **10**(5), 350–352 (2006)
29. T-H Pham, Y-C Liang, A Nallanathan, A joint channel estimation and data detection receiver for multiuser MIMO IFDMA systems. *IEEE Trans. Commun.* **57**(6), 1857–1865 (2009)
30. J Gao, H Li, Low-complexity MAP channel estimation for mobile MIMO-OFDM systems. *IEEE Trans. Wirel. Commun.* **7**(3), 774–780 (2008)
31. RD Souza, J Garcia-Frias, A Haimovich M, Semiblind EM based iterative receivers for space-time-coded modulation and quasi-static frequency-selective fading channels. *IEEE Trans. Veh. Technol.* **55**(4), 1259–1268 (2006)
32. YZ Xie, CN Georghiades, Two EM-type channel estimation algorithms for OFDM with transmitter diversity. *IEEE Trans. Commun.* **51**(1), 106–115 (2003)
33. X Ma, G Giannakis B, S Ohno, Optimal training for block transmissions over doubly selective wireless fading channels. *IEEE Trans. Signal Process.* **51**(5), 1351–1366 (2003)
34. Z Tang, G Leus, RC Cannizzaro, P Banelli, Pilot-assisted timevarying channel estimation for OFDM systems. *IEEE Trans. Signal Process.* **55**(5), 2226–2238 (2007)
35. H Stark, JW Woods, *Probability and Random Processes with Applications to Signal Processing* Prentice Hall (Prentice-Hall, Upper Saddle River, 2002), pp. 1–699
36. SM Kay, *Fundamental of Statistical Signal Processing: Estimation Theory* (Prentice-Hall, Upper Saddle River, 1993), pp. 1–625
37. A Kannu, P Schniter, Design and analysis of MMSE pilot-aided cyclic-prefixed block transmissions for doubly selective channels. *IEEE Trans. Signal Process.* **56**(3), 1148–1160 (2008)
38. H Tree, K Bell, *Bayesian Bounds for Parameter Estimation and Nonlinear Filtering/Tracking* (Wiley-IEEE Press, New York, 2007), pp. 1–951

doi:10.1186/1687-1499-2013-182

Cite this article as: Zhong et al.: Iterative channel estimation and data detection for MIMO-OFDM systems operating in time-frequency dispersive channels under unknown background noise. *EURASIP Journal on Wireless Communications and Networking* 2013 **2013**:182.