Expectation-Maximization-based Channel Estimation for Multiuser MIMO Systems

Sunho Park*, Jun Won Choi†, Ji-Yun Seol‡, and Byonghyo Shim*

*Institute of New Media and Communications and School of Electrical and Computer Engineering, Seoul National University, Seoul, Korea

†Electrical-Bio Engineering Department, Hanyang University, Seoul, Korea

‡Communications Research Team, DMC R&D Center, Samsung Electronics Co., Ltd., Suwon, Korea

Abstract

Multiuser multiple-input multiple-output (MU-MIMO) transmission techniques have been popularly used to improve the spectral efficiency and user experience. However, because of the coarse knowledge of channel state information at the transmitter (CSIT), the quality of transmit precoding to control multiuser interference is degraded and hence co-scheduled user equipment (UE) may suffer from large residual multiuser interference. In this paper, we propose a new channel estimation technique employing reliable soft symbols to improve the channel estimation and subsequent detection quality of MU-MIMO systems. To this end, we pick a small number of reliable data tones from both desired and interfering users and then use them as pilots to re-estimate the channel. In order to estimate the channel and data symbols jointly, we employ the expectation maximization (EM) algorithm where the channel estimation and data decoding are performed iteratively. From numerical experiments in realistic MU-MIMO scenarios, we show that the proposed method achieves substantial performance gain in channel estimation and detection quality over conventional channel estimation approaches.

This work was sponsored by Communications Research Team, DMC R&D Center, Samsung Electronics Co. Ltd, the MSIP (Ministry of Science, ICT&Future Planning), Korea in the ICT R&D Program 2013 (No. 1291101110-130010100), and the National Research Foundation of Korea (NRF) grant funded by the Korean government (MSIP) (2014R1A5A1011478).

This paper was presented in part at the Personal indoor mobile radio communications (PIMRC) symposium, 2015 [1] and Vehicular Technology Conference (VTC), 2016 [2].

Expectation-Maximization-based Channel Estimation for Multiuser MIMO Systems

I. Introduction

In most wireless systems, multiuser multiple-input multiple-output (MIMO) techniques have been used to improve the spectral efficiency and user experience [3]. In contrast to the traditional single user MIMO (SU-MIMO) systems where the time-frequency resource element is dedicated to a single user, multiuser MIMO system allows multiple users to use the same time-frequency resources via a proper control of the interference among co-scheduled users at the base station. Control of this multiuser interference is achieved by applying precoding to the symbol vectors of all users scheduled in the same time-frequency resources. Since the precoding matrix is generated using the downlink channel state information (CSI) which relies on the feedback information from the mobile users, inaccurate precoding operation from imperfect CSI causes a severe degradation in multiuser interference cancellation at the transmitter side and the channel estimation and detection at the receiver side, undermining the benefits of multiuser MIMO in the end.

In order to mitigate the degradation of channel estimation quality caused by multiuser interference, dedicated pilots (e.g., demodulation reference signal (DM-RS) has been introduced in the Long Term Evolution Advanced (LTE-Advanced) standard [4]). While the purpose of common pilots is to serve all users in the cell, dedicated pilots are literally dedicated to a single or a group of users for the better estimation of channels. The precoding matrix applied to the dedicated pilots is the same as that applied to the information vector and hence each user estimates the precoded channel which is the compound of the precoding matrix and the (transfer function of) physical channels.

As the wireless cellular industry is moving fast towards the fifth generation (5G) communication systems, an attempt to use more number of antennas at the base station than that of current systems has received much attention recently. For example, LTE-Advanced-Pro, the recent standard of 3rd Generation Partnership Project (3GPP) LTE, considers using up to 64 antennas at the base station [5]. In this scenario, owing to the increased spatial degrees of freedom, multiuser MIMO systems can accommodate tens of users in the same time-frequency resource.

Since the pilot symbols should be assigned for all users, significant amount of radio resources would be consumed for the pilot transmission. Since such large pilot overhead encroaches on the data resources and therefore obstructs the throughput enhancement, one needs to consider the reduction of pilot overhead. However, reduction of pilot density will cause severe degradation in the channel estimation quality and eventual loss of the system performance. Therefore, an approach to enhance the channel estimation quality without increasing pilot overhead is desired.

There have been some studies to improve the channel estimation quality of multiuser MIMO systems [6]–[12]. In [6], multiuser interferences are suppressed using the channel estimates predicted from adjacent symbols. Using the interference suppressed signals, tentative decisions are performed on data symbols. In [7], a joint detection algorithm for multiuser environment has been suggested. In this work, effect of the residual interference was considered to improve the performance of joint demodulation and decoding. Also, channel estimation (CE) techniques accounting for the effect of multiuser interferences have been proposed. Notable examples include the maximum a posteriori (MAP)-based CE [8], joint maximum likelihood CE and detection [9], Kalman filter-based soft decision CE [10], and the CE combined with interference suppression [11], [12].

An aim of this paper is to propose an improved channel estimation technique for multiuser MIMO systems. The proposed method exploits the channel information at the data tones to improve the channel estimation and subsequent detection quality. Towards this end, we deliberately choose reliable data tones and then use them for *virtual* pilots to generate the refined channel estimates. Our framework is based on the expectation maximization (EM) algorithm [13], where the channel estimation and data decoding are performed iteratively to generate the joint estimate of the channel and data symbols. The EM algorithm is computationally effective means to solve the maximum likelihood (ML) and maximum a posteriori (MAP) estimation problems [14]. There have been some studies using the EM technique for the channel estimation purposes [15]–[17]. In [15], EM-based channel estimation has been proposed for the frequency-selective channel environment with inter-symbol interference. In [16], an iterative receiver technique using the EM algorithm was proposed for the single user MIMO system. In [17], EM-based channel estimator performing the interference suppression using the sample covariance of the received signal has been proposed.

The proposed method is distinct from previous efforts in the following two aspects. First, we modify the original EM algorithm such that the soft information delivered from the channel

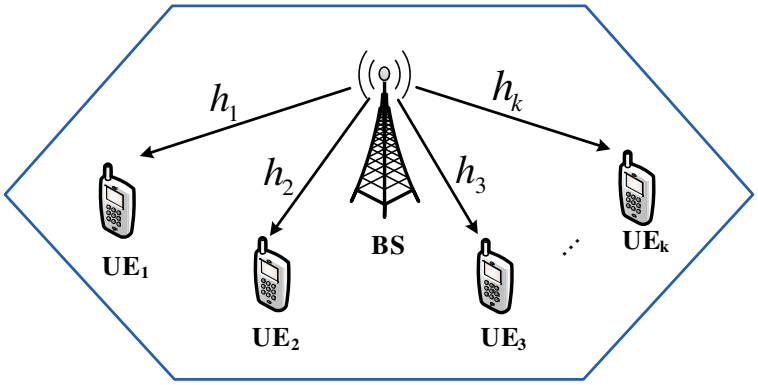


Fig. 1. Illustration of multiuser MIMO-OFDM system

decoder is incorporated into the cost metric in the E-step. Since the soft information generated from the channel decoder is more accurate than the output of the MIMO detector, we employ the feedback from the channel decoder as *prior* information of the data symbols. We observe from numerical evaluations that use of feedback from the channel decoder improves both the convergence speed and the quality of channel estimation. Second, in order to reduce the computational complexity associated with the virtual pilot selection, we choose a small group of reliable data tones making a dominant contribution to the channel estimation quality. To do so, we design a mean square error (MSE) based data tone selection strategy. We show from numerical simulations in LTE-Advanced and LTE-Advanced-Pro scenarios that the proposed channel estimator outperforms conventional MMSE scheme and existing EM-based channel estimation technique, especially in the scenario where the present multiuser MIMO systems fails to operate due to the insufficient pilot resources and the inaccurate precoding operation.

The rest of this paper is organized as follows. In Section II, we provide the brief summary of the multiuser MIMO-orthogonal frequency division multiplex (OFDM) systems and also review the conventional channel estimation schemes. In Section III, we present the proposed EM-based joint pilot and data tone channel estimation techniques and the reliable data symbol selection criterion. Section IV, we provide the simulation results and conclude the paper in Section VI.

II. System description

A. Multiuser MIMO-OFDM Systems

In this subsection, we briefly describe the basic system model of OFDM-based multiuser MIMO systems. Fig. 1 depicts a multiuser MIMO system where the base station equipped

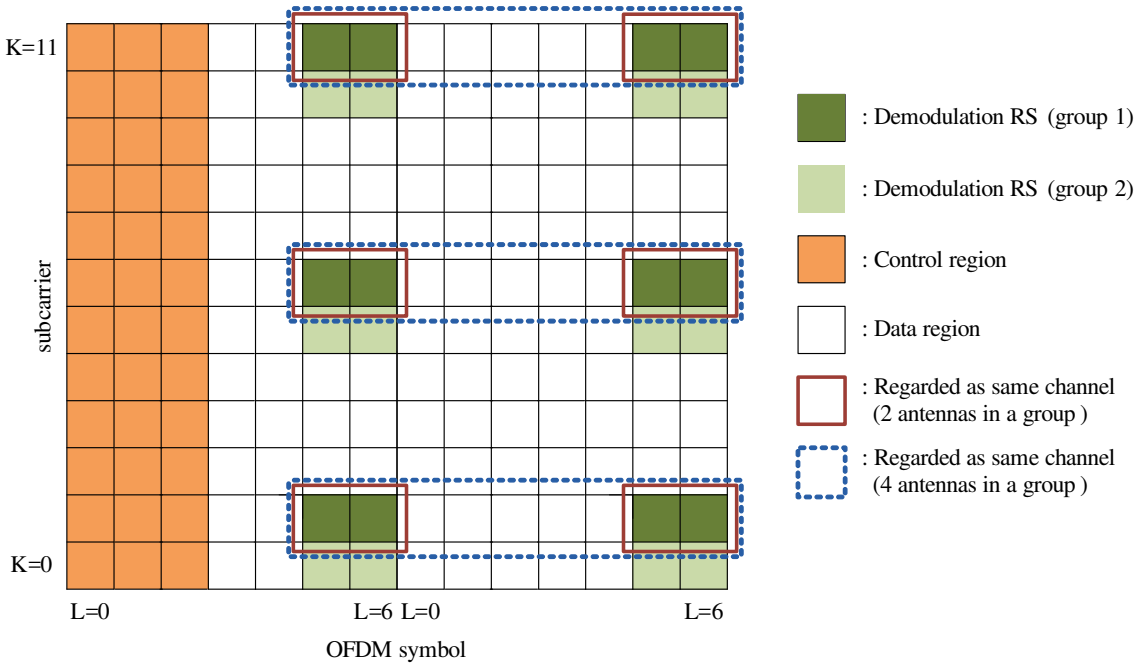


Fig. 2. Example of window for channel estimation in multiuser MIMO system

with N_T antennas transmits data to M user equipments (UEs) with a single receive antenna. The information bits $\{b_i\}$ for each user are encoded by a rate C channel encoder, producing the coded bit sequence $\{c_i\}$. The sequence $\{c_i\}$ is permuted using a random interleaver and Q interleaved bits are mapped to a finite alphabet symbol $\{x_i\}$ in 2^Q quadrature amplitude modulation (QAM) constellations. Both the pilot symbols and data symbols are allocated in two-dimensional resource grid indexed by the frequency index k and OFDM symbol index l (see Fig. 2). The subcarriers on the resource grid are divided into pilot and data tones. At the (k,l)th pilot tone, the pilot symbols for M users, $[p_{k,l}^{(0)} \quad \cdots \quad p_{k,l}^{(M-1)}]^T$ are allocated over the M layers. In a similar manner, M data symbols $[x_{k',l'}^{(0)} \quad \cdots \quad x_{k',l'}^{(M-1)}]^T$ are assigned to the (k',l')th data tone. We consider a time-frequency resource block (RB) occupying K frequency subcarriers and L OFDM symbols (see the example of LTE-A systems with $K = 12$ and $L = 7$ in Fig. 2) [18], [19]. Among them, there are N_p pilot tones and $N_d'(= KL - N_p)$ data tones in the processing block. We assume a block fading channel where the channel gain remains constant within a processing block. Each of M users receives the pilot sequence with length N_p in the designated pilot tones. The channels of different users are distinguished by the orthogonal pilot sequences referred to as *orthogonal cover codes* (OCC) (see Fig. 3).

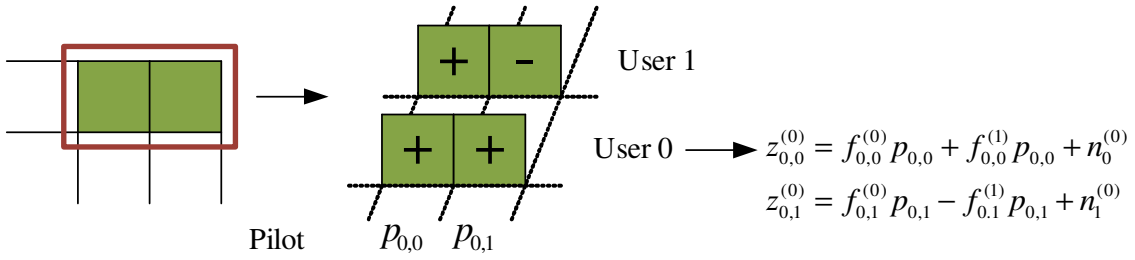


Fig. 3. Example of orthogonal cover codes (OCC) for two UEs

The $N_T \times M$ precoding matrix $\mathbf{W}$ to control the multiuser interference is multiplied to both $[p_{k,l}^{(0)} \quad \cdots \quad p_{k,l}^{(M-1)}]^T$ and $[x_{k',l'}^{(0)} \quad \cdots \quad x_{k',l'}^{(M-1)}]^T$, generating the N_T streams of frequency-domain OFDM symbols. These symbols are transformed into time-domain signals via the inverse discrete Fourier transform (IDFT). After the addition of the cyclic prefix (CP), a time-domain signal is transmitted by the N_T transmit antennas over the frequency selective channels.

On the user side, after the removal of the CP, the received signal is transformed back to the frequency-domain OFDM symbols via the DFT operation. Assuming that the channel delay spread is less than the CP length, the received signal $z_{k,l}^{(m)}$ at the *pilot tone* is given by

$$z_{k,l}^{(m)} = \left(\mathbf{h}_{k,l}^{(m)}\right)^H \mathbf{w}_m p_{k,l}^{(m)} + \left(\mathbf{h}_{k,l}^{(m)}\right)^H \sum_{j=0, j\neq m}^{M-1} \mathbf{w}_j p_{k,l}^{(m)} + n_{k,l}^{(m)} \tag{1}$$

$$= f_{k,l}^{(m)} p_{k,l}^{(m)} + \sum_{j=0, j\neq m}^{M-1} f_{k,l}^{(j)} p_{k,l}^{(j)} + n_{k,l}^{(m)}, \tag{2}$$

where (k,l) is the resource index used for pilot tones, $f_{k,l}^{(j)} = \left(\mathbf{h}_{k,l}^{(m)}\right)^H \mathbf{w}_j$, $\mathbf{h}_{k,l}^{(m)}$ is the $N_T \times 1$ vector of the channel gains from the base station to the mth user, $n_{k,l}^{(m)} (\sim \mathcal{CN}(0,1))$ is the complex Gaussian noise, and $\mathbf{w}_m$ is the mth column of the precoding matrix $\mathbf{W}$.

In a similar way, the received signal at the *data tone* can be expressed as

$$y_{k',l'}^{(m)} = \left(\mathbf{h}_{k',l'}^{(m)}\right)^H \mathbf{w}_m x_{k',l'}^{(m)} + \left(\mathbf{h}_{k',l'}^{(m)}\right)^H \sum_{j=0, j\neq m}^{M-1} \mathbf{w}_j x_{k',l'}^{(j)} + n_{k',l'}^{(m)}, \tag{3}$$

$$= g_{k',l'}^{(m)} x_{k',l'}^{(m)} + \sum_{j=0, j\neq m}^{M-1} g_{k',l'}^{(j)} x_{k',l'}^{(j)} + n_{k',l'}^{(m)}, \tag{4}$$

where (k',l') is the resource index used for data tones and $g_{k',l'}^{(j)} = \left(\mathbf{h}_{k',l'}^{(m)}\right)^H \mathbf{w}_j$. Note that the first term $g_{k,l}^{(m)} x_{k,l}^{(m)}$ in (4) is the desired signal of the mth user and the rest is the residual

interference plus noise. Both pilot and data symbols are multiplied by the same precoding matrix $\mathbf{W}$. Hence, the effective channel gain at the pilot and data tones includes the transfer function of the precoding matrix. Under the assumption that the residual interference plus noise is Gaussian, the log-likelihood ratio (LLR) of the coded bits c_i (i.e., the ith coded bit of the data symbol $x_{k,l}^{(m)}$) can be expressed as [20]

$$L\left(c_i\right) = \underbrace{\ln\frac{P(c_{i+})}{P(c_{i-})}}_{\textit{a priori}} + \underbrace{\ln\frac{\sum_{x_{k,l,+}^{(m)}} \exp\left(-\frac{\|y_{k,l}^{(m)}-\hat{g}_{k,l}^{(m)}x_{k,l}^{(m)}\|^2}{2\sigma_n^2}+\frac{1}{2}\mathbf{c}_{\bar{i}}^T\mathbf{L}_{A,\bar{i}}\right)}{\sum_{x_{k,l,-}^{(m)}} \exp\left(-\frac{\|y_{k,l}^{(m)}-\hat{g}_{k,l}^{(m)}x_{k,l}^{(m)}\|^2}{2\sigma_n^2}+\frac{1}{2}\mathbf{c}_{\bar{i}}^T\mathbf{L}_{A,\bar{i}}\right)}}_{\text{extrinsic}}, \tag{5}$$

where $\hat{g}_{k,l}^{(m)}$ is the estimate of the channel gain $g_{k,l}^{(m)}$, $P(c_{i+}) = P(c_i = +1)$ and $P(c_{i-}) = P(c_i = -1)$, $\mathbf{c}_{\bar{i}} = [c_1 \quad \cdots \quad c_{i-1} \quad c_{i+1} \quad \cdots]^T$, $\mathbf{L}_{A,\bar{i}} = [L_A(c_1) \quad \cdots \quad L_A(c_{i-1}) \quad L_A(c_{i+1}) \quad \cdots]^T$, $L_A(c_i) = \ln\frac{P(c_{i+})}{P(c_{i-})}$, and σ_n^2 is the variance of the residual interference plus noise. In the detection process, soft symbols in a form of LLRs are generated and then delivered to the channel decoder. When the receiver performs the iterative detection and decoding, the channel decoder feeds back the extrinsic LLRs to the detector to improve the system performance.

B. Conventional Channel Estimation

In this subsection, we briefly review the MMSE-based channel estimation, which is a typical process to estimate the channel for the multiuser MIMO systems. In this approach, each user estimates the channel using the pilot signals in the processing block (see Fig. 2). We assume that the mth user collects the N_p pilot measurements for the channel estimation. For notational simplicity, we reorder the resource index of the pilot positions $\{p_{k,l}^{(m)}\}_{k=1,\ldots,K,l=1,\ldots,L}$ as $\{p_i^{(m)}\}_{i=0,\ldots,N_p-1}$. Then, the corresponding received signals and channel gains can be expressed as $\{z_i^{(m)}\}_{i=0,\ldots,N_p-1}$ and $\{f_i^{(m)}\}_{i=0,\ldots,N_p-1}$, respectively. If we denote $\mathbf{z}_m = [z_0^{(m)} \quad \cdots \quad z_{N_p-1}^{(m)}]^T$ and $\mathbf{f}_m = [f_0^{(m)} \quad \cdots \quad f_{N_p-1}^{(m)}]^T$, then the vector form of the received pilot signal in (1) can be expressed as

$$\mathbf{z}_m = \mathbf{P}_m\mathbf{f}_m + \sum_{j=0,j\neq m}^{M-1}\mathbf{P}_j\mathbf{f}_j + \mathbf{n}_m, \tag{6}$$

where $\mathbf{P}_m = \text{diag}([p_0^{(m)} \quad \cdots \quad p_{N_p-1}^{(m)}]^T)$ is the diagonal matrix of the pilot symbols and $\mathbf{n}_m = [n_0^{(m)} \quad \cdots \quad n_{N_p-1}^{(m)}]^T$. Since the pilot symbol has a unit magnitude (i.e., $\mathbf{P}_m^H\mathbf{P}_m = \mathbf{I}$), the pilot

symbols can be removed from the received pilot vector $\mathbf{z}_m$ as

$$\mathbf{z}'_m = \mathbf{P}_m^H \mathbf{z}_m = \mathbf{f}_m + \mathbf{n}'_m \tag{7}$$

where $\mathbf{n}'_m = \mathbf{P}_m^H \mathbf{n}_m$. The linear MMSE estimate $\hat{\mathbf{f}}_m$ of $\mathbf{f}_m$ is given by

$$\hat{\mathbf{f}}_m = \mathrm{Cov}\left(\mathbf{f}_m, \mathbf{z}'_m\right) \mathrm{Cov}\left(\mathbf{z}'_m, \mathbf{z}'_m\right)^{-1} \mathbf{z}'_m \tag{8}$$

$$= \mathrm{Cov}\left(\mathbf{f}_m, \mathbf{f}_m\right) \left(\mathrm{Cov}\left(\mathbf{f}_m, \mathbf{f}_m\right) + \sigma_n^2 \mathbf{I}\right)^{-1} \mathbf{z}'_m, \tag{9}$$

where $\mathrm{Cov}(\mathbf{a}, \mathbf{b}) = E[\mathbf{a}\mathbf{b}^H] - E[\mathbf{a}]E[\mathbf{b}^H]$ and $\mathbf{n}'_m \sim \mathcal{CN}(0, \mathbf{I})$. Note that the covariance matrix $C_{f,f} = \mathrm{Cov}\left(\mathbf{f}_m, \mathbf{f}_m\right)$ can be obtained using the well-known Jake's model [21]

$$E\left[f_{k,l}^{(m)} f_{k+\Delta k, l+\Delta l}^{(m)}\right] = \left(\sum_{i=0}^{P-1} P_i e^{j2\pi \Delta k f_s \tau_i}\right) J_0(2\pi f_d T_s \Delta t), \tag{10}$$

where $E\left[f_{k,l}^{(m)} f_{k+\Delta k, l+\Delta l}^{(m)}\right]$ is the correlation of channel gains spaced by Δk frequency tones and Δl symbol times, f_s and f_d are the spacing between adjacent frequency tones and Doppler frequency respectively, T_s is the period for one OFDM symbol, and $J_0(x)$ is the 0-th order Bessel function. The MSE of the estimated channel $\hat{\mathbf{g}}$ for the MMSE-based channel estimation is [14]

$$\begin{aligned} E\left[||\mathbf{f}_m - \hat{\mathbf{f}}_m||^2\right] =& \mathrm{tr}\left(\mathrm{Cov}\left(\mathbf{f}_m - \hat{\mathbf{f}}_m, \mathbf{f}_m - \hat{\mathbf{f}}_m\right)\right) \\ =& \mathrm{tr}\left(C_{f,f} - C_{f,f}\left(C_{f,f} + \sigma_n^2 \mathbf{I}\right)^{-1} C_{f,f}\right). \end{aligned} \tag{11}$$

III. Proposed Channel estimation for multiuser MIMO Systems

Fig. 4 depicts the block diagram of the proposed EM-based joint pilot and data channel estimation method. First, UE estimates the channels for all users and then delivers them to the multiuser MIMO detector. Multiuser MIMO detector jointly estimates the channels and produces the soft symbols for all users. Using the generated soft symbols, the channel decoder computes the soft information of information bits in a form of log likelihood ratio (LLR) and then feeds it back to the channel estimator. Using this LLR information, small number of reliable data tones is picked for the channel re-estimation (see Fig. 5). The soft symbols of selected tones are used to generate the refined channel estimates. Such an iterative process is repeated until the termination condition (e.g., max iteration number or performance convergence) is satisfied.

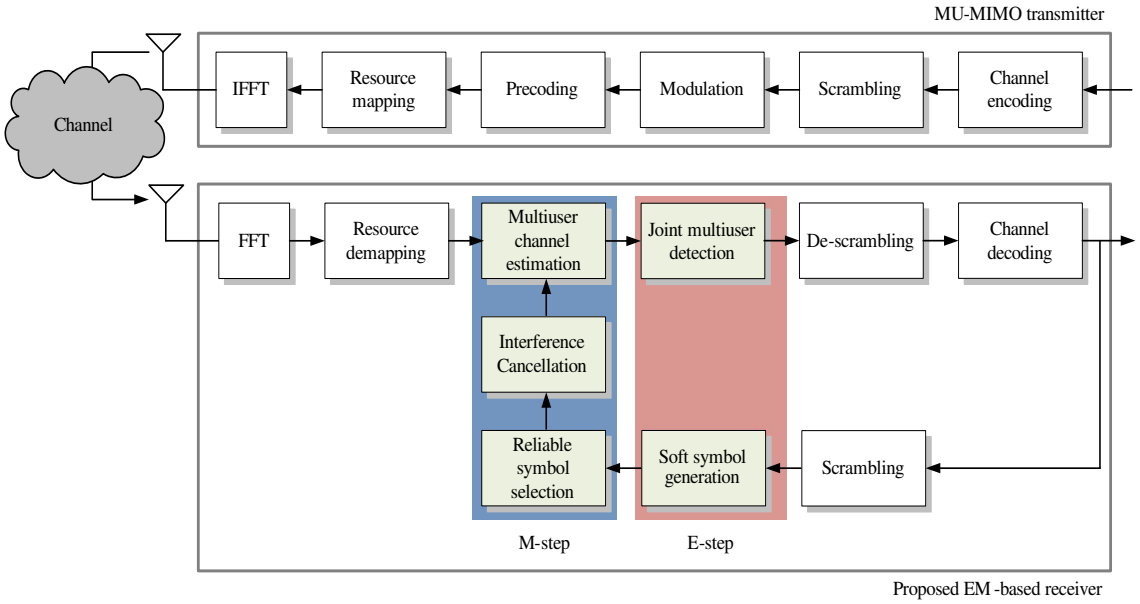


Fig. 4. Block diagram of the proposed EM-based joint pilot and data channel estimation method

A. Proposed EM-based Joint Pilot and Data Channel Estimation

Let $\mathbf{y}_m$ be the vector of N_d measurements (i.e., $\mathbf{y}_m = [y_0^{(m)} \quad \cdots \quad y_{N_d-1}^{(m)}]^T$), then we construct the observation vector by concatenating the received pilot vector $\mathbf{z}'_m$ in (7) and the received data vector $\mathbf{y}_m$ as

$$\begin{bmatrix} \mathbf{z}'_m \\ \mathbf{y}_m \end{bmatrix} = \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{X}_m \end{bmatrix} \mathbf{g}_m + \sum_{j=0, j\neq m}^{M-1} \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{X}_j \end{bmatrix} \mathbf{g}_j + \mathbf{n}_m \tag{12}$$

where $\mathbf{X}_m = \text{diag}([x_0^{(m)} \quad \cdots \quad x_{N_d-1}^{(m)}]^T)$ is the diagonal matrix whose diagonal entries are the data symbols for the mth user, $\mathbf{g}_m = [f_0^{(m)} \quad \cdots \quad f_{N_p-1}^{(m)} \quad g_0^{(m)} \quad \cdots \quad g_{N_d-1}^{(m)}]^T$ is the concatenated vector of pilot channels $[f_0^{(m)} \quad \cdots \quad f_{N_p-1}^{(m)}]^T$ and data channels $[g_0^{(m)} \quad \cdots \quad g_{N_d-1}^{(m)}]^T$, and $\mathbf{n}_m$ is the noise vector. While the multiuser interferences at the pilot tones can be removed perfectly (see Section II-B), the effect of multiuser interferences at the data tones (the second term in the right-hand side of (12)) would be non-negligible since the precoding operation at the base station is imperfect. Since this multiuser interferences cause a performance degradation, we cancel the estimate of the interuser interference terms in (12) using the soft estimate of the interferer's data

symbols. That is,

$$\begin{bmatrix} \tilde{\mathbf{z}}_m \\ \tilde{\mathbf{y}}_m \end{bmatrix} = \begin{bmatrix} \mathbf{z}'_m \\ \mathbf{y}_m \end{bmatrix} - \sum_{j=0, j\neq m}^{M-1} \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \overline{\mathbf{X}}_j \end{bmatrix} \hat{\mathbf{g}}_j^{(l)} \tag{13}$$

where $\hat{\mathbf{g}}_j^{(l)}$ is the estimated channel vector from the l-th iteration, and $\overline{\mathbf{X}}_j = \mathrm{diag}([\overline{x}_0^{(j)} \quad \cdots \quad \overline{x}_{N_d-1}^{(j)}]^T)$ is the soft estimate of the inteferer's symbol $\mathbf{X}_j$. Note that the first and second order statistics of dth diagonal element in $\mathbf{X}_j$ are given by [22]

$$\begin{aligned} \overline{x}_d^{(j)} = E[x_d^{(j)}] &= \sum_{\theta\in\Theta} \theta \prod_{i=0}^{Q-1} \Pr(c_{i,d}^{(j)}) \\ &= \sum_{\theta\in\Theta} \theta \prod_{i=0}^{Q-1} \frac{1}{2}\left(1 + c_{i,d}^{(j)} \tanh\left(\frac{1}{2} L\left(c_{i,d}^{(j)}\right)\right)\right), \end{aligned} \tag{14}$$

$$\lambda_d^{(j)} = E[(x_d^{(j)})^2] = \sum_{\theta\in\Theta} \theta^2 \prod_{i=0}^{Q-1} \frac{1}{2}\left(1 + c_{i,d}^{(j)} \tanh\left(\frac{1}{2} L\left(c_{i,d}^{(j)}\right)\right)\right), \tag{15}$$

where $c_{i,d}^{(j)}$ is i-th bit information comprising the symbol $x_d^{(j)}$, Q is the number of information bits for $x_d^{(j)}$, and $L(c_{i,d}^{(j)})$ is the LLR value for $c_{i,d}^{(j)}$, respectively. From (12) and (13), we have

$$\begin{bmatrix} \tilde{\mathbf{z}}_m \\ \tilde{\mathbf{y}}_m \end{bmatrix} = \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{X}_m \end{bmatrix} \mathbf{g}_m + \mathbf{R} + \mathbf{n}_m. \tag{16}$$

where $\mathbf{R} = \sum_{j=0, j\neq m}^{M-1} \left(\begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{X}_j \end{bmatrix} \left(\mathbf{g}_j - \hat{\mathbf{g}}_j^{(l)}\right) - \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{X}_j - \overline{\mathbf{X}}_j \end{bmatrix} \hat{\mathbf{g}}_j^{(l)} \right)$ is the residual interference after the cancellation.

The goal of the proposed scheme is to estimate the channels $\mathbf{g}_m$ using these interference cancelled received vectors $\tilde{\mathbf{z}}_m$ and $\tilde{\mathbf{y}}_m$. The maximum a posteriori (MAP) estimate of the channel $\mathbf{g}_m$ is expressed as

$$\hat{\mathbf{g}}_m = \arg\max_{\mathbf{g}_m} \Pr\left(\mathbf{g}_m | \, \tilde{\mathbf{z}}_m, \tilde{\mathbf{y}}_m\right) \tag{17}$$

$$= \arg\max_{\mathbf{g}_m} \int_{\mathbf{X}_m} \Pr\left(\tilde{\mathbf{z}}_m, \tilde{\mathbf{y}}_m | \, \mathbf{X}_m, \mathbf{g}_m\right) \Pr(\mathbf{X}_m) \Pr(\mathbf{g}_m) d\mathbf{X}_m. \tag{18}$$

In (18), we see that the data symbols $\mathbf{X}_m$ need to be marginalized to generate the MAP estimate. Since the computational complexity of this marginalization process increases exponentially with the dimension of $\mathbf{X}_m$, this process requires humongous computational burden. To overcome the problem, we employ the EM algorithm for estimating $\mathbf{g}_m$. The EM algorithm is an efficient means to find out the Bayesian estimate of an unknown signal in the presence of unobserved latent

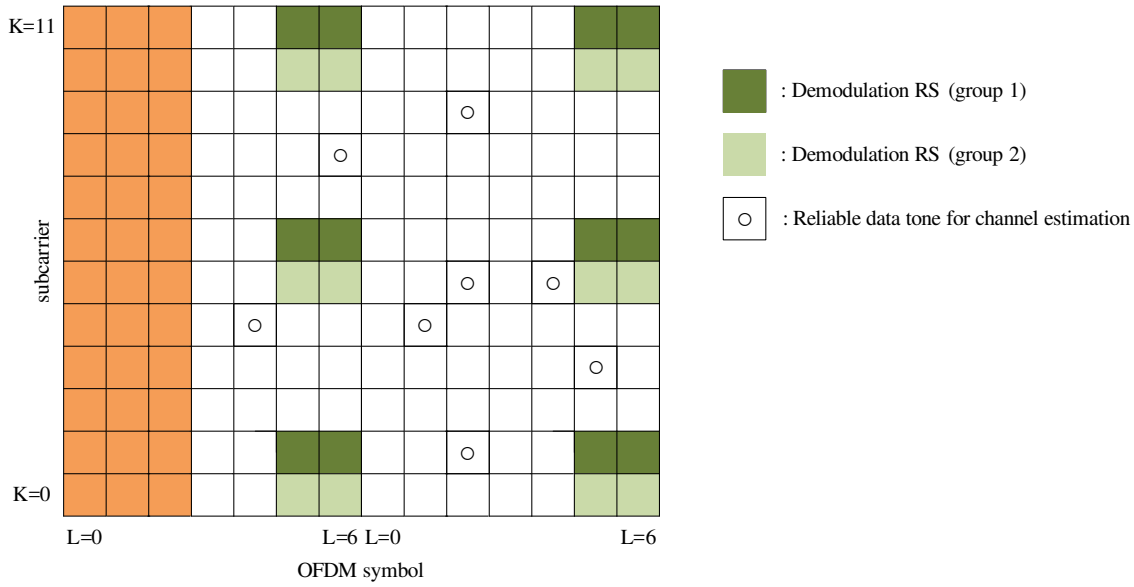


Fig. 5. Example of reliable data selection for proposed EM-based joint pilot and data channel estimation technique

variables [14]. Specifically, the EM algorithm alternates the following two steps: 1) computing the expected log-likelihood function of the complete data (E-step) and 2) maximizing the expectation result with respect to the parameters (M-step). Using $\mathbf{X}_m, \tilde{\mathbf{y}}_m$, and $\tilde{\mathbf{z}}_m$ as inputs, the EM algorithm iteratively computes the following two steps;

- Expectation step (E-step)

$$Q(\mathbf{g}_m; \hat{\mathbf{g}}_m^{(l)}) = E_{\mathbf{X}_m}\left[\log \Pr\left(\mathbf{g}_m | \mathbf{X}_m, \tilde{\mathbf{y}}_m, \tilde{\mathbf{z}}_m\right) | \hat{\mathbf{g}}_m^{(l)}, \tilde{\mathbf{y}}_m, \tilde{\mathbf{z}}_m\right] \tag{19}$$

- Maximization step (M-step)

$$\hat{\mathbf{g}}_m^{(l+1)} = \arg\max_{\mathbf{g}_m} Q(\mathbf{g}_m; \hat{\mathbf{g}}_m^{(l)}) \tag{20}$$

In the E-step, we evaluate the expected conditional probability when the channel estimates from the previous iteration are given. Since the expectation is with respect to the latent variable $\mathbf{X}_m$ and $\mathbf{X}_m$ is generated from the soft information of the decoder, quality of the estimated channel $\hat{\mathbf{g}}_m^{(l)}$ improves with the number of iterations. In the M-step, we find the channel estimates $\hat{\mathbf{g}}_m^{(l+1)}$ for the next iteration which corresponds to the argument maximizing the cost function $Q(\mathbf{g}_m; \hat{\mathbf{g}}_m^{(l)})$.

B. The E-step

The first task of the E-step is to reformulate the cost metric $Q(\mathbf{g}_m; \hat{\mathbf{g}}_m^{(l)})$ in a simple form to handle. Using Bayes' rule, we have

$$Q(\mathbf{g}_m; \hat{\mathbf{g}}_m^{(l)}) = E_{\mathbf{X}_m}\left[\log \Pr\left(\mathbf{g}_m | \mathbf{X}_m, \tilde{\mathbf{y}}_m, \tilde{\mathbf{z}}_m\right) | \hat{\mathbf{g}}_m^{(l)}, \tilde{\mathbf{y}}_m, \tilde{\mathbf{z}}_m\right] \tag{21}$$

$$= E_{\mathbf{X}_m}\left[\log \Pr\left(\tilde{\mathbf{y}}_m, \tilde{\mathbf{z}}_m | \mathbf{g}_m, \mathbf{X}_m\right) - \log \Pr\left(\tilde{\mathbf{y}}_m, \tilde{\mathbf{z}}_m\right) + \log \Pr\left(\mathbf{g}_m\right) | \xi\right], \tag{22}$$

where $\xi = \{\hat{\mathbf{g}}_m^{(l)}, \tilde{\mathbf{y}}_m, \tilde{\mathbf{z}}_m\}$. Assuming that the residual interference plus noise terms in (16) follow a Gaussian distribution, which is true when the number of co-scheduled users is large, the conditional probability density $\Pr\left(\tilde{\mathbf{y}}_m, \tilde{\mathbf{z}}_m | \mathbf{g}_m, \mathbf{X}_m\right)$ becomes

$$\begin{aligned}
&\Pr(\tilde{\mathbf{y}}_m, \tilde{\mathbf{z}}_m | \mathbf{g}_m, \mathbf{X}_m) \\
&= \frac{1}{\sqrt{(2\pi)^n |\Omega|}} \exp\left(- \left(\begin{bmatrix} \tilde{\mathbf{z}}_m \\ \tilde{\mathbf{y}}_m \end{bmatrix} - \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{X}_m \end{bmatrix} \mathbf{g}_m \right)^H \Omega^{-1} \left(\begin{bmatrix} \tilde{\mathbf{z}}_m \\ \tilde{\mathbf{y}}_m \end{bmatrix} - \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{X}_m \end{bmatrix} \mathbf{g}_m \right) \right) \\
&= \frac{1}{\sqrt{(2\pi)^n |\Omega|}} \exp\left(- \begin{bmatrix} \tilde{\mathbf{z}}_m \\ \tilde{\mathbf{y}}_m \end{bmatrix}^H \Omega^{-1} \begin{bmatrix} \tilde{\mathbf{z}}_m \\ \tilde{\mathbf{y}}_m \end{bmatrix} + \begin{bmatrix} \tilde{\mathbf{z}}_m \\ \tilde{\mathbf{y}}_m \end{bmatrix}^H \Omega^{-1} \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{X}_m \end{bmatrix} \mathbf{g}_m \right. \\
&\qquad \left. + \mathbf{g}_m^H \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{X}_m \end{bmatrix}^H \Omega^{-1} \begin{bmatrix} \tilde{\mathbf{z}}_m \\ \tilde{\mathbf{y}}_m \end{bmatrix} - \mathbf{g}_m^H \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{X}_m \end{bmatrix}^H \Omega^{-1} \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{X}_m \end{bmatrix} \mathbf{g}_m \right),
\end{aligned} \tag{23}$$

where

$$\begin{aligned}
\Omega =& \mathrm{Cov}\left(\sum_{j \neq m} \left(\begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{X}_j \end{bmatrix} \left(\mathbf{g}_j - \hat{\mathbf{g}}_j^{(l)} \right) + \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{X}_j - \overline{\mathbf{X}}_j \end{bmatrix} \hat{\mathbf{g}}_j^{(l)} \right) + \mathbf{n}_m \right) \\
=& E\left[\sum_{j \neq m} \left(\begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{X}_j \end{bmatrix} \left(\mathbf{g}_j - \hat{\mathbf{g}}_j^{(l)} \right) \left(\mathbf{g}_j - \hat{\mathbf{g}}_j^{(l)} \right)^H \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{X}_j \end{bmatrix}^H \right. \right. \\
&\qquad \left. \left. + \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{X}_j - \overline{\mathbf{X}}_j \end{bmatrix} \hat{\mathbf{g}}_j^{(l)} (\hat{\mathbf{g}}_j^{(l)})^H \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{X}_j - \overline{\mathbf{X}}_j \end{bmatrix}^H \right) + \sigma^2 \mathbf{I} \right]
\end{aligned} \tag{24}$$

$$= \sum_{j \neq m} \left(\begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \Lambda_j \end{bmatrix} \odot C_{e_j, e_j}^{(l)} + \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \Lambda_j - \overline{\mathbf{x}}_j \overline{\mathbf{x}}_j^H \end{bmatrix} \odot \left(\hat{\mathbf{g}}_j^{(l)} (\hat{\mathbf{g}}_j^{(l)})^H \right) \right) + \sigma^2 \mathbf{I}, \tag{25}$$

where the operator $\odot$ is the Hadamard product[1], $C_{e_j, e_j}^{(l)} = \mathrm{Cov}\left(\mathbf{g}_j - \hat{\mathbf{g}}_j^{(l)} \right)$ is the error covariance matrix (see Appendix A), and $\Lambda_j = \mathrm{diag}([\lambda_0^{(j)} \quad \lambda_1^{(j)} \quad \cdots \quad \lambda_{N_d-1}^{(j)}]^T) = E[\mathbf{X}_j \mathbf{X}_j^H]$ is the second

[1]The hadamard product is an element-by-element product between two matrices, e.g., $\begin{bmatrix} a & b \\ c & d \end{bmatrix} \odot \begin{bmatrix} e & f \\ g & h \end{bmatrix} = \begin{bmatrix} ae & bf \\ cg & dh \end{bmatrix}$.

order statistic of $\mathbf{X}_j$, respectively. In (25), we assume that the estimation error $\mathbf{g}_j - \hat{\mathbf{g}}_j^{(l)}$ and the estimated channel $\hat{\mathbf{g}}_j^{(l)}$ are uncorrelated. If we assume that the *a priori* distribution of $\mathbf{g}_m$ is Gaussian, we have

$$\Pr\left(\mathbf{g}_m\right) = \frac{1}{\sqrt{(2\pi)^n |C_{g,g}|}} \exp\left(-\mathbf{g}_m^H C_{g,g}^{-1} \mathbf{g}_m\right), \tag{26}$$

where $C_{g,g}$ is obtained using the correlation model for fading channels (see (10) in Section II-B). Using (23) and (26), $Q(\mathbf{g}_m; \hat{\mathbf{g}}_m^{(l)})$ in (22) can be expressed as

$$\begin{aligned} Q(\mathbf{g}_m; \hat{\mathbf{g}}_m^{(l)}) =& E_{\mathbf{X}_m} \left[\begin{bmatrix} \tilde{\mathbf{z}}_m \\ \tilde{\mathbf{y}}_m \end{bmatrix}^H \Omega^{-1} \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{X}_m \end{bmatrix} \mathbf{g}_m + \mathbf{g}_m^H \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{X}_m \end{bmatrix}^H \Omega^{-1} \begin{bmatrix} \tilde{\mathbf{y}}_m \\ \tilde{\mathbf{z}}_m \end{bmatrix} \right. \\ & \left. - \mathbf{g}_m^H \begin{bmatrix} \mathbf{I} & 0 \\ 0 & \mathbf{X}_m \end{bmatrix}^H \Omega^{-1} \begin{bmatrix} \mathbf{I} & 0 \\ 0 & \mathbf{X}_m \end{bmatrix} \mathbf{g}_m - \mathbf{g}_m^H C_{g,g}^{-1} \mathbf{g}_m + C \middle| \xi \right] \qquad (27) \\ =& \begin{bmatrix} \tilde{\mathbf{z}}_m \\ \tilde{\mathbf{y}}_m \end{bmatrix}^H \Omega^{-1} \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & E_{\mathbf{X}_m}[\mathbf{X}_m|\xi] \end{bmatrix} \mathbf{g}_m + \mathbf{g}_m^H \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & E_{\mathbf{X}_m}[\mathbf{X}_m|\xi]^H \end{bmatrix} \Omega^{-1} \begin{bmatrix} \tilde{\mathbf{y}}_m \\ \tilde{\mathbf{z}}_m \end{bmatrix} \\ & - \mathbf{g}_m^H \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & E_{\mathbf{X}_m}[\mathbf{X}_m^H \mathbf{X}_m|\xi] \end{bmatrix} \odot \Omega^{-1} \mathbf{g}_m - \mathbf{g}_m^H C_{g,g}^{-1} \mathbf{g}_m + C, \qquad (28) \end{aligned}$$

where the constant C contains the terms unrelated to the variable $\mathbf{g}_m$. Note that $E_{\mathbf{X}_m}[\mathbf{X}_m|\xi]$ and $E_{\mathbf{X}_m}[\mathbf{X}_m^H \mathbf{X}_m|\xi]$ are the statistics of the data symbols $\mathbf{X}_m$ when the observation and the previous estimate $\hat{\mathbf{g}}_m^{(l)}$ are provided. In the standard EM formulation, the terms $E_{\mathbf{X}_m}[\mathbf{X}_m|\xi]$ and $E_{\mathbf{X}_m}[\mathbf{X}_m^H \mathbf{X}_m|\xi]$ are obtained from the soft output of the joint multiuser MIMO detector. In our derivation, we compute $E_{\mathbf{X}_m}[\mathbf{X}_m|\xi]$ and $E_{\mathbf{X}_m}[\mathbf{X}_m^H \mathbf{X}_m|\xi]$ using the soft information produced by the channel decoder, i.e., $E_{\mathbf{X}_m}[\mathbf{X}_m|\xi] = \overline{\mathbf{X}}_m$ and $E_{\mathbf{X}_m}[\mathbf{X}_m^H \mathbf{X}_m|\xi] = \mathbf{\Lambda}_m$. By substituting $E_{\mathbf{X}_m}[\mathbf{X}_m|\xi] = \overline{\mathbf{X}}_m$ and $E_{\mathbf{X}_m}[\mathbf{X}_m^H \mathbf{X}_m|\xi] = \mathbf{\Lambda}_m$, we have

$$\begin{aligned} Q(\mathbf{g}_m; \hat{\mathbf{g}}_m^{(l)}) =& \begin{bmatrix} \tilde{\mathbf{z}}_m \\ \tilde{\mathbf{y}}_m \end{bmatrix}^H \Omega^{-1} \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \overline{\mathbf{X}}_m \end{bmatrix} \mathbf{g}_m + \mathbf{g}_m^H \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \overline{\mathbf{X}}_m^H \end{bmatrix} \Omega^{-1} \begin{bmatrix} \tilde{\mathbf{y}}_m \\ \tilde{\mathbf{z}}_m \end{bmatrix} \\ & - \mathbf{g}_m^H \left(\begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{\Lambda}_m \end{bmatrix} \odot \Omega^{-1} + C_{g,g}^{-1} \right) \mathbf{g}_m + C. \end{aligned} \tag{29}$$

C. The M-step

The goal of the M-step is to find the channel vector $\mathbf{g}_m$ maximizing the cost function $Q(\mathbf{g}_m; \hat{\mathbf{g}}_m^{(l)})$:

$$\hat{\mathbf{g}}_m^{(l+1)} = \arg\max_{\mathbf{g}_m} Q(\mathbf{g}_m; \hat{\mathbf{g}}_m^{(l)}) \tag{30}$$

Since $Q(\mathbf{g}_m; \hat{\mathbf{g}}_m^{(l)})$ is a concave function of $\mathbf{g}_m$[2], $\hat{\mathbf{g}}_m^{(l+1)}$ can be found by setting the gradient of $Q(\mathbf{g}_m; \hat{\mathbf{g}}_m^{(l)})$ to zero. From (29), we have

$$\nabla_{\mathbf{g}_m} Q(\mathbf{g}_m; \hat{\mathbf{g}}_m^{(l)}) = \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \overline{\mathbf{X}}_m^H \end{bmatrix} \Omega^{-1} \begin{bmatrix} \tilde{\mathbf{z}}_m \\ \tilde{\mathbf{y}}_m \end{bmatrix} - \left(\begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{\Lambda}_m \end{bmatrix} \odot \Omega^{-1} + C_{g,g}^{-1} \right) \mathbf{g}_m. \tag{31}$$

Solving $\nabla_{\mathbf{g}_m} Q(\mathbf{g}_m; \hat{\mathbf{g}}_m^{(l)}) = 0$ yields

$$\hat{\mathbf{g}}_m^{(l+1)} = \left(\begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{\Lambda}_m \end{bmatrix} \mathrm{diag}\left(\Omega^{-1}\right) + C_{g,g}^{-1} \right)^{-1} \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \overline{\mathbf{X}}_m^H \end{bmatrix} \Omega^{-1} \begin{bmatrix} \tilde{\mathbf{z}}_m \\ \tilde{\mathbf{y}}_m \end{bmatrix} \tag{32}$$

$$= C_{g,g} \Sigma^{-1} \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \overline{\mathbf{X}}_m^H \end{bmatrix} \Omega^{-1} \begin{bmatrix} \tilde{\mathbf{z}}_m \\ \tilde{\mathbf{y}}_m \end{bmatrix} \tag{33}$$

where $\Sigma = \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{\Lambda}_m \end{bmatrix} \mathrm{diag}\left(\Omega^{-1}\right) C_{g,g} + \mathbf{I}$. Note that since Σ is a diagonal matrix, computation of inverse Σ^{-1} is trivial, and hence the complexity of the channel estimation in (33) is dominated by the inverse operation of Ω.

D. Reliable data symbol selection criterion

Although employing all reliable data tones in the processing block might be the best choice to improve the channel estimation quality, this approach will increase the dimension of the matrix Ω, causing a significant burden in the computation of C_{e_j,e_j} in Ω (see (25)). To address this issue, we choose only small number of data tones that are highly likely to be beneficial in the channel re-estimation. In order to choose the best possible data tones among all candidates, we

[2]Since the differentiation of $Q(\mathbf{g}_m; \hat{\mathbf{g}}_m^{(l)})$ is a monotonic decreasing function of $\mathbf{g}_m$ in (31), $Q(\mathbf{g}_m; \hat{\mathbf{g}}_m^{(l)})$ is a concave function for $\mathbf{g}_m$ [23].

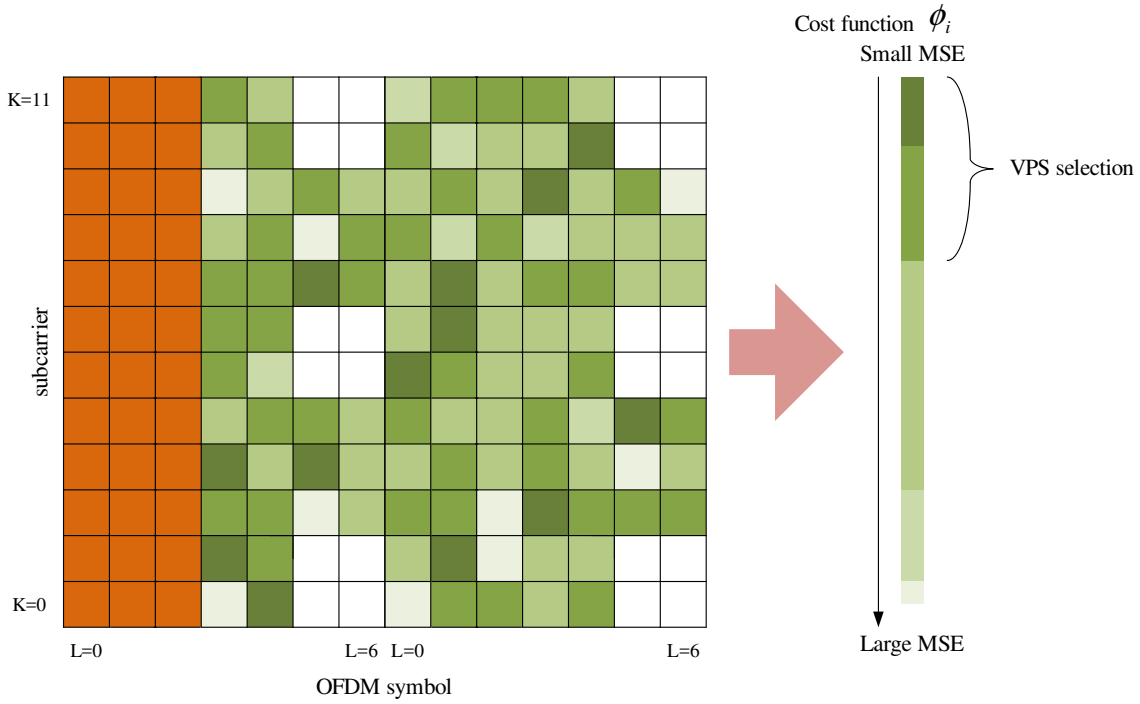


Fig. 6. Illustration of reliable data tone selection process

basically need to evaluate the MSE of all possible N_d data tone combinations and then choose the best combination generating the minimum MSE, i.e.,

$$\mathbf{X}_m = \arg \min_{\mathbf{X}_{\{N_d\}} \in \Phi_X} E\left[\|\mathbf{g}_m - \hat{\mathbf{g}}_m\|^2\right] \tag{34}$$

$$= \arg \min_{\mathbf{X}_{\{N_d\}} \in \Phi_X} tr\left(\mathrm{Cov}\left(\mathbf{g}_m - \hat{\mathbf{g}}_m, \mathbf{g}_m - \hat{\mathbf{g}}_m\right)\right) \tag{35}$$

$$= \arg \min_{\mathbf{X}_{\{N_d\}} \in \Phi_X} tr\left(C_{g,g} - C_{g,g}\left(C_{g,g} + \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{\Lambda}_{\{N_d\}} \end{bmatrix} \odot \Omega\right)^{-1} C_{g,g}\right), \tag{36}$$

where $\mathbf{X}_{\{N_d\}}$ is the possible combination of N_d tones selection, Φ_X is the set of all combinations, $\mathbf{\Lambda}_{\{N_d\}}$ is the second order statistic of $\mathbf{X}_{\{N_d\}}$, respectively. Since such procedure requires a combinatorial search[3], we instead use a simple approach computing the MSE of the estimated channel when only a single data tone is used. Among all reliable data tones, we pick N_d-best data tones generating the minimum MSE. In Fig. 6, we illustrate the simple example of the reliable

[3]Since the total number of data tones in the processing block is $KL - N_p$, the number of possible combinations for N_d data tone selection is $\binom{KL-N_p}{N_d}$.

data symbol selection for the proposed channel estimation algorithm. Although this heuristic might not be optimal since it does not capture the correlation among N_d virtual pilot tones, the computational complexity is very small ($KL - N_p$ comparisons) and also is effective in improving the channel estimation quality (see Section IV-B).

Let $\mathbf{g}_m = [g_0^{(m)} \quad \cdots \quad g_{N_p-1}^{(m)}]^T$ be the channel vector associated with pilot tones and g_i be the channel gain for the single data tone candidate. Then, the MSE associated with the channel estimation is (see Appendix A)

$$\phi_i = E\left[\left\|\begin{bmatrix}\mathbf{g}_m \\ g_i\end{bmatrix} - \begin{bmatrix}\hat{\mathbf{g}}_m^{(l)} \\ \hat{g}_i^{(l)}\end{bmatrix}\right\|^2\right] \tag{37}$$

$$= tr\left(\text{Cov}\left(\begin{bmatrix}\mathbf{g}_m \\ g_i\end{bmatrix} - \begin{bmatrix}\hat{\mathbf{g}}_m^{(l)} \\ \hat{g}_i^{(l)}\end{bmatrix}, \begin{bmatrix}\mathbf{g}_m \\ g_i\end{bmatrix} - \begin{bmatrix}\hat{\mathbf{g}}_m^{(l)} \\ \hat{g}_i^{(l)}\end{bmatrix}\right)\right) \tag{38}$$

$$= tr\left(C_{G,G} - C_{G,G}\left(\begin{bmatrix}\mathbf{I} & \mathbf{0} \\ \mathbf{0} & \lambda_i^{-1}\end{bmatrix} \odot \begin{bmatrix}\Omega_{11} & \Omega_{12} \\ \Omega_{21} & \Omega_{22}\end{bmatrix} + C_{G,G}\right)^{-1} C_{G,G}\right) \tag{39}$$

$$= tr\left(C_{G,G} - C_{G,G}\begin{bmatrix}\Phi_{11} & \Phi_{12} \\ \Phi_{21} & \Phi_{22}\end{bmatrix} C_{G,G}\right) \tag{40}$$

$$\begin{aligned} =& tr\left(C_{g,g} - C_{g,g}\Phi_{11}C_{g,g} - C_{g,g'}\Phi_{21}C_{g,g} - C_{g,g}\Phi_{12}C_{g',g} - C_{g,g'}\Phi_{22}C_{g',g}\right) \\ &+ tr\left(1 - C_{g',g}\Phi_{11}C_{g,g'} - \Phi_{21}C_{g,g'} - C_{g',g}\Phi_{12} - \Phi_{22}\right) \end{aligned} \tag{41}$$

where $C_{G,G} = \begin{bmatrix}C_{g,g} & C_{g,g'} \\ C_{g',g} & C_{g',g'}\end{bmatrix} = \begin{bmatrix}Cov(\mathbf{g}_m, \mathbf{g}_m) & Cov(\mathbf{g}_m, g_i) \\ Cov(g_i, \mathbf{g}_m) & Cov(g_i, g_i)\end{bmatrix}$ is the autocovariance matrix of $\begin{bmatrix}\mathbf{g}_m \\ g_i\end{bmatrix}$, $C_{g',g'} = 1$, and $\begin{bmatrix}\Phi_{11} & \Phi_{12} \\ \Phi_{21} & \Phi_{22}\end{bmatrix} = \begin{bmatrix}\mathbf{I} \odot \Omega_{11} + C_{g,g} & C_{g,g'} \\ C_{g',g} & \lambda_i^{-1}\Omega_{22} + C_{g',g'}\end{bmatrix}^{-1}$. Denoting

$$\phi_{i,1} = tr\left(C_{g,g} - C_{g,g}\Phi_{11}C_{g,g} - C_{g,g'}\Phi_{21}C_{g,g} - C_{g,g}\Phi_{12}C_{g',g} - C_{g,g'}\Phi_{22}C_{g',g}\right) \tag{42}$$

$$\phi_{i,2} = tr\left(1 - C_{g',g}\Phi_{11}C_{g,g'} - \Phi_{21}C_{g,g'} - C_{g',g}\Phi_{12} - \Phi_{22}\right), \tag{43}$$

then $\phi_i = \phi_{i,1} + \phi_{i,2}$. After some manipulations, we have (see Appendix B)

$$\phi_i = \text{tr}\left[C_{g,g} - C_{g,g}\Psi C_{g,g}\right] + \frac{\left(\Gamma_{22} - 1\right)\left(1 - C_{g',g}\Psi C_{g,g'}\right) - \left|C_{g,g'} - C_{g,g}\Psi C_{g,g'}\right|^2}{\Gamma_{22} - C_{g',g}\Psi C_{g,g'}}, \tag{44}$$

where $\Psi = \Gamma_{11}^{-1}$, $\Gamma_{11} = \Omega_{11} + C_{g,g}$, and $\Gamma_{22} = \lambda_i^{-1}\Omega_{22} + 1$. Since the computation of $\Psi = (\Omega_{11} + C_{g,g})^{-1}$ in the cost function (44) is cumbersome, we assume that the pilot tones are

spaced apart, which is true for the LTE-Advanced systems [4]. Under this assumption, the correlation between different pilot signals is negligible ($C_{g,g} \approx \mathbf{I}$) and hence

$$\phi_i = \mathrm{tr}\left[\frac{\sigma^2}{1+\sigma^2}\mathbf{I}\right] + \frac{(\Gamma_{22}-1)\left(1-(1+\sigma^2)^{-1}C_{g',g}C_{g,g'}\right) - \left|C_{g,g'} - (1+\sigma^2)^{-1}C_{g,g'}\right|^2}{\Gamma_{22} - (1+\sigma^2)^{-1}C_{g',g}C_{g,g'}}, \quad (45)$$

where $\Omega_{11} = \sigma^2\mathbf{I}$. In the high SNR regime ($\sigma^2 \approx 0$), we also have

$$\phi_i \approx \frac{(\Gamma_{22}-1)\left(1-C_{g',g}C_{g,g'}\right) - \left|C_{g,g'} - C_{g,g'}\right|^2}{\Gamma_{22} - C_{g',g}C_{g,g'}} \quad (46)$$

$$= \frac{\left(1-\Gamma_{22}^{-1}\right)\left(1-\left\|C_{g,g'}\right\|^2\right)}{1-\Gamma_{22}^{-1}\left\|C_{g,g'}\right\|^2} \quad (47)$$

$$= \frac{1}{\frac{1}{1-\Gamma_{22}^{-1}} + \frac{1}{1-\left\|C_{g,g'}\right\|^2} - 1}. \quad (48)$$

Since the cost function ϕ_i is the MSE of proposed channel estimation when the data symbol x_i is used for the virtual pilot, the desired data tone x_{i^*} ($i^* = \arg\min_i \phi_i$) should be that maximizing the denominator of (48). In other words, to minimize the cost function ϕ_i, both Γ_{22}^{-1} and $C_{g,g'}$ should be maximized. Since $\Gamma_{22}^{-1} = \frac{1}{\lambda_i^{-1}\Omega_{22}+1}\left(\approx \frac{\lambda_i}{\Omega_{22}}\right)$ is proportional to the second order statistic of the data symbol λ_i in (15) and inversely proportional to the amount of residual interference Ω_{22} in (25), we can expect that the data tone maximizing Γ_{22}^{-1} can improve the channel estimation quality. Also, we need to choose a data tone maximizing the correlation $C_{g,g'}$ between the channel $\mathbf{g}_m$ at the pilot position and g_i at given data tone position, since otherwise the selected data tone would not be helpful. In summary, the data tones with 1) high reliability Γ_{22}^{-1} and 2) large absolute correlation $\|C_{g,g'}\|^2$ need to be chosen as virtual pilots for the channel re-estimation. The proposed channel estimation algorithm is summarized in Table I.

E. Comments on Complexity

In this subsection, we analyze the computational complexity of the proposed EM-based channel estimation method and the conventional MMSE channel estimation. Note that since the detection and decoding operation at the receiver is common, we focus on computations associated with the channel estimation exclusively. While the major operation of the conventional MMSE estimation is the inversion and multiplication of matrix associated with pilot signals, the interference cancellation of the received signal in (13) and calculation of coefficients associated with data signal Ω and Σ in (33) are additionally required for the proposed scheme. Denoting the complexity associated with conventional MMSE estimation, interference cancellation, calculation

TABLE I

SUMMARY OF THE PROPOSED CHANNEL ESTIMATION ALGORITHM

Input: The LLRs on the transmitted symbols $\{\mathbf{X}_m\}$, the received vectors $\{\mathbf{z}_m\}$ and $\{\mathbf{y}_m\}$, the pilot sequence $\{\mathbf{P}_m\}$

STEP 1: (**Pilot removal**) Obtain the received signal $\mathbf{z}'_m$ whose pilot signal is removed using (7).

$$\mathbf{z}'_m = \mathbf{P}_m^H \mathbf{z}_m = \mathbf{f}_m + \mathbf{n}'_m$$

STEP 2: (**Reliable symbol selection**) Select the data tones according to the cost function in (48).

$$\phi_i = \frac{1}{\frac{1}{1-\Gamma_{22}^{-1}} + \frac{1}{1-\|C_{g,g'}\|^2} - 1}$$

STEP 3: (**Interference subtraction**) Subtract the effect of interfering symbols $\mathbf{X}_j$ from other streams using (13).

$$\begin{bmatrix} \tilde{\mathbf{z}}_m \\ \tilde{\mathbf{y}}_m \end{bmatrix} = \begin{bmatrix} \mathbf{z}'_m \\ \mathbf{y}_m \end{bmatrix} - \sum_{j=0, j\neq m}^{M-1} \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \overline{\mathbf{X}}_j \end{bmatrix} \hat{\mathbf{g}}_j^{(l)}$$

STEP 4: (**Covariance matrix update**) Update the error covariance matrix for the channel estimation from (53).

$$C_{e_j,e_j}^{(l)} = C_{g,g} - C_{g,g} \left(C_{g,g} + \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{\Lambda}_j \end{bmatrix} \odot \Omega \right)^{-1} C_{g,g}$$

STEP 5: (**Channel estimation**) Calculate the channel estimate of $\hat{\mathbf{g}}_m$ using (33).

$$\hat{\mathbf{g}}_m = C_{g,g} \Sigma^{-1} \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \overline{\mathbf{X}}_m^H \end{bmatrix} \Omega^{-1} \begin{bmatrix} \tilde{\mathbf{y}}_m \\ \tilde{\mathbf{z}}_m \end{bmatrix}$$

STEP 7: (**Interference stream estimation**) Perform **STEP 2** $\sim$ **6** for jth interference stream.

of Ω and Σ, and the proposed scheme as C_{conv}, C_{IC}, C_{Ω}, C_{Σ}, and C_{prop}, respectively, then the number of required floating-point operations (flops) is as follows [24]:

- C_{conv} requires $\frac{4}{3}N_p^3 + \frac{3}{2}N_p^2 + \frac{7}{6}N_p$ flops for $\left(\mathrm{Cov}\left(\mathbf{f}_m, \mathbf{f}_m\right) + \sigma_n^2 \mathbf{I}\right)^{-1}$ and $2N_p^3 + N_p^2 - N_p$ flops for the rest part of multiplication.
- C_{IC} requires $2N_d(M-1)$ flops for computing (13).
- C_{Ω} requires $(M-1)(N_p^2 + 2N_p + 2N_pN_d + 2N_d^2 + 5N_d) + N_p + N_d$ flops for matrix elements computation associated with (25).
- C_{Σ} requires $\frac{4}{3}(N_p+N_d)^3 + \frac{3}{2}(N_p+N_d)^2 - \frac{5}{6}(N_p+N_d)$ flops for inverting Ω and $2(N_p + N_d)^2 + (N_p + N_d)$ flops for computing $\left(\begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{\Lambda}_m \end{bmatrix} \mathrm{diag}\left(\Omega^{-1}\right) C_{g,g} + \mathbf{I} \right)$.
- C_{prop} requires C_{IC}, C_{Ω}, C_{Σ}, and $9(N_p+N_d)^2 - (N_p+N_d)$ flops for matrix inversion of Σ and matrix multiplication of $C_{g,g}\Sigma^{-1} \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \overline{\mathbf{X}}_m^H \end{bmatrix} \Omega^{-1} \begin{bmatrix} \tilde{\mathbf{z}}_m \\ \tilde{\mathbf{y}}_m \end{bmatrix}$.

Since the proposed method has more number of pilots including virtual pilot signals, computa-

TABLE II

COMPUTATIONAL COMPLEXITY (FLOPS) OF THE PROPOSED METHOD AND CONVENTIONAL MMSE ESTIMATION IN CASE OF $N_p = 12$ AND $M = 4$

Operations	$N_d = 2$	$N_d = 4$	$N_d = 8$	$N_d = 16$
Interference cancellation	12	24	48	96
Ω computation	716	964	1604	3460
Σ computation	4347	6360	12070	32018
Proposed method	6825	9636	17302	42602
Conventional MMSE estimation	6122			

tional complexity of proposed method is higher than that of conventional method. However, since the complexity of proposed method depends strongly on the number of virtual pilot signals, by choosing an appropriate number of virtual pilot signals, computational burden can be minimized. Table II summarizes flops of the proposed method and conventional MMSE estimation.

IV. SIMULATIONS

A. Simulation Setup

In this section, we evaluate the performance of the proposed channel estimation technique through numerical simulations. The simulation setup is based on a multiuser MIMO system in the 3GPP LTE-A standard [4] and LTE-Advanced-Pro [25]. We consider 4×4, 8×8, and 16×16 multiuser MIMO systems. As a channel model, we use extended vehicular A (EVA) channel model [26]. The basic pilot allocation follows the DM-RS sequence specified in the 3GPP LTE-A standard. Since 16×16 multiuser MIMO is not defined in the standard, we allocate the pilot signals by adding two DM-RS groups in 9th to 16th transmit antennas. We assume that the number of co-scheduled users M equals the number of antennas N_T. When we generate the precoding vector $\mathbf{w}_{i'}$ of the desired user, we assume that the receiver feeds back 8-bit quantized channel vector. In performing the channel vector quantization, we use DFT-based codebook given by

$$i' = \arg\max_i \left\| \left(\mathbf{h}_{(k,l)}^{(m)} \right)^H \mathbf{w}_i \right\|^2,$$

where $\mathbf{w}_i = \frac{1}{\sqrt{M}}[1 \quad e^{-j\frac{2\pi i}{M}} \quad e^{-j\frac{4\pi i}{M}} \quad \cdots \quad e^{-j\frac{2\pi(M-1)i}{M}}]^T$ is i-th column vector of the DFT-based codebook matrix [27].

For the channel coding, we use Turbo code with the $1/2$ rate, feedback polynomial $1+D+D^2$, and feedforward polynomial $1+D^2$. Each codeblock spans a length of 15 physical RBs, which consists of 12 subcarrier tones and 14 OFDM symbols (see Fig. 5). As a multiuser MIMO detector, we employ a linear MMSE detector with parallel interference canceller (MMSE-PIC) [28]. As a performance measure, we consider the MSE of the estimated channel as well as the block error rate (BLER) of the codeblocks.

In our simulations, we study the performance of the following algorithms:

- SUD with MMSE-based channel estimator: The receiver performs the MMSE-based channel estimation. Desired user detection is performed after the channel estimation.
- MUD with MMSE-based channel estimator: Using the DM-RS structure, MMSE-based channel estimation and multiuser detection are performed for all interfering users.
- EM-based channel estimator [17]: In each iteration, the interference covariance matrix used to suppress the multiuser interference is generated using the received signals and the desired symbol of previous iteration. The soft symbols of data tones are used as pilot symbols.
- Proposed method: We pick 32 reliable data tones for 4×4 multiuser MIMO systems and 16 reliable data tones for 8×8 and 16×16 multiuser MIMO systems.
- SUD with perfect CSI: SUD is performed under the assumption that the receiver knows the desired channel only. This scheme provides the best possible bound of the conventional approaches.
- MUD with perfect CSI: The receiver performs MUD with perfect CSI of all interfering users. This scheme provides the best possible bound of the proposed method.

The details of the simulation setup are summarized in Table III.

B. Simulation Results

In Fig. 7(a), we investigate the MSE performance of the channel estimation schemes for 4×4 MIMO systems. Since all data tones including unreliable or inaccurate symbols are used in the conventional EM-based channel estimation scheme, the performance gain of the conventional EM-based channel estimation over the conventional MMSE-based channel estimator is marginal. Conversely, the proposed method outperforms the conventional LS and MMSE channel estimation techniques, achieving $15\% \sim 70\%$ reduction in MSE over the MMSE-based

TABLE III

DETAILS OF THE SIMULATION SETUP

Parameter	Test1	Test2	Test 3	Test 4
Number of transmit antennas	4	4	8	16
Number of scheduled users	4	4	8	16
Number of DM-RS group	2	2	2	4
Number of receiver antennas (per user)	1			
Number of assigned RBs	15			
Channel model	Extended Vehicular A			
Doppler frequency	70 Hz			
Modulation	QPSK			
Channel coding	Turbo code			
Code rate	$1/2$			
MIMO detector	MMSE-PIC	MMSE-SIC	MMSE-PIC	MMSE-PIC
Number of inner iterations (Turbo decoder)	8			
Number of outer iterations (IDD iteration)	5	5	7	5
Number of virtual RS	32	32	16	16
Power ratio between desired and interference	-3 dB	-3 dB	0 dB	0 dB

channel estimation. We also observe that the performance of the proposed scheme improves with the number of iterations. In Fig. 7(b), we plot the BLER performance of all techniques under consideration. Since the conventional EM-based algorithm treats the multiuser interference as a noise, performance of the conventional EM-based channel estimation is not so appealing. However, since the proposed method employs deliberately chosen reliable symbols to cancel the multiuser interference and estimate the channel, it achieves substantial performance gain. For example, the proposed scheme achieves more than 3 dB gain over the conventional EM-based channel estimation algorithm and about 0.7 dB gain over the MMSE channel estimation with MUD receiver at 10^{-1} BLER point.

Since the proposed scheme exploits the soft symbols generated from the detector and decoder as virtual pilot signals, it is of interest to investigate the impacts of the detection scheme on performance. There are a number of MIMO detection schemes outperforming MMSE-PIC [29]–[32]. Among these, we consider MMSE with soft interference cancellation (MMSE-SIC) detector [30] for 4×4 MIMO systems. In Fig. 8, we plot the MSE and BLER performance of the proposed approach when the MMSE-SIC is used as the MIMO detector. We observe from Fig. 8(a) that the

proposed method using the MMSE-SIC detector performs similar to that using the MMSE-PIC in Fig. 7(a). However, since the performance of proposed method depends heavily on the accuracy of virtual pilot signals and the MMSE-SIC detector outperforms MMSE-PIC, we observe from Fig. 8(b) that the BLER performance of the proposed algorithm is better than that of MMSE-PIC detector. Also, we observe that the performance gain of the proposed method is 1 dB over the MMSE channel estimation with MUD receiver and more than 3.5 dB over the conventional EM-based channel estimation algorithm at 10^{-1} BLER point.

We next consider the performance of the proposed method for 8×8 multiuser MIMO systems. Due to the increase in the number of interferers, we observe from Fig. 9(a) that the performance of all receivers under consideration is worse than that of 4×4 multiuser MIMO system. In this case, MSE performance of the proposed scheme is better than that of conventional approaches, resulting in less than 40% of MSE of the conventional EM-based channel estimation. In Fig. 9(b), we plot the BLER performance of channel estimation algorithms. Since most of SUD techniques treats the interference as a noise, when the number of interference streams increases, the effective signal-to-interference-plus-noise-ratio (SINR) of the SUD technique decreases. As shown in Fig. 9, performance gain of the proposed scheme is more than 4 dB over the conventional MMSE channel estimation with SUD receiver and 2.5 dB over the conventional EM-based channel estimation algorithm at 10^{-1} BLER point.

In Fig. 10(a), we investigate the MSE performance of the proposed and conventional channel estimation schemes for 16×16 multiuser MIMO systems. When compared to 8×8 multiuser MIMO systems, interferers are now doubled and the effective number of pilot tones is reduced by half. In other words, 4 resource elements share one pilot tone in 8×8 multiuser MIMO systems and 8 resource elements share one pilot tone in 16×16 multiuser MIMO systems. In this very low pilot density scenario, we expect that the proposed approach using the virtual pilot signals has clear benefit over the conventional scheme using pilot signals only. In fact, the proposed scheme outperforms the conventional MMSE and EM-based channel estimation by a large margin for whole range of SNR under test. For example, we observe that the proposed scheme maintains the performance gain over the conventional EM-based channel estimation and MMSE channel estimation with MUD, achieving more than 3 dB and 2 dB at 10^{-1} BLER point.

In Fig. 11, we plot the throughput for 8×8 and 16×16 MIMO systems. Note that the throughput is the practical measure to evaluate the effectiveness of the MIMO technique in the link-level simulations and the real field trials. Since 15 resource blocks are used for the data transmission,

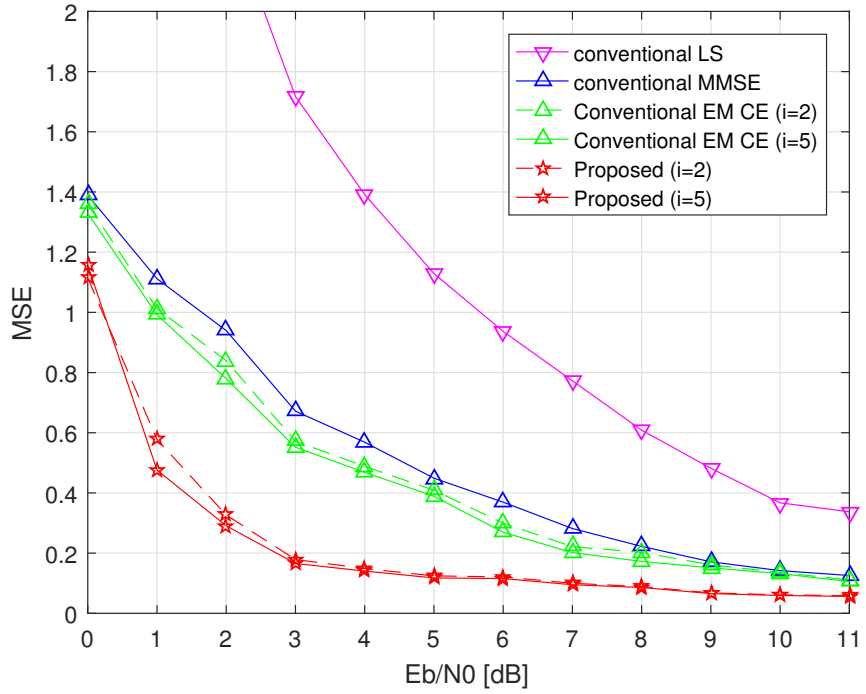
the maximum throughput is about 3.1 Mbps per scheduled user. While the conventional approach achieves the maximum throughput after 11 dB, the proposed method achieves it at 10 dB. In Fig. 11(b), since the number of interfering users increases for 16×16 systems, overall performance is worse than that of 8×8 MIMO system. In this scenario, performance gain of the proposed scheme is pronounced, achieving the maximum throughput at 16 dB, which is 3 dB faster than that of conventional approach.

Finally, we investigate the convergence behavior of the proposed method through extrinsic information transfer (EXIT) chart analysis [33], [34]. Fig. 12 provides the EXIT chart of the 8×8 MIMO system where QPSK transmission is considered and E_b/N_0 is set to 10 dB. In Fig. 12, $I_{Det,in}$ and $I_{Det,out}$ mean the mutual information at the input and output of the MIMO detector, which are measured between extrinsic LLRs and transmitted coded bits and $I_{Dec,in}$ and $I_{Dec,out}$ are those of the channel decoder. A graph of $I_{Det,in}$ versus $I_{Det,out}$ is an information transfer curve (ITC) of MIMO detectors and that of $I_{Dec,in}$ versus $I_{Dec,out}$ is the ITC of the channel decoder. In Fig. 12, we plot the ITCs of the MIMO detector when perfect CSI, the proposed scheme, the conventional LMMSE channel estimator are used. The ITC for the $(13, 15)$ RSC channel decoder is drawn in the same plot. From Fig. 12, we show that the proposed channel estimator achieves higher mutual information, thus exhibiting to faster convergence behavior compared to the standard channel estimator.

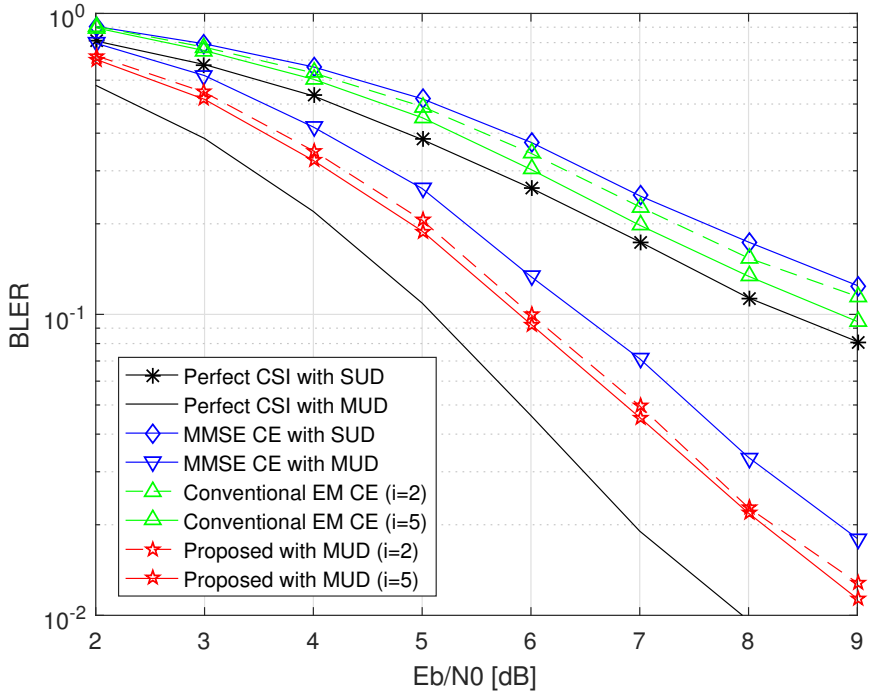
V. Concluding Remarks

In this paper, we proposed an EM-based joint pilot and data channel estimation algorithm for the multiuser MIMO systems. Our work is motivated by the observation that the inaccurate CSI caused by the insufficient pilot signals brings severe link performance degradation in multiuser MIMO systems. By using deliberately chosen data tones for pilot purpose, the proposed method achieves better channel estimation and eventual link performance gain. Although our study focused on the state of the art cellular systems (LTE-Advanced and LTE-Advanced-Pro), we expect that the effectiveness of the proposed can be readily extended to future wireless systems such as machine-type communication with ultra-small packets for internet of things (IoT) network. In this scenario, often referred to as ultra reliability and low latency (uRLLC) and machine type massive communication (mMTC), devices exchange small amount of information with the base station to achieve the ultra low latency and energy efficiency so that the size of packet is much smaller than that of human-centric communications. Accordingly, the number

of pilot tones in a packet is very small, and the proposed virtual pilot-based channel estimation strategy would be effective in improving the quality of service (QoS).

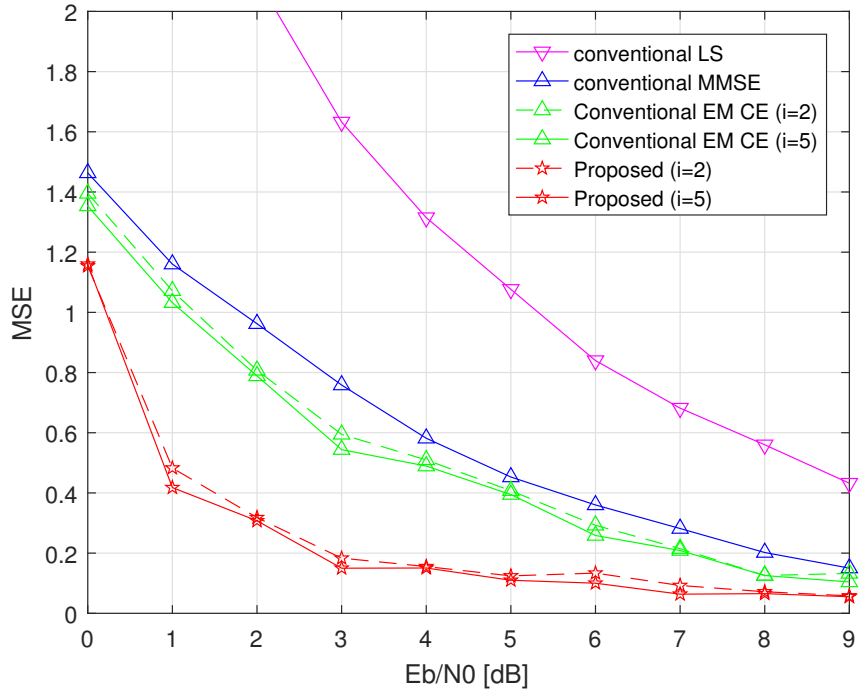


(a)

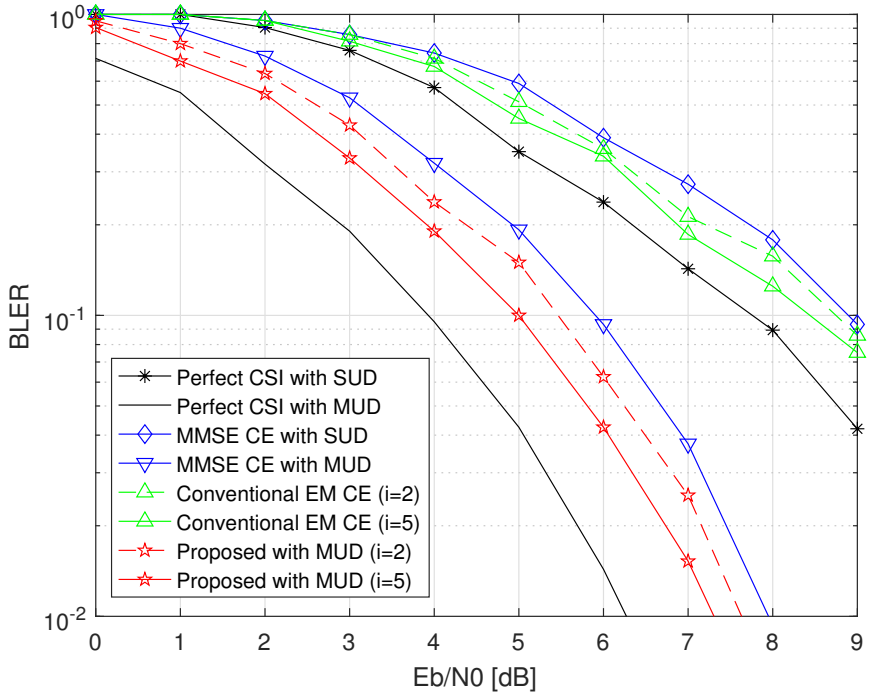


(b)

Fig. 7. MSE Performance of conventional and proposed scheme as a function of E_b/N_0 in 4×4 MIMO system : (a) MSE and (b) BLER performance after 2 and 5 iterations.

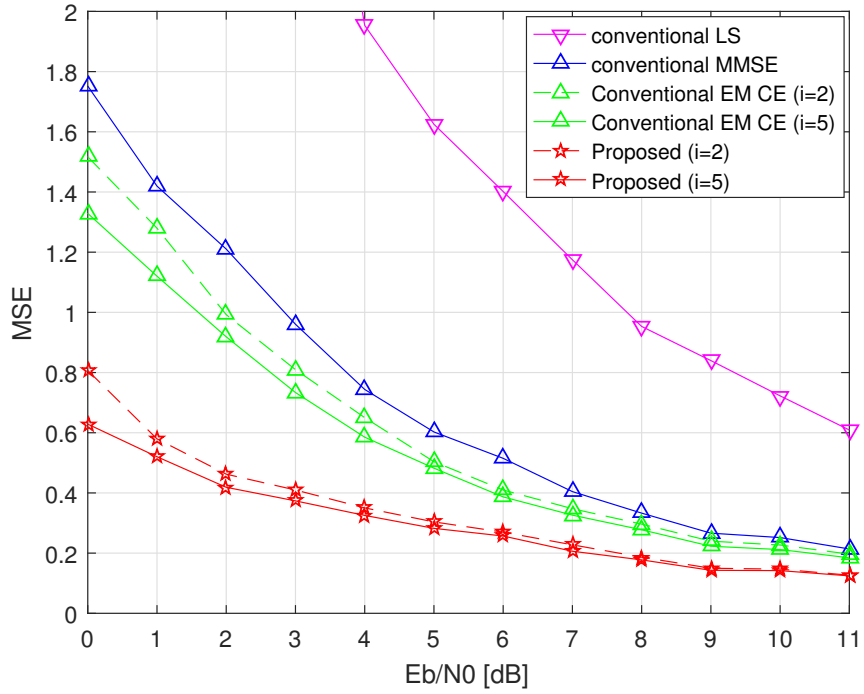


(a)

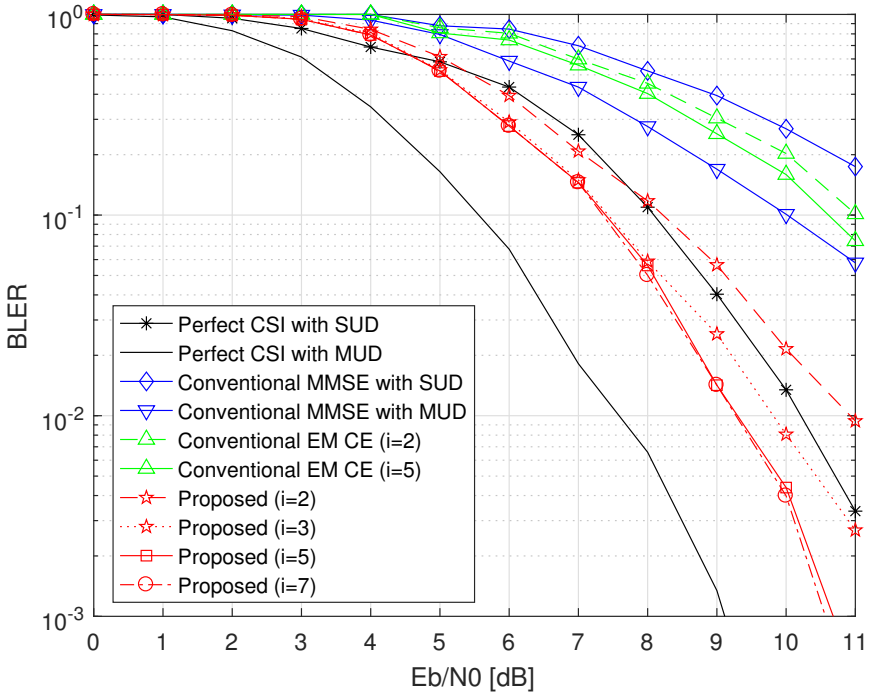


(b)

Fig. 8. MSE Performance of conventional and proposed scheme as a function of E_b/N_0 in 4×4 MIMO system with MMSE-SIC detector : (a) MSE and (b) BLER performance after 2 and 5 iterations.

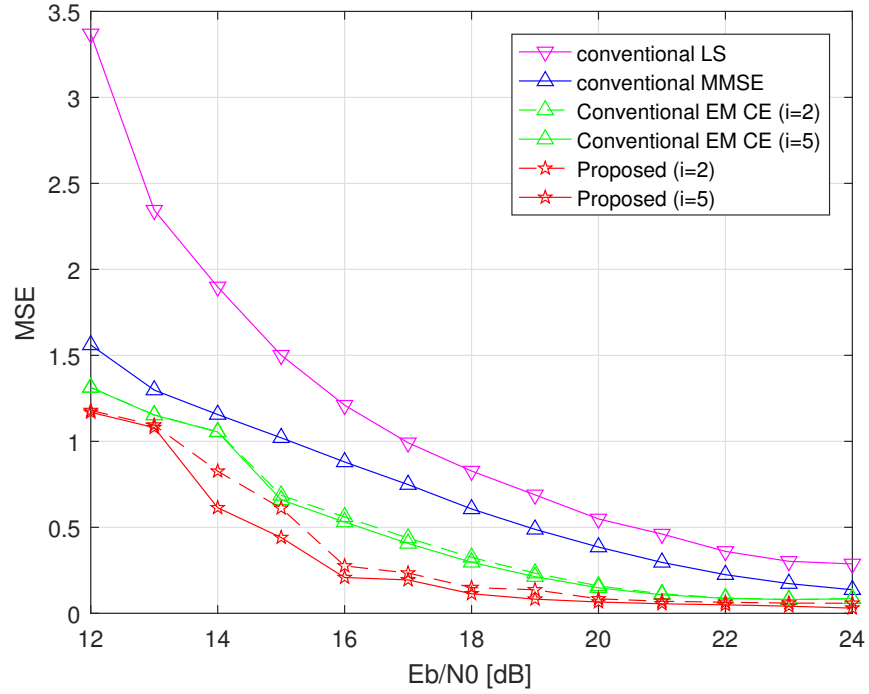


(a)

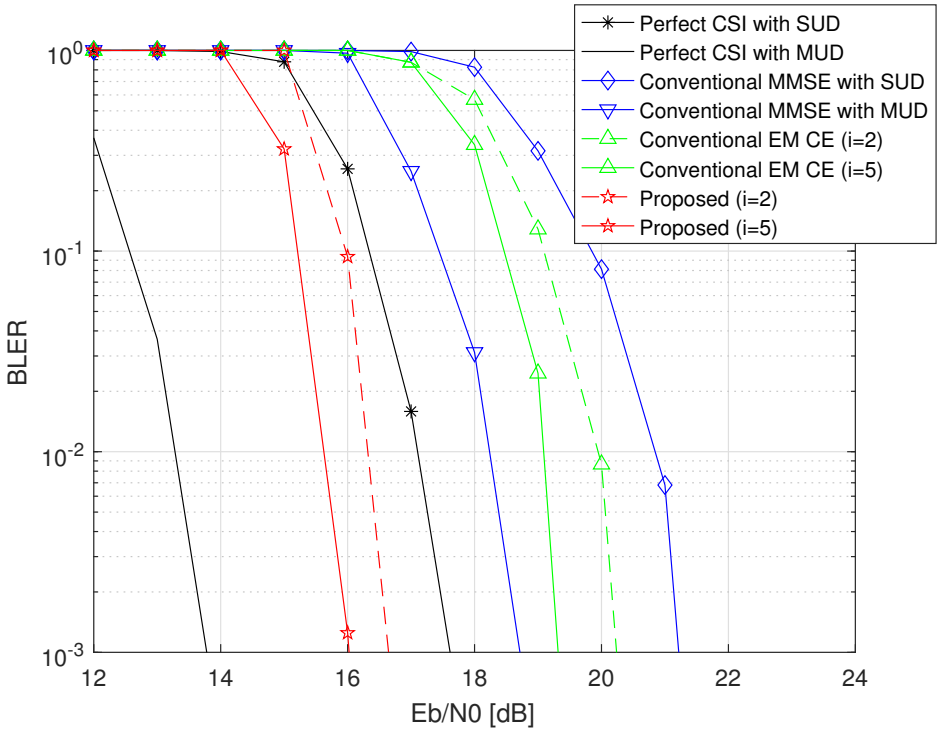


(b)

Fig. 9. Performance of conventional and proposed scheme as a function of E_b/N_0 in 8×8 MIMO system : (a) MSE and (b) BLER performance after 2 and 5 iterations.

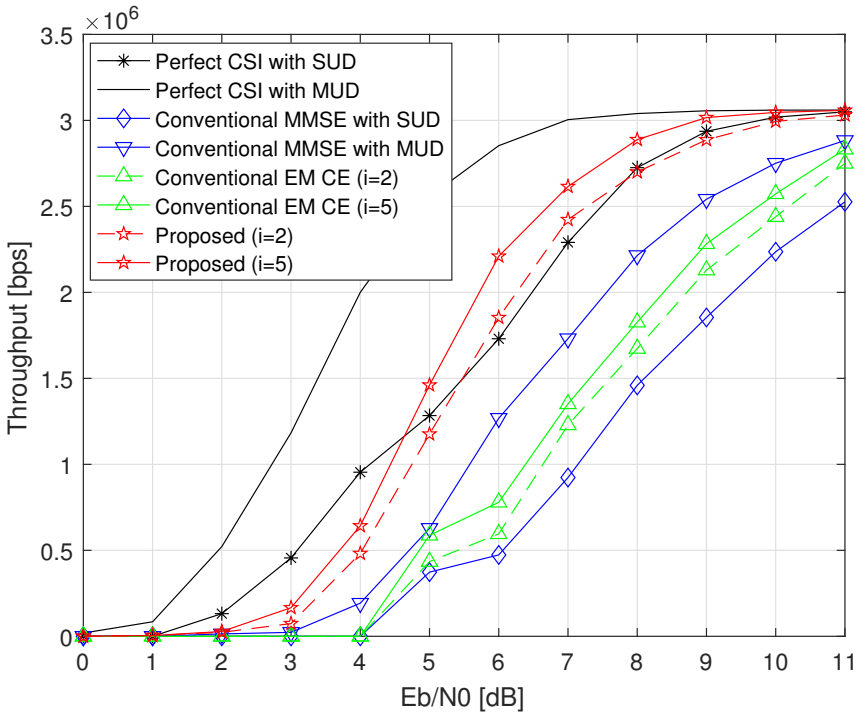


(a)

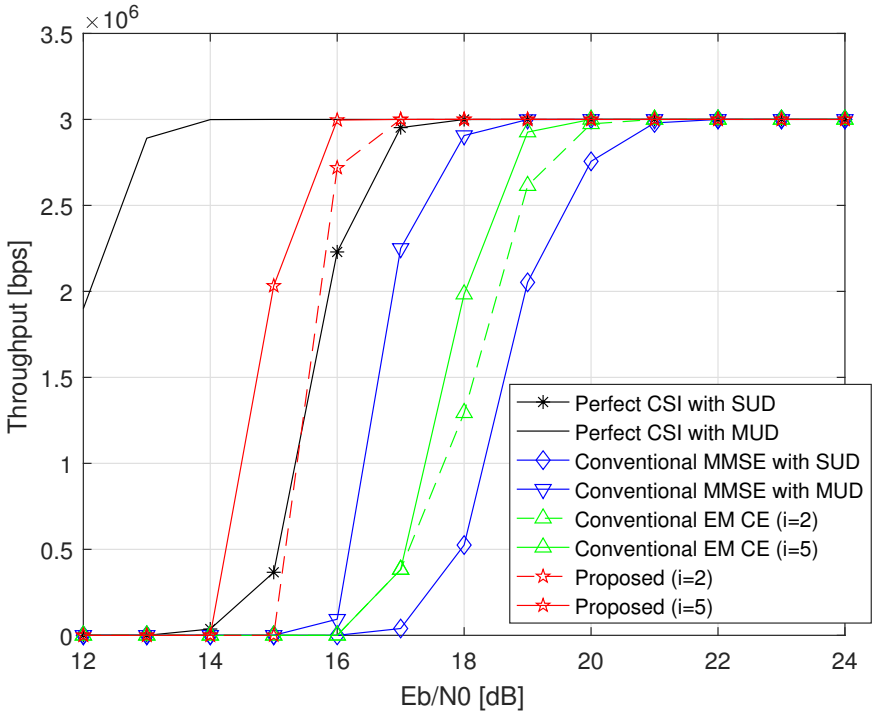


(b)

Fig. 10. Performance of conventional and proposed scheme as a function of E_b/N_0 in 16×16 MIMO system : (a) MSE and (b) BLER performance after 2 and 5 iterations.



(a)



(b)

Fig. 11. Throughput performance of conventional and proposed scheme as a function of E_b/N_0 in (a) 8×8 and (b) 16×16 MIMO systems.

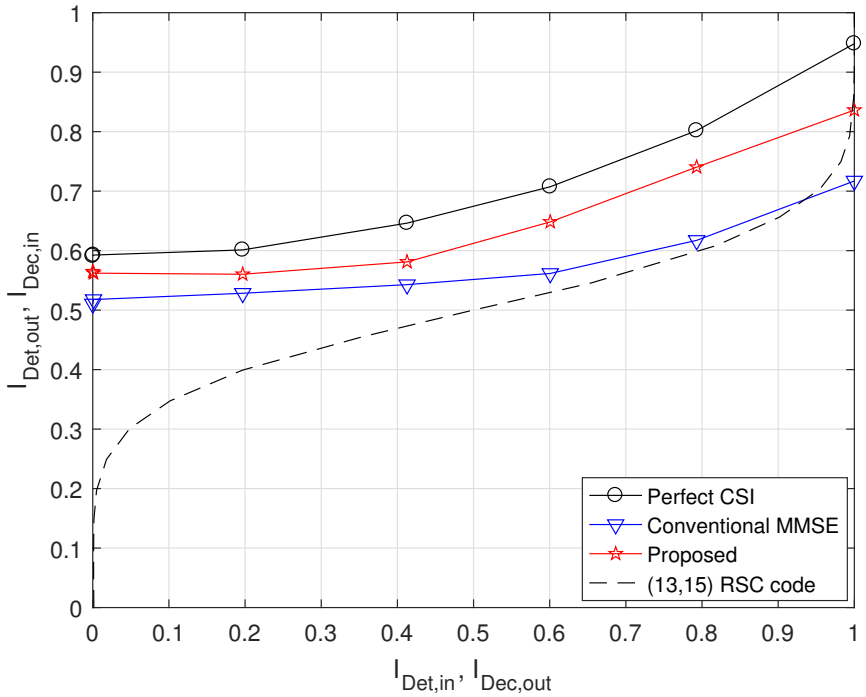


Fig. 12. EXIT chart for 8×8 MIMO system at the E_b/N_0 of 10 dB.

APPENDIX A
DERIVATION OF ERROR COVARIANCE $C_{e_j,e_j}^{(l)}$

Let $\hat{\mathbf{g}}_j^{(l)}$ be the channel estimate of $\mathbf{g}_j$ obtained in the l-th iteration and $\mathbf{e}_j^{(l)} = \mathbf{g}_j - \hat{\mathbf{g}}_j^{(l)}$ be the estimation error of channel $\mathbf{g}_j$, then the error covariance matrix $C_{e_j,e_j}^{(l)}$ becomes

$$C_{e_j,e_j}^{(l)} = \text{Cov}\left(\mathbf{e}_j, \mathbf{e}_j^H\right) \tag{49}$$

$$= E\left[\left(\mathbf{g}_j - \mathbf{A}_j \begin{bmatrix}\tilde{\mathbf{y}}_m \\ \tilde{\mathbf{z}}_m\end{bmatrix}\right)\left(\mathbf{g}_j^H - \begin{bmatrix}\tilde{\mathbf{y}}_m \\ \tilde{\mathbf{z}}_m\end{bmatrix}^H \mathbf{A}_j^H\right)\right], \tag{50}$$

where $A = \left(\begin{bmatrix}\mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{X}_j^H\end{bmatrix} \Omega^{-1} \begin{bmatrix}\mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{X}_j\end{bmatrix} + C_{g,g}^{-1}\right)^{-1} \begin{bmatrix}\mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{X}_j^H\end{bmatrix} \Omega^{-1}$ is the filter weight of proposed scheme. Under the assumption that the estimation error $\mathbf{e}_j^{(l)}$ and $\hat{\mathbf{g}}_j$ are uncorrelated, we further have

$$C_{e_j,e_j}^{(l)} = E\left[C_{g,g} - C_{g,g}\begin{bmatrix}\mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{X}_j^H\end{bmatrix}\mathbf{A}_j^H - \mathbf{A}_j\begin{bmatrix}\mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{X}_j\end{bmatrix} C_{g,g} + \mathbf{A}_j\left(\begin{bmatrix}\mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{X}_j\end{bmatrix} C_{g,g} \begin{bmatrix}\mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{X}_j^H\end{bmatrix} + \Omega\right)\mathbf{A}_j^H\right]. \tag{51}$$

Let $B = \begin{bmatrix}\mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{X}_j^H\end{bmatrix} \Omega^{-1} \begin{bmatrix}\mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{X}_j\end{bmatrix}$, then (51) becomes

$$C_{e_j,e_j}^{(l)} = E\left[C_{g,g} - C_{g,g}\left(C_{g,g} + B^{-1}\right)^{-1} C_{g,g}\right] \tag{52}$$

$$= C_{g,g} - C_{g,g}\left(C_{g,g} + \begin{bmatrix}\mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{\Lambda}_j\end{bmatrix} \odot \Omega\right)^{-1} C_{g,g}, \tag{53}$$

where $\mathbf{\Lambda}_j$ is the second order statistic of $\mathbf{X}_j$. Note that $\mathbf{\Lambda}_j$ and Ω can be obtained from the previous iteration.

APPENDIX B
DERIVATION OF EQUATION (44)

Let $\begin{bmatrix}\Phi_{11} & \Phi_{12}\\ \Phi_{21} & \Phi_{22}\end{bmatrix} = \begin{bmatrix}\Gamma_{11} & \Gamma_{12}\\ \Gamma_{21} & \Gamma_{22}\end{bmatrix}^{-1}$, then using the inversion formula of partitioned matrices[4], we have

$$\Phi_{11} = \left(\Gamma_{11} - \Gamma_{11}\Gamma_{22}^{-1}\Gamma_{21}\right)^{-1} = \Gamma_{11}^{-1} + \frac{\Gamma_{11}^{-1}\Gamma_{12}\Gamma_{21}\Gamma_{11}^{-1}}{\Gamma_{22} - \Gamma_{21}\Gamma_{11}^{-1}\Gamma_{12}} \tag{54}$$

$$\Phi_{12} = -\frac{\Gamma_{11}^{-1}\Gamma_{12}}{\Gamma_{22} - \Gamma_{21}\Gamma_{11}^{-1}\Gamma_{12}} \tag{55}$$

$$\Phi_{21} = -\Gamma_{22}^{-1}\Gamma_{21}\left(\Gamma_{11}^{-1} + \frac{\Gamma_{11}^{-1}\Gamma_{12}\Gamma_{21}\Gamma_{11}^{-1}}{\Gamma_{22} - \Gamma_{21}\Gamma_{11}^{-1}\Gamma_{12}}\right) \tag{56}$$

$$\Phi_{22} = \frac{1}{\Gamma_{22} - \Gamma_{21}\Gamma_{11}^{-1}\Gamma_{12}} \tag{57}$$

where (54) is from the matrix inversion lemma ($(A-BD^{-1}C)^{-1} = A^{-1}+A^{-1}B(D-CA^{-1}B)^{-1}CA^{-1}$). From (42) and (43), we further have

$$\begin{aligned}\phi_{i,1} =& \mathrm{tr}\left[C_{g,g} - C_{g,g}\Gamma_{11}^{-1}C_{g,g} - \frac{C_{g,g}\Gamma_{11}^{-1}\Gamma_{12}\Gamma_{21}\Gamma_{11}^{-1}C_{g,g}}{\Gamma_{22} - \Gamma_{21}\Gamma_{11}^{-1}\Gamma_{12}} + \frac{1}{\Gamma_{22}}C_{g,g'}\Gamma_{21}\Gamma_{11}^{-1}C_{g,g}\right.\\ &\left.+\frac{1}{\Gamma_{22}}\frac{C_{g,g'}\Gamma_{21}\Gamma_{11}^{-1}\Gamma_{12}\Gamma_{21}\Gamma_{11}^{-1}C_{g,g}}{\Gamma_{22} - \Gamma_{21}\Gamma_{11}^{-1}\Gamma_{12}} + \frac{C_{g,g}\Gamma_{11}^{-1}\Gamma_{12}C_{g',g}}{\Gamma_{22} - \Gamma_{21}\Gamma_{11}^{-1}\Gamma_{12}} - \frac{C_{g,g'}C_{g',g}}{\Gamma_{22} - \Gamma_{21}\Gamma_{11}^{-1}\Gamma_{12}}\right]\end{aligned} \tag{58}$$

$$\begin{aligned}\phi_{i,2} =& \mathrm{tr}\left[1 - C_{g',g}\Gamma_{11}^{-1}C_{g,g'} - \frac{C_{g',g}\Gamma_{11}^{-1}\Gamma_{12}\Gamma_{21}\Gamma_{11}^{-1}C_{g,g'}}{\Gamma_{22} - \Gamma_{21}\Gamma_{11}^{-1}\Gamma_{12}} + \frac{1}{\Gamma_{22}}\Gamma_{21}\Gamma_{11}^{-1}C_{g,g'}\right.\\ &\left.+\frac{1}{\Gamma_{22}}\frac{\Gamma_{21}\Gamma_{11}^{-1}\Gamma_{12}\Gamma_{21}\Gamma_{11}^{-1}C_{g,g'}}{\Gamma_{22} - \Gamma_{21}\Gamma_{11}^{-1}\Gamma_{12}} + \frac{C_{g',g}\Gamma_{11}^{-1}\Gamma_{12}}{\Gamma_{22} - \Gamma_{21}\Gamma_{11}^{-1}\Gamma_{12}} - \frac{1}{\Gamma_{22} - \Gamma_{21}\Gamma_{11}^{-1}\Gamma_{12}}\right].\end{aligned} \tag{59}$$

Using $\Gamma_{12} = C_{g,g'}$ and $\Gamma_{21} = C_{g',g}$, $\phi_{i,1}$ becomes

$$\begin{aligned}\phi_{i,1} =& \mathrm{tr}\left[C_{g,g} - C_{g,g}\Gamma_{11}^{-1}C_{g,g} - \frac{1}{\Gamma_{22} - C_{g',g}\Gamma_{11}^{-1}C_{g,g'}}\right.\\ &- C_{g,g'}C_{g',g}\Gamma_{11}^{-1}C_{g,g} + C_{g,g}\Gamma_{11}^{-1}C_{g,g'}C_{g',g}\Gamma_{11}^{-1}C_{gg}\\ &\left.-C_{gg}\Gamma_{11}^{-1}C_{g,g'}C_{g',g} + C_{g,g'}C_{g',g}\right)\Big].\end{aligned} \tag{60}$$

[4]Let $\begin{bmatrix}B_{11} & B_{12}\\ B_{21} & B_{22}\end{bmatrix} = \begin{bmatrix}A_{11} & A_{12}\\ A_{21} & A_{22}\end{bmatrix}^{-1}$, $\det(A_{11}) \neq 0$, and $\det(A_{22}) \neq 0$, then $B_{11} = (A_{11} - A_{12}A_{22}^{-1}A_{21})^{-1}$, $B_{12} = -A_{11}^{-1}A_{12}(A_{22} - A_{21}A_{11}^{-1}A_{12})^{-1}$, $B_{21} = -A_{22}^{-1}A_{21}(A_{11} - A_{12}A_{22}^{-1}A_{21})^{-1}$, and $B_{22} = (A_{22} - A_{21}A_{11}^{-1}A_{12})^{-1}$.

Denoting $\Psi = \Gamma_{11}^{-1}$, we have

$$\phi_{i,1} = \mathrm{tr}\left[C_{g,g} - C_{g,g}\Psi C_{g,g} - \frac{\left(C_{g,g'} - C_{g,g}\Psi C_{g,g'}\right)\left(C_{g',g} - C_{g',g}\Psi C_{g,g}\right)}{\Gamma_{22} - C_{g',g}\Psi C_{g,g'}}\right] \tag{61}$$

$$= \mathrm{tr}\left[C_{g,g} - C_{g,g}\Psi C_{g,g}\right] - \frac{\left(C_{g',g} - C_{g',g}\Psi C_{g,g}\right)\left(C_{g,g'} - C_{g,g}\Psi C_{g,g'}\right)}{\Gamma_{22} - C_{g',g}\Psi C_{g,g'}} \tag{62}$$

where (62) is because $\mathrm{tr}(AB) = \mathrm{tr}(BA)$ and $\mathrm{tr}(a) = a$ for scalar a. In a similar way, $\phi_{i,2}$ becomes

$$\phi_{i,2} = \mathrm{tr}\left[1 - C_{g',g}\Psi C_{g,g'} - \frac{\left(1 - C_{g',g}\Psi C_{g,g'}\right)^2}{\Gamma_{22} - C_{g',g}\Psi C_{g,g'}}\right] \tag{63}$$

$$= \frac{\left(\Gamma_{22} - 1\right)\left(1 - C_{g',g}\Psi C_{g,g'}\right)}{\Gamma_{22} - C_{g',g}\Psi C_{g,g'}}. \tag{64}$$

Finally, by adding (62) and (64), we have

$$\begin{aligned}\phi_i =& \mathrm{tr}\left[C_{g,g} - C_{g,g}\Sigma C_{g,g}\right] + \frac{\left(\Gamma_{22} - 1\right)\left(1 - C_{g',g}\Psi C_{g,g'}\right)}{\Gamma_{22} - C_{g',g}\Psi C_{g,g'}} \\ & - \frac{\left(C_{g',g} - C_{g',g}\Psi C_{gg}\right)\left(C_{g,g'} - C_{g,g}\Psi C_{g,g'}\right)}{\Gamma_{22} - C_{g',g}\Psi C_{g,g'}}.\end{aligned} \tag{65}$$

Note that the first term in the right-hand side of (65) is a function of $C_{g,g}$ and hence independent of data tones.

REFERENCES

[1] S. Park, J. Choi, K. Lee, and B. Shim, "Soft decision-directed channel estimation for multiuser MIMO systems," *Proc. of IEEE Personal indoor mobile radio communications (PIMRC) symposium*, 2015.

[2] S. Park, J. Choi, J. Seol, and B. Shim, "Virtual pilot-based channel estimation and multiuser detection for multiuser MIMO in LTE-Advanced," *Proc. of IEEE Vehicular technology conference (VTC)*, 2016.

[3] C. Lim, T. Yoo, B. Clercks, B. Lee, and B. Shim, "Recent trend of multiuser MIMO in LTE-Advanced," *IEEE Comm. Mag.*, pp 127-135, March 2013.

[4] 3GPP TS 36.211 V12.0.0 (2013-12): "Evolved Universal Terrestrial Radio Access (E-UTRA); Physical Channels and Modulation (Release 12)"

[5] H. Ji, Y. Kim, J. Lee, E. Onggosanusi, Y. Nam, Z. Jianzhong, B. Lee, and B. Shim, "Overview of Full-Dimension MIMO in LTE-Advanced Pro," To appear in *IEEE Comm. Mag.*

[6] P. S. Rossi and R. R. Muller, "Joint twofold-iterative channel estimation and multiuser detection for MIMO-OFDM systems," in *IEEE Trans. on Wire. Commun.*, vol. 7, no. 11, pp. 4719-4729, November 2008.

[7] C. Koike, D. Ogawa, T. Seyama, T. Dateki, "MLD-Based MU-MIMO detection scheme for LTE downlink," *Proc. IEEE Veh. Tech. Conf.*, pp. 1-5, 2012.

[8] J. Gao and H. Liu, "Decision-directed estimation of MIMO time-varying Rayleigh fading channels," *IEEE Trans. Wire. Commun.*, vol. 4, pp. 1412-1417, July 2005.

[9] H. Vikalo, B. Hassibi, and P. Stoica, "Efficient joint maximum-likelihood channel estimation and signal detection," *IEEE Trans. Wire. Commun.*, vol. 5, pp. 1838-1845, July 2006.

[10] R. Otnes and M. Tuchler, "Soft iterative channel estimation for turbo equalization: Comparison of channel estimation algorithms," *Proc. Int. Conf. Commun. Syst.*, Singapore, Nov. 2002, pp. 72-76.

[11] Y. Li, "Simplified channel estimation for OFDM systems with multiple transmit antennas," *IEEE Trans. Wire. Commun.,* vol. 1, pp. 67-75, Jan. 2002.

[12] Y. Liu and S. Sezginer, "Iterative compensated MMSE channel estimation in LTE systems," *Proc. IEEE Int. Conf. Commun.*, pp.4862-4866, 2012.

[13] A. P. Demster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *J. R. Statist. Soc. B*, vol. 39, pp. 1-38, 1977.

[14] S. M. Kay, *Fundamentals of statistical signal processing: estimation theory,* Prentice Hall, 1998.

[15] X. Wautelet, C. Herzet, A. Dejonghe, J. Louveaux and L. Vandendorpe, "Comparison of EM-Based algorithms for MIMO channel estimation," *IEEE Trans. Commun.*, vol. 55, no. 1, pp. 216-226, Jan. 2007.

[16] M. Khalighi and J. Boutros, "Semi-blind channel estimation using the EM algorithm in iterative MIMO APP detectors," *IEEE Trans. Wire. Commun.*, vol. 5, no. 11, pp. 3165-3173, November 2006.

[17] J. Choi, "An EM-based iterative receiver for MIMO-OFDM under interference-limited environments," *IEEE Trans. Wire. Commun.*, vol. 6, no. 11, pp. 3994-4003, November 2007.

[18] H. Zhu and J. Wang, "Chunk-based resource allocation in OFDMA systems - part I: chunk allocation," *IEEE Trans. Commun.*, vol. 57, no. 9, pp. 2734-2744, September 2009.

[19] H. Zhu and J. Wang, "Chunk-Based Resource Allocation in OFDMA SystemsPart II: Joint Chunk, Power and Bit Allocation," *IEEE Trans. Commun.*, vol. 60, no. 2, pp. 499-509, February 2012.

[20] J. Lee, B. Shim, and I. Kang, "Soft-input soft-output list sphere detection with a probabilistic radius tightening," *IEEE Trans. Wire. Commun.*, vol. 11, no. 8, pp. 2848-2857, August 2012.

[21] W. C. Jakes, "Microwave Mobile Communications," New York, John Wiley & Sons Inc, 1975.

[22] B. M. Hochwald and S. ten Brink, "Achieving near-capacity on a multiple-antenna channel," *IEEE Trans. Commun.*, vol. 51, pp. 389-399, March 2003.

[23] S. Boyd, L. Vandenberghe, *Convex Optimization,* Cambridge University Press, 2008.

[24] G. H. Golub and C. F. V. Loan, "Matrix Computations," 4th ed. The Johns Hopkins University Press, 2013.

[25] 3GPP TS 36.211 V13.0.0 (2015-12): "Evolved Universal Terrestrial Radio Access (E-UTRA); Physical Channels and Modulation (Release 13)"

[26] 3GPP TS 36.101 V12.0.0 (2013-07): "Evolved Universal Terrestrial Radio Access (E-UTRA); User Equipment (UE) radio transmission and reception (Release 12)"

[27] 3GPP TS 36.213 V12.0.0 (2013-12): "Evolved Universal Terrestrial Radio Access (E-UTRA); Physical layer procedures (Release 12)"

[28] M. Sellathurai and S. Haykin, "Turbo-BLAST for wireless communications: Theory and experiments," *IEEE Trans. Sig. Proc.*, vol. 50, no. 10, pp. 25382546, Oct. 2002.

[29] H. Lee, B. Lee, and I. Lee, "Iterative detection and decoding with an improved V-BLAST for MIMO-OFDM Systems," *IEEE J. Sel. Areas Commun.*, vol. 24, pp. 504-513, Mar. 2006.

[30] J. W. Choi, A. C. Singer, J Lee, N. I. Cho, "Improved linear soft-input soft-output detection via soft feedback successive interference cancellation," *IEEE Trans. Commun.*, vol.58, no.3, pp. 986-996, March 2010.

[31] S. Yang, T. Lv, R. Maunder, and L. Hanzo, "Unified Bit-Based Probabilistic Data Association Aided MIMO Detection for High-Order QAM Constellations," *IEEE Trans. Veh. Tech.*, vol. 60, no. 3, pp. 981-991, 2011.

[32] R. C. de Lamare, "Adaptive and Iterative Multi-Branch MMSE Decision Feedback Detection Algorithms for Multi-Antenna Systems," *IEEE Trans. Wire. Commun.*, vol. 14, no. 10, October 2013.

[33] S. ten Brink, "Convergence of iterative decoding," *Electron. Lett.*, vol. 35, pp. 806-808, May 1999.

[34] S. ten Brink, "Convergence behavior of iteratively decoded parallel concatenated codes," *IEEE Trans. Commun.*, vol. 49, pp. 1727-1737, Oct. 2001.