Variational Bayes Assisted Joint Signal Detection, Noise Covariance Estimation and Channel Tracking in MIMO-OFDM Systems

Xiaoying Zhang, Pei Xiao, *Senior Member, IEEE,* Dongtang Ma,
Jibo Wei, *Member, IEEE,*

Abstract

This paper introduces an improved variational bayes (Improved-VB) receiver algorithm for joint signal detection, noise covariance matrix estimation and channel impulse response (CIR) tracking in MIMO-OFDM systems over time varying channels. The variational bayes (VB) framework and turbo principle are combined to accomplish the parameter estimation and data detection. In the proposed Improved-VB receiver, a modified linear minimum mean-square-error interference cancellation (LMMSE-IC) soft detector is developed based on the VB theory, which adaptively sets the log-likelihood ratio (LLR) clipping value according to the reliability of detection on each subcarrier to mitigate the error propagation. Following the signal detection, an adaptive noise covariance matrix estimator is derived for the effective noise covariance estimation. Furthermore, in order to track time varying channels, a VB soft-input Kalman filter (VB-Soft-KF) is first derived. However, unreliable soft symbols introduce outliers, which degrade the performance of VB-Soft-KF. To tackle this problem, we propose a robust VB soft-input Kalman filter (VB-Robust-KF) based on the Huber M estimation theory. Finally, the performance

The material has been presented in parts at IEEE PIMRC 2010, Istanbul, Turkey, Sep. 26-30. This work has been supported in part by the National Natural Science Foundation of China under Grant 61101096, 61372099 and the Hunan Provincial Natural Science Foundation of China under Grant 11jj4055.

X. Zhang, D. Ma and J. Wei are with the School of Electronic Science and Engineering, National University of Defense Technology, P. R. China (e-mail: zxy_nudt@163.com, Dongtangma@nudt.edu.cn, windflow_cfd@163.com).

P. Xiao is with the center of communication systems research (CCSR), University of Surrey, Guilford, Surrey, GU2 7HX, UK (e-mail: p.xiao@surrey.ac.uk).

of the proposed algorithm is assessed via simulations, showing superior performance compared to the other benchmark receiver algorithms.

Index Terms

MIMO-OFDM, variational bayes (VB) algorithm, kalman filter, error propagation

I. Introduction

Multiple-Input-Multiple-Output (MIMO) and Orthogonal Frequency-Division Multiplexing (OFDM) are effective techniques for high data rate transmission. One of the important advantages of OFDM is its robustness to frequency-selective channels [1]. Multiple antennas can be used in conjunction with OFDM to improve the channel capacity and quality of communication systems [2]. MIMO-OFDM has been adopted in many wireless communication standards, such as the Worldwide Interoperability for Microwave Access (WiMAX) IEEE 802.16 e/m standard and the Third Generation Partnership Project long term evolution (3GPP LTE). All these emerging wireless systems are expected to offer both the high data rate and high-mobility transmissions. For example, LTE Release 10 promises to support a peak data rate of 100Mbps for high mobility up to 350 km/h. The high vehicular speeds of the user terminals make the channels subjected to significant time selectivity, rendering channel estimation a challenging task. The traditional pilot aided channel estimation algorithms need an abundance of extra resources to transmit pilot symbols to track the fast fading channels, which may significantly reduce the spectral efficiency and the data transmission rate.

In order to save the pilot overhead and improve the efficiency of communications, joint detection and channel estimation schemes have received considerable attentions [3]–[5]. In these schemes, pilot symbols are used to obtain initial channel estimation; the hard or soft decision symbols are exploited to re-estimate the channels which in turn refine the data detection. The performance of channel estimation and data detection is improved in an iterative manner. For example, the QR-decomposition (i.e., orthogonal triangular decomposition) M (QRD-M) algorithm combined with a decision-directed (DD) Kalman channel estimator was proposed for MIMO-OFDM systems in [3]. In [4], Kashima et al. developed iterative receivers which employ the data-aided recursive least squares (RLS) algorithm to improve the channel tracking capability. In the above mentioned schemes, hard decision symbols are used for channel estimation, which are sensitive to error propagation as a result of erroneous detection. The receiver algorithm in [5] considered the soft DD channel estimation by employing the expectation maximization (EM) algorithm. However, as reported in [6]–[8], both soft and hard DD channel tracking methods are susceptible to error propagation, especially in fast fading channels. Furthermore, the noise variance is usually assumed to be known to the receiver, which may not be a realistic assumption in many practical situations.

It is well known that the optimal joint estimation of multiple parameters is intractable in most applications, necessitating the use of alternative suboptimal approaches. The Variational Bayes (VB) algorithm [9] effectively solves this problem by iteratively finding the optimal set of marginal distributions in the sense of minimizing the Kullback-Leibler (KL) distance. In [10] and [11], variational inference theory was applied to derive iterative receiver for Code-Division Multiple Access (CDMA) systems. The VB algorithm was also applied to joint data detection and channel estimation for single-input single-output OFDM systems [12], [13] and MIMO-OFDM systems [14]. We have studied the VB assisted iterative receiver with joint data detection and channel estimation for MIMO-OFDM systems in [15]–[18]. In [12]–[18], after applying the VB algorithm, the soft symbols and the corresponding estimation variances calculated from the signal detector were fed to the channel estimators. However, in the case of fast fading, the initial channel estimates used by the detectors may be poor, leading to unreliable soft symbol estimates, which in turn degrade the channel estimation performance and give rise to error propagation.

To overcome the shortcomings of the conventional VB receiver, we propose an Improved-VB receiver algorithm for joint data detection, noise covariance estimation and channel estimation, which can combat the error propagation effectively. The proposed VB receiver achieves a reliable system performance over fast fading channels even in the case where only one training symbol is transmitted at the start of the data frame with moderate length. It provides an efficient solution for high data rate and high-mobility wireless transmissions. The main contributions of this paper are listed as follows.

Firstly, a modified linear minimum mean-square-error interference cancellation (MLMMSE-IC) detector is derived based on the VB theory for data detection. In contrast to the conventional LMMSE-IC detector [19], [20], the MLMMSE-IC detector takes into account the mean square error matrix of the channel estimation to combat the channel uncertainty. Furthermore, to reduce the effect of error propagation, we introduce a log likelihood ratio (LLR) clipping component in the MLMMSE-IC detector at the first turbo iteration. This component sets the upper bound of the LLRs according to the reliability measure of data detection, i.e., the estimated minimum bit error probability (BEP) at each subcarrier. Simulation results show that the proposed MLMMSE-IC algorithm is more robust to the channel uncertainty and achieves better performance than its conventional counterpart.

Secondly, we propose an algorithm for estimating the effective noise covariance matrix for each subcarrier under the VB framework. This algorithm can adaptively track the effective noise covariance matrix and account for the residual spatial interference, thermal noise and uncertainties from channel estimation and symbol detection. Note that, in the related previous work, noise is often assumed to be additive white Gaussian noise with constant variance [12], [13], [15]–[18], or remain stationary during one data frame [11], [14]. Similar noise variance tracking method was proposed in [21] which, however, only considers the covariance matrix caused by channel estimation errors while neglecting the one caused by the signal detection error. Simulation results show that the proposed noise estimation algorithm is effective in mitigating error propagation.

Thirdly, a robust version of the VB soft-input Kalman filter (VB-Soft-KF) algorithm, called VB-Robust-KF algorithm, is proposed based on the robust statistical theory [22] to reduce the effect of unreliable soft symbols in fast fading channels. As shown in [6]–[8], the VB-Soft-KF algorithm suffers from unreliable soft symbols, which are regarded as outliers in the KF algorithm. To solve this problem, we derive an equivalent observation equation for the VB-Soft-KF algorithm, which allows reformulating the VB-Soft-KF in a similar way to that for the standard KF. This enables us to extend the robust standard KF [23]–[25] to the VB-Robust-KF under the VB framework. Note that a similar approach of setting equivalent observation equation is also applied to the VB based data detection in this paper and to derive the soft-input Kalman filter in [5].

The rest of the paper is organized as follows. In Sec. II, the system and channel models under investigation are presented. Sec. III describes the algorithm for joint data detection, noise covariance estimation and channel estimation under the VB framework. Sec. IV shows the performance results obtained by simulations. Finally, conclusions are drawn based on the presented results in Sec. V.

Notations: $\mathbf{I}_N$ denotes an identity matrix of dimension $N \times N$. The element at the i-th row and the j-th column of matrix $\mathbf{X}(n)$ is denoted as $[\mathbf{X}(n)]_{(i,j)}$, the i-th row of matrix $\mathbf{X}(n)$ is denoted as $[\mathbf{X}(n)]_{(i,:)}$, and the i-th element of the vector $\mathbf{x}(n,k)$ is denoted by $[\mathbf{x}(n,k)]_{(i)}$; $Tr\{\cdot\}$ denotes the trace operation, $diag(\mathbf{x})$ represents a diagonal matrix with the diagonal elements taken from the vector $\mathbf{x}$, $D_{iag}(\mathbf{X})$ denotes a diagonal matrix with its diagonal elements the same as that of $\mathbf{X}$; $\otimes$ represents the kronecker product. $CN(\mathbf{m},\mathbf{C})$ denotes the complex, circularly symmetric, multivariate Gaussian probability density function (pdf) with mean $\mathbf{m}$ and a covariance matrix $\mathbf{C}$. $E_{P(x)}\{f(x)\}$ denotes the expectation of the function $f(x)$ with respect to the pdf $P(x)$. The proportionality $x \propto y$ denotes $x = \alpha y$, where α is a scalar.

II. SYSTEM DESCRIPTION

First, let us consider a MIMO-OFDM system with N_t transmit antennas, N_r receive antennas and K subcarriers. There are N_s OFDM symbols in a transmission frame. The transmitter structure is shown in Fig. 1. The input information sequence is first encoded and then bit-wise interleaved. Every m_c-tuple interleaved binary code bits are mapped to a symbol chosen from the complex-valued finite alphabet $\mathbf{\Omega}$ with 2^{m_c} possible signal points. The complex symbols are demultiplexed into N_t sub-streams through the serial-to-parallel (S/P) converter. The symbol vector $\mathbf{x}_t(n) = [x_t(n,0), \cdots, x_t(n,K-1)]^T$ is transmitted through the t-th antenna after OFDM modulation. The length of cyclic prefix (CP) is L_{CP}.

The channel is a frequency-selective fading channel, and is assumed to remain constant during an OFDM symbol period T_s , but varies from one OFDM symbol to another. The discrete channel impulse response (CIR) from the t-th transmit antenna to the r-th receive antenna during the n-th OFDM symbol is expressed as

$$\mathbf{h}_{r,t}(n) = [h_{r,t}(n,0), \cdots, h_{r,t}(n,L-1)]^T, \tag{1}$$

where L is the channel delay spread ($L \leq L_{CP}$). We assume that the channel correlation can be described by the autoregressive (AR) filter of order one [5], i.e.,

$$\mathbf{h}_r(n+1) = \mathbf{A}\mathbf{h}_r(n) + \mathbf{v}(n), \tag{2}$$

where $\mathbf{h}_r(n) = \left[\mathbf{h}_{r,1}^T(n), \cdots, \mathbf{h}_{r,N_t}^T(n)\right]^T$, $\mathbf{v}(n)$ is an $N_tL \times 1$ complex white Gaussian vector which excites the AR filter, and $\mathbf{A}$ is an $N_tL \times N_tL$ diagonal matrix containing the AR parameters. The received signal at the r-th receive antenna after CP removal and discrete fourier transform

(DFT) operation is given as

$$\begin{aligned} \mathbf{y}_r(n) &= \sum_{t=1}^{N_t} \mathbf{X}_t(n)\mathbf{F}\mathbf{h}_{r,t}(n) + \boldsymbol{\omega}_r(n) \\ &= \mathbf{X}(n)\mathbf{W}\mathbf{h}_r(n) + \boldsymbol{\omega}_r(n), \end{aligned} \tag{3}$$

where $\mathbf{y}_r(n) = [y_r(n,0), \cdots, y_r(n,K-1)]^T$ is a K-dimensional frequency domain observation vector, $\mathbf{X}(n) = [\mathbf{X}_1(n), \cdots, \mathbf{X}_{N_t}(n)]$ is constructed from $\mathbf{X}_t(n) = diag(\mathbf{x}_t(n))$, which contains the symbols transmitted from the t-th transmit antenna. In (3), $\mathbf{F}$ is constructed by the first L columns of the $K \times K$ DFT matrix and $\mathbf{W} = \mathbf{I}_{N_t} \otimes \mathbf{F}$, $\boldsymbol{\omega}_r(n) \sim CN(\mathbf{0}, \boldsymbol{\Sigma}(\mathbf{n}))$ is a K-dimensional Gaussian noise vector, where $\boldsymbol{\Sigma}(n)$ is a $K \times K$ diagonal matrix with the k-th diagonal element $\sigma^2(n,k)$, $0 \leq k \leq K-1$, which denotes the variance of the noise on the k-th subcarrier over the n-th OFDM symbol. Let's define $\mathbf{x}(n) = [\mathbf{x}_1^T(n), \cdots, \mathbf{x}_{N_t}^T(n)]^T$ and $\mathbf{H}_r(n) = [diag(\mathbf{F}\mathbf{h}_{r,1}(n)), \cdots, diag(\mathbf{F}\mathbf{h}_{r,N_t}(n))]$. It can be shown that $\mathbf{y}_r(n)$ can also be expressed as

$$\mathbf{y}_r(n) = \mathbf{H}_r(n)\mathbf{x}(n) + \boldsymbol{\omega}_{\boldsymbol{r}}(n). \tag{4}$$

Denoting $\mathbf{y}(n,k) = [y_1(n,k), \cdots, y_{N_r}(n,k)]^T$ as the received signals at the k-th subcarrier from all receive antennas, we have

$$\mathbf{y}(n,k) = \mathbf{H}(n,k)\mathbf{x}(n,k) + \boldsymbol{\omega}(n,k), \tag{5}$$

where $\boldsymbol{\omega}(n,k)$ is a $N_r \times 1$ complex Gaussian noise vector, $\mathbf{x}(n,k) = [x_1(n,k), \cdots, x_{N_t}(n,k)]^T$ is the transmitted signals at the k-th subcarrier from all the transmit antennas, and

$$\mathbf{H}(n,k) = \begin{bmatrix} H_{1,1}(n,k) & \cdots & H_{1,N_t}(n,k) \\ \vdots & \ddots & \vdots \\ H_{N_r,1}(n,k) & \cdots & H_{N_r,N_t}(n,k) \end{bmatrix}.$$

III. JOINT SIGNAL DETECTION, NOISE COVARIANCE ESTIMATION AND CHANNEL TRACKING

In this section, we derive the algorithm for achieving joint signal detection, noise covariance estimation and channel tracking based on the VB algorithm.

Let $\boldsymbol{\Phi} = \{\boldsymbol{\alpha}_l\}_{l=1}^m$ denote the vector containing all the unknown parameters to be estimated. Let $P(\boldsymbol{\Phi}\,|\boldsymbol{\theta}) = P(\boldsymbol{\alpha}_1, \cdots, \boldsymbol{\alpha}_m\,|\boldsymbol{\theta})$ denote the joint posterior distribution of $\boldsymbol{\Phi}$, given the observation $\boldsymbol{\theta}$. Since the optimal joint maximum a posteriori probability (MAP) estimation is usually too complicated to implement in practice, feasible suboptimal algorithms are therefore desirable. In this paper, we introduce the VB algorithm which approximates $P(\mathbf{\Phi}\,|\boldsymbol{\theta})$ by an auxiliary distribution $\hat{P}(\mathbf{\Phi})$, i.e., $\hat{P}(\mathbf{\Phi}) = \prod_{l=1}^{m} \hat{P}(\boldsymbol{\alpha}_l)$ under the mean-field approximation [9]. By minimizing the Kullback-Leibler (KL) divergence between $P(\mathbf{\Phi}\,|\boldsymbol{\theta})$ and $\hat{P}(\mathbf{\Phi})$, the optimized marginal distribution $\hat{P}(\boldsymbol{\alpha}_j)$, $j = 1, \cdots, m$, can be obtained by the following equations [12] (up to some additive constant)

$$\begin{aligned} \ln \hat{P}(\boldsymbol{\alpha}_j) &= E_{\hat{P}(\mathbf{\Phi}_{/j})}\{\ln P(\boldsymbol{\alpha}_1 \cdots \boldsymbol{\alpha}_m\,|\boldsymbol{\theta})\} \\ &= \int \prod\nolimits_{l=1, l\neq j}^{m} \hat{P}(\boldsymbol{\alpha}_l) \ln P(\boldsymbol{\alpha}_1 \cdots \boldsymbol{\alpha}_m\,|\boldsymbol{\theta}) d\mathbf{\Phi}_{/j}, \end{aligned} \tag{6}$$

where $\hat{P}(\mathbf{\Phi}_{/j}) = \prod_{l=1, l\neq j}^{m} \hat{P}(\boldsymbol{\alpha}_l)$. A closed form derivation of (6) is intractable, as such the optimal set of marginal distributions has to be found iteratively by alternately minimizing the KL divergence with respect to one of the distribution $\hat{P}(\boldsymbol{\alpha}_j)$, while retaining the other marginal distributions $\hat{P}(\mathbf{\Phi}_{/j})$ fixed. The iterative VB algorithm is at least guaranteed to converge to a local optimal value [26]. In the context of this work, the observations of the first n OFDM symbols are $\mathbf{Y}_0^n = \{\mathbf{Y}(l)\}_{l=0}^{n}$, where $\mathbf{Y}(l) = \{\mathbf{y}_1(l), \cdots, \mathbf{y}_{N_r}(l)\}$ and $\mathbf{y}_r(l)$ is given by (3). The unknown parameters to be estimated include the transmitted data $\mathbf{X}(n)$, the channel coefficients $\mathbf{h}(n) = \{\mathbf{h}_r(n)\}_{r=1}^{N_r}$ and the inverse of noise covariance matrix $\mathbf{\Sigma}^{-1}(n)$. After ignoring the terms independent of the unknown parameters, the joint posterior pdf $P(\mathbf{X}(n), \mathbf{h}(n), \mathbf{\Sigma}^{-1}(n)\,|\mathbf{Y}_0^n)$ can be factorized as

$$\begin{aligned} & P(\mathbf{X}(n), \mathbf{h}(n), \mathbf{\Sigma}^{-1}(n)\,|\mathbf{Y}_0^n) \\ & \propto P\left(\mathbf{Y}(n)\,|\mathbf{X}(n), \mathbf{h}(n)\,, \mathbf{\Sigma}^{-1}(n)\right) P(\mathbf{X}(n)\left|\mathbf{Y}_0^{n-1}\right.) P(\mathbf{\Sigma}^{-1}(n)\left|\mathbf{Y}_0^{n-1}\right.) P(\mathbf{h}(n)\left|\mathbf{Y}_0^{n-1}\right.). \end{aligned} \tag{7}$$

By invoking the VB algorithm, we can approximate $P(\mathbf{X}(n), \mathbf{h}(n), \mathbf{\Sigma}^{-1}(n)\,|\mathbf{Y}_0^n)$ as

$$P(\mathbf{X}(n), \mathbf{h}(n), \mathbf{\Sigma}^{-1}(n)\,|\mathbf{Y}_0^n) \approx \hat{P}(\mathbf{X}(n))\hat{P}(\mathbf{h}(n))\hat{P}(\mathbf{\Sigma}^{-1}(n)),$$

where $\hat{P}(\mathbf{X}(n))$, $\hat{P}(\mathbf{h}(n))$ and $\hat{P}(\mathbf{\Sigma}^{-1}(n))$ represent the approximations of the true marginal distributions $P(\mathbf{X}(n)\,|\mathbf{Y}_0^n)$, $P(\mathbf{h}(n)\,|\mathbf{Y}_0^n)$, and $P(\mathbf{\Sigma}^{-1}(n)\,|\mathbf{Y}_0^n)$, respectively. Next we shall describe our proposed algorithm in detail.

A. Signal Detection

According to (6), $\hat{P}\left(\mathbf{X}(n)\right)$ can be estimated based on the fixed distributions $\hat{P}\left(\mathbf{h}(n)\right)$ and $\hat{P}\left(\mathbf{\Sigma}^{-1}(n)\right)$ by solving the equation

$$\ln \hat{P}(\mathbf{X}(n)) = E_{\hat{P}(\mathbf{h}(n))} E_{\hat{P}(\mathbf{\Sigma}^{-1}(n))}\left[\ln P(\mathbf{X}(n), \mathbf{h}(n), \mathbf{\Sigma}^{-1}(n)\,|\mathbf{Y}_0^n)\right]. \tag{8}$$

Using (7) and assuming $\mathbf{X}(n)$ is independent of $\mathbf{Y}_0^{n-1}$, (8) can be reformed as

$$\ln \hat{P}(\mathbf{X}(n)) = \ln P^a(\mathbf{X}(n)) + E_{\hat{P}(\mathbf{h}(n))} E_{\hat{P}(\mathbf{\Sigma}^{-1}(n))} \left[\ln P(\mathbf{Y}(n) \,|\mathbf{X}(n), \mathbf{h}(n) \ , \mathbf{\Sigma}^{-1}(n)) \right], \tag{9}$$

where $P^a(\cdot)$ denotes the prior distribution, which is computed from the SISO channel decoder. Using (3), the log-likelihood function (LLF) in (9) can be expressed as

$$\begin{aligned} &\ln P(\mathbf{Y}(n) \,|\mathbf{X}(n), \mathbf{h}(n), \mathbf{\Sigma}^{-1}(n)) \\ &= N_r \ln \left(\left| \mathbf{\Sigma}^{-1}(n) \right| \right) - \sum_{r=1}^{N_r} \left\{ [\mathbf{y}_r(n) - \mathbf{X}(n)\mathbf{W}\mathbf{h}_r(n)]^H \mathbf{\Sigma}^{-1}(n) \, [\mathbf{y}_r(n) - \mathbf{X}(n)\mathbf{W}\mathbf{h}_r(n)] \right\}, \end{aligned} \tag{10}$$

up to some constant additive term. Assume that $\hat{P}(\mathbf{h}(n))$ has a joint Gaussian distribution with mean $\{\hat{\mathbf{h}}_r(n)\}_{r=1}^{N_r}$ and covariance matrix $\{\mathbf{\Xi}_r(n)\}_{r=1}^{N_r}$, and let $\hat{\mathbf{\Sigma}}^{-1}(n) = E_{\hat{P}(\mathbf{\Sigma}^{-1}(n))}\left[\mathbf{\Sigma}^{-1}(n)\right]$ be a diagonal matrix with diagonal elements $\hat{\sigma}^{-2}(n,k)$, $0 \le k \le K-1$. Furthermore, we assume that the transmitted symbols on different subcarriers are uncorrelated. Then, it can be shown from (9) that $\hat{P}(\mathbf{X}(n))$ can be decoupled into subcarrier components as

$$\hat{P}(\mathbf{X}(n)) = \prod_{k=0}^{K-1} \hat{P}(\mathbf{x}(n,k)).$$

For the k-th subcarrier, $\ln \hat{P}(\mathbf{x}(n,k))$ can be written as

$$\ln \hat{P}(\mathbf{x}(n,k)) = \ln P^a(\mathbf{x}(n,k)) + E_{LLF}(\mathbf{x}(n,k)) \tag{11}$$

with

$$E_{LLF}(\mathbf{x}(n,k)) = \frac{-1}{\hat{\sigma}^2(n,k)} \left\{ \left\| \mathbf{y}(n,k) - \hat{\mathbf{H}}(n,k)\mathbf{x}(n,k) \right\|^2 + \mathbf{x}^H(n,k)\mathbf{Q}(n)\mathbf{x}(n,k) \right\}. \tag{12}$$

where $E_{LLF}(\mathbf{x}(n,k))$ represents the expectation of LLF at the k-th subcarrier with respect to $\hat{P}\left(\mathbf{h}(n)\right)$ and $\hat{P}\left(\mathbf{\Sigma}^{-1}(n)\right)$. In (12), $\hat{\mathbf{H}}(n,k)$ is the mean of $\mathbf{H}(n,k)$ and its element $\hat{H}_{r,t}(n,k)$ can be calculated from the DFT of $\hat{\mathbf{h}}_r(n)$. The matrix $\mathbf{Q}(n)$ is given by

$$\mathbf{Q}(n) = \sum_{r=1}^{Nr} \begin{bmatrix} \left[\ddot{\mathbf{\Xi}}^r_{(1,1)}(n)\right]_{(k,k)} & \left[\ddot{\mathbf{\Xi}}^r_{(2,1)}(n)\right]_{(k,k)} \cdots & \left[\ddot{\mathbf{\Xi}}^r_{(N_t,1)}(n)\right]_{(k,k)} \\ \left[\ddot{\mathbf{\Xi}}^r_{(1,2)}(n)\right]_{(k,k)} & \ddots & \vdots \\ \left[\ddot{\mathbf{\Xi}}^r_{(1,N_t)}(n)\right]_{(k,k)} & \cdots & \left[\ddot{\mathbf{\Xi}}^r_{(N_t,N_t)}(n)\right]_{(k,k)} \end{bmatrix}, \tag{13}$$

where $\ddot{\mathbf{\Xi}}^r_{(t,t')}(n) = \mathbf{F}\mathbf{\Xi}_{(r,t);(r,t')}(n)\mathbf{F}^H$ denotes the frequency-domain mean square error (MSE) matrix and $\mathbf{\Xi}_{(r,t);(r,t')}(n) = E\{[\mathbf{h}_{r,t}(n) - \hat{\mathbf{h}}_{r,t}(n)][\mathbf{h}_{r,t'}(n) - \hat{\mathbf{h}}_{r,t'}(n)]^H\}$ denotes the time-domain MSE matrix , whose elements are given by $[\mathbf{\Xi}_{(r,t);(r,t')}(n)]_{(i,j)} = [\mathbf{\Xi}_r(n)]_{((t-1)L+i,(t'-1)L+j)}$ for

$1 \leq i, j \leq L$. Detailed derivations for (11) can be found in Appendix A. Note that $\{\Xi_r(n)\}_{r=1}^{N_r}$ represents the estimation errors of the CIRs, thus the uncertainties of the channel state information are incorporated in LLF for data detection. The computation of $\{\Xi_r(n)\}_{r=1}^{N_r}$ is described in details in Section III-C.

Based on the variational inference interpretation of the LMMSE-IC detector [10] and (11), we are ready to derive a modified LMMSE-IC (MLMMSE-IC) detector which considers the covariance matrix of channel estimation. To detect the t-th transmitted signal $x_t(n,k)$ on the k-th subcarrier, we first ignore the prior information of $x_t(n,k)$ and approximate the prior distribution of $\mathbf{x}(n,k)$ to be Gaussian, i.e.,

$$P^a(\mathbf{x}(n,k)) = CN(\bar{\mathbf{x}}^{(n,k)}, \bar{\mathbf{\Gamma}}^{(n,k)}), \tag{14}$$

where $\bar{\mathbf{x}}^{(n,k)} = [\bar{x}_1^{(n,k)}, \cdots, \bar{x}_{t-1}^{(n,k)}, 0, \bar{x}_{t+1}^{(n,k)}, \cdots, \bar{x}_{N_t}^{(n,k)}]$, $\bar{\mathbf{\Gamma}}^{(n,k)} = diag([V_{ar}(x_1(n,k)), \cdots,$ $V_{ar}(x_{t-1}(n,k)), 1, V_{ar}(x_{t+1}(n,k)), \cdots V_{ar}(x_{N_t}(n,k))]$. The prior expectation $\bar{x}_i^{(n,k)}$ and the corresponding variance $V_{ar}(x_i(n,k))$, $1 \leq i \leq N_t$, can be found to be

$$\begin{aligned} \bar{x}_i^{(n,k)} &= \sum_{x_m \in \mathbf{\Omega}} x_m p^a(x_i(n,k) = x_m), & (15) \\ V_{ar}(x_i(n,k)) &= \sum_{x_m \in \mathbf{\Omega}} \left| x_m - \bar{x}_i^{(n,k)} \right|^2 p^a(x_i(n,k) = x_m), & (16) \end{aligned}$$

where $p^a(x_i(n,k))$ denotes the a priori probability mass function (pmf) of $x_i(n,k)$. Under the assumption that $x_i(n,k)$ is mapped from bits $[b_i^1, \cdots, b_i^{m_c}]$, $p^a(x_i(n,k))$ can be written as

$$p^a(x_i(n,k)) = \prod\nolimits_{m=1}^{m_c} p^a(b_i^m), \tag{17}$$

where $p^a(b_i^m)$ denotes the a priori pmf of bit b_i^m which is obtained from the channel decoder. Substituting (14) in (11), the posterior distribution of $\mathbf{x}(n,k)$, which does not consider the prior information of $x_t(n,k)$, can be written as

$$P(\mathbf{x}(n,k)) = CN(\breve{\mathbf{x}}^{(n,k)}, \mathbf{\Gamma}^{(n,k)}). \tag{18}$$

In (18), the covariance matrix $\mathbf{\Gamma}^{(n,k)}$ is obtained by taking the second derivative of (11) with respect to $\mathbf{x}(n,k)$, which is given by

$$\mathbf{\Gamma}^{(n,k)} = \left[\frac{1}{\hat{\sigma}^2(n,k)} \left(\hat{\mathbf{H}}(n,k)^H \hat{\mathbf{H}}(n,k) + \mathbf{Q}(n) \right) + \left(\bar{\mathbf{\Gamma}}^{(n,k)} \right)^{-1} \right]^{-1}, \tag{19}$$

The mean $\breve{\mathbf{x}}^{(n,k)}$ is obtained by setting the first order partial derivative of (11) with respect to the $\mathbf{x}(n,k)$ to zero, which is written as

$$\breve{\mathbf{x}}^{(n,k)} = \mathbf{\Gamma}^{(n,k)} \left(\frac{1}{\hat{\sigma}^2(n,k)} \hat{\mathbf{H}}(n,k)^H \mathbf{y}(n,k) + \left(\bar{\mathbf{\Gamma}}^{(n,k)}\right)^{-1} \bar{\mathbf{x}}^{(n,k)} \right). \tag{20}$$

Note that, by setting $\mathbf{y}_{eq}^{(n,k)} = [\mathbf{y}(n,k)^T \mathbf{0}_{N_t \times 1}^T]^T$ and $\hat{\mathbf{H}}_{eq}^{(n,k)} = [\ \hat{\mathbf{H}}(n,k)^T \quad \sqrt{\mathbf{Q}(n)}^T\ ]^T$, the following equivalent observation equation has the same LLF as (12) (up to some additive constant)

$$\mathbf{y}_{eq}^{(n,k)} = \hat{\mathbf{H}}_{eq}^{(n,k)} \mathbf{x}(n,k) + \boldsymbol{\varpi}'(n,k), \tag{21}$$

where $\boldsymbol{\varpi}'(n,k) \sim CN(\mathbf{0}_{(N_t+N_r)\times 1}, \hat{\sigma}^2(n,k)\mathbf{I})$ is the virtual noise vector at the k-th subcarrier. Based on (21), it can be shown that (19) and (20) can also be obtained by the conventional interference cancellation (IC) and LMMSE filtering process [19], i.e.,

$$\begin{aligned}
\mathbf{\Gamma}^{(n,k)} &= \left[\frac{1}{\hat{\sigma}^2(n,k)} \left(\hat{\mathbf{H}}_{eq}^{(n,k)}\right)^H \hat{\mathbf{H}}_{eq}^{(n,k)} + \left(\bar{\mathbf{\Gamma}}^{(n,k)}\right)^{-1} \right]^{-1}, && (22)\\
\breve{\mathbf{x}}^{(n,k)} &= \bar{\mathbf{x}}^{(n,k)} + \bar{\mathbf{\Gamma}}^{(n,k)} \left(\hat{\mathbf{H}}_{eq}^{(n,k)}\right)^H \\
&\quad \cdot \left[\hat{\sigma}^2(n,k)\mathbf{I} + \hat{\mathbf{H}}_{eq}^{(n,k)} \bar{\mathbf{\Gamma}}^{(n,k)} \left(\hat{\mathbf{H}}_{eq}^{(n,k)}\right)^H \right]^{-1} (\mathbf{y}_{eq}^{(n,k)} - \hat{\mathbf{H}}_{eq}^{(n,k)} \bar{\mathbf{x}}^{(n,k)}). && (23)
\end{aligned}$$

Let $[\breve{\mathbf{x}}^{(n,k)}]_t$ denote the t-th element of $\breve{\mathbf{x}}^{(n,k)}$ and $[\mathbf{\Gamma}^{(n,k)}]_{t,t}$ denote the t-th diagonal element of $\mathbf{\Gamma}^{(n,k)}$, the marginal distribution of $x_t(n,k)$ can be written as

$$P(x_t(n,k)) = CN([\breve{\mathbf{x}}^{(n,k)}]_t, [\mathbf{\Gamma}^{(n,k)}]_{t,t}). \tag{24}$$

Since (24) is derived without using the prior information of $x_t(n,k)$, we have

$$P(\mathbf{y}(n,k)\,|x_t(n,k)\,) \propto P(x_t(n,k)). \tag{25}$$

Using Bayes's rule, we can further express the posterior distribution of $x_t(n,k)$ as

$$\hat{P}\left(x_t(n,k)\right) = P(\mathbf{y}(n,k)\,|x_t(n,k)\,) p^a(x_t(n,k)). \tag{26}$$

After deriving $\hat{P}\left(x_t(n,k)\right)$, the mean and variance of $x_t(n,k)$ can be calculated, respectively, as

$$\begin{aligned}
\hat{x}_t(n,k) &= \sum_{x \in \mathbf{\Omega}} x \hat{P}\left(x_t(n,k) = x\right), && (27)\\
\varepsilon_t^2(n,k) &= E_{\hat{P}(x_t(n,k))}\{|x_t(n,k) - \hat{x}_t(n,k)|^2\}, && (28)
\end{aligned}$$

where $\hat{x}_t(n,k)$ represents the soft symbol estimation of $x_t(n,k)$, the corresponding MSE $\varepsilon_t^2(n,k)$ indicates the uncertainty of the soft symbol. The extrinsic LLRs of the coded bits $L^e(b_t^m)$, $1 \leq m \leq m_c$, can be computed as

$$\begin{aligned} L^e(b_t^m) &= \ln \frac{\sum\limits_{x_t \in \mathbf{\Omega}_m^+} P(\mathbf{y}(n,k) \,| x_t(n,k)) p^a(x_t(n,k))}{\sum\limits_{x_t \in \mathbf{\Omega}_m^-} P(\mathbf{y}(n,k) \,| x_t(n,k)) p^a(x_t(n,k))} - \ln \frac{p^a(b_t^m = 1)}{p^a(b_t^m = -1)} \\ &= \ln \frac{\sum\limits_{x_t \in \mathbf{\Omega}_m^+} \hat{P}(x_t(n,k)) \prod\limits_{i=1, i \neq m}^{m_c} p^a(b_t^i)}{\sum\limits_{x_t \in \mathbf{\Omega}_m^-} \hat{P}(x_t(n,k)) \prod\limits_{i=1, i \neq m}^{m_c} p^a(b_t^i)}, \end{aligned}$$

where $\mathbf{\Omega}_m^+$ and $\mathbf{\Omega}_m^-$ denote the subsets of $\mathbf{\Omega}$ with $b_t^m = +1$ and $b_t^m = -1$, respectively. The exchange of extrinsic information between the MLMMSE-IC detector and channel decoder obeys turbo principle, as depicted in Fig. 2. At each iteration, the SISO decoder takes in the deinterleaved extrinsic LLRs $L^e(\mathbf{b})$ for the coded bits $\mathbf{b}$ from the SISO detector and computes the updated extrinsic LLRs $L_{Dec}^e(\mathbf{b})$, which is then interleaved and sent back to the SISO detector as the a priori information $L^a(\mathbf{b})$ for the next turbo iteration.

In the first turbo iteration, $L^a(\mathbf{b}) = \mathbf{0}$ and $\bar{\mathbf{x}}^{(n,k)} = \mathbf{0}_{N_t \times 1}$. In this case, as seen in (22) and (23), the MLMMSE-IC detector degenerates to a MMSE linear soft detector based on effective observation model (21). We can estimate the detector's performance at the k-th subcarrier by evaluating its BEP. According to [27], the maximum output Signal-to-Noise Ratio (SNR) of the MMSE data estimate $[\breve{\mathbf{x}}^{(n,k)}]_t$, $1 \leq t \leq N_t$, at the k-th subcarrier can be calculated as

$$\gamma_{\max}^{(n,k)} = \frac{1}{\min\limits_{1 \leq t \leq N_t} \left\{ \left[\left(\mathbf{I} + \frac{1}{\hat{\sigma}^2(n,k)} \left(\hat{\mathbf{H}}_{eq}^{(n,k)} \right)^H \hat{\mathbf{H}}_{eq}^{(n,k)} \right)^{-1} \right]_{t,t} \right\}} - 1. \tag{29}$$

Based on (29), the minimum BEP in the case of QPSK modulation can be obtained as [28]

$$P_{\min}^{(n,k)} = P_{QPSK}^b \approx \frac{1}{2} erfc \left(\sqrt{\gamma_{\max}^{(n,k)} / 2} \right).$$

In the case of 16QAM modulation, the minimum BEP can be approximated as [28]

$$P_{\min}^{(n,k)} = P_{16QAM}^b \approx \frac{1}{4} \left[1 - (1 - P_{4PAM}^s)^2 \right],$$

where $P_{4PAM}^s = \frac{3}{4} erfc \left(\sqrt{\gamma_{\max}^{(n,k)} / 10} \right)$ denotes the symbol error rate of 4PAM modulation with SNR $\gamma_{\max}^{(n,k)}$. In order to reduce the error propagation, it is desirable to limit the magnitude of the

LLRs calculated from those subcarriers with high estimated BEP. Note that the error probability of bit b with a posteriori LLR $L(b)$ can be written as [29]

$$P^b(b) = (1 + \exp(|L(b)|))^{-1}. \tag{30}$$

With (30) and using $P_{\min}^{(n,k)}$ as the reliability measure of the detector, we can adaptively set the LLR clipping value $L_c^{(n,k)}$ for the kth subcarrier as

$$L_c^{(n,k)} = \ln(1 \Big/ P_{\min}^{(n,k)} - 1). \tag{31}$$

Note that this LLR clipping value is only used in the first turbo iteration because, in this case, there is no feedback information from channel decoder and the channel estimation results are usually poor. The possible over-optimistic LLRs need to be constrained to an appropriate level by the clipping value. After the first turbo iteration, clipping is no longer applied in order to fully utilize channel decoder's error correction capability. In the rest of this paper, we will use MLMMSE-IC to represent the proposed modified LMMSE-IC detector with LLR clipping at the first iteration.

B. Estimation of Noise Covariance Matrix Distribution

According to (6) and (7), $\hat{P}(\boldsymbol{\Sigma}^{-1}(n))$ is updated based on the fixed $\hat{P}(\mathbf{X}(n))$ and $\hat{P}(\mathbf{h}(n))$ as

$$\begin{aligned} \ln \hat{P}(\boldsymbol{\Sigma}^{-1}(n)) &= E_{\hat{P}(\mathbf{h}(n))} E_{\hat{P}(\mathbf{X}(n))} \left[\ln P(\mathbf{X}(n), \mathbf{h}(n), \boldsymbol{\Sigma}^{-1}(n) \,|\mathbf{Y}_0^n) \right] \\ &= \ln P(\boldsymbol{\Sigma}^{-1}(n) \left| \mathbf{Y}_0^{n-1} \right.) + E_{\hat{P}(\mathbf{h}(n))} E_{\hat{P}(\mathbf{X}(n))} \left[\ln P(\mathbf{Y}(n) \left| \mathbf{X}(n), \mathbf{h}(n), \boldsymbol{\Sigma}^{-1}(n) \right.) \right]. \end{aligned} \tag{32}$$

In (32), the expectation of LLF with respect to the distribution function $\hat{P}(\mathbf{h}(n))$ and $\hat{P}(\mathbf{X}(n))$ can be expressed as

$$E_{\hat{P}(\mathbf{h}(n))} E_{\hat{P}(\mathbf{X}(n))} \left[\ln P(\mathbf{Y}(n) \left| \mathbf{X}(n), \mathbf{h}(n), \boldsymbol{\Sigma}^{-1}(n) \right.) \right] = N_r \ln \left(\left| \boldsymbol{\Sigma}^{-1}(n) \right| \right) - Tr \left[\boldsymbol{\Sigma}^{-1}(n) \boldsymbol{\Theta} \right], \tag{33}$$

where

$$\begin{aligned}\mathbf{\Theta} = \sum_{r=1}^{N_r} \Bigg\{ & \left[\mathbf{y}_r(n) - \sum_{t=1}^{N_t} \hat{\mathbf{X}}_t(n)\mathbf{F}\hat{\mathbf{h}}_{r,t}(n)\right]\left[\mathbf{y}_r(n) - \sum_{t=1}^{N_t} \hat{\mathbf{X}}_t(n)\mathbf{F}\hat{\mathbf{h}}_{r,t}(n)\right]^H \\ & + \sum_{t=1}^{N_t}\sum_{t'=1}^{N_t} \hat{\mathbf{X}}_t(n)\ddot{\mathbf{\Xi}}^r_{(t,t')}(n)\hat{\mathbf{X}}^H_{t'}(n) + \sum_{t=1}^{N_t} D_{iag}\left(\mathbf{F}\hat{\mathbf{h}}_{r,t}(n)\hat{\mathbf{h}}^H_{r,t}(n)\mathbf{F}^H\right)\mathbf{D}_t(n) \\ & + \sum_{t=1}^{N_t} D_{iag}\left(\ddot{\mathbf{\Xi}}^r_{(t,t)}(n)\right)\mathbf{D}_t(n)\Bigg\},\end{aligned} \tag{34}$$

with $\mathbf{D}_t(n) = diag\left([\varepsilon_t^2(n,0), \cdots, \varepsilon_t^2(n,K-1)]\right)$ and $\hat{\mathbf{X}}_t(n) = E_{\hat{P}(\mathbf{X}(n))}[\mathbf{X}_t(n)]$, which can be computed from (27-28). The mean $\hat{\mathbf{h}}_{r,t}(n)$ and covariance matrix $\ddot{\mathbf{\Xi}}^r_{(t,t')}(n)$ can be obtained from $\hat{P}\left(\mathbf{h}(n)\right)$. The detailed derivations are given in Appendix B.

Assume $P(\mathbf{\Sigma}^{-1}(n)\left|\mathbf{Y}_0^{n-1}\right.)$ obeys the Wishart distribution [11] as $P(\mathbf{\Sigma}^{-1}(n)\left|\mathbf{Y}_0^{n-1}\right.) \sim W_K(\tilde{\boldsymbol{o}}_n, \tilde{\mathbf{\Sigma}}^{-1}(n)/\tilde{\boldsymbol{o}}_n)$, i.e.,

$$\ln\left(P\left(\mathbf{\Sigma}^{-1}(n)\left|\mathbf{Y}_0^{n-1}\right.\right)\right) \propto (\tilde{\boldsymbol{o}}_n - K - 1)\ln\left|\mathbf{\Sigma}^{-1}(n)\right| - Tr[\tilde{\boldsymbol{o}}_n\tilde{\mathbf{\Sigma}}(n)\mathbf{\Sigma}^{-1}(n)], \tag{35}$$

where $W_K(\boldsymbol{o}, \mathbf{\Sigma})$ denotes a Wishart distribution with the $K \times K$ precision matrix $\mathbf{\Sigma}$ and $\boldsymbol{o}$ degree of freedom. Upon applying (33) and (35) to (32), we can show that $\hat{P}(\mathbf{\Sigma}^{-1}(n))$ also follows the Wishart distribution $\hat{P}(\mathbf{\Sigma}(n)^{-1}) \sim W_K(\boldsymbol{o}_n, \hat{\mathbf{\Sigma}}^{-1}(n)/\boldsymbol{o}_n)$, where

$$\begin{aligned}\boldsymbol{o}_n &= \tilde{\boldsymbol{o}}_n + N_r, \\ \hat{\mathbf{\Sigma}}^{-1}(n) &= \frac{\boldsymbol{o}_n}{\mathbf{\Theta} + \tilde{\boldsymbol{o}}_n\tilde{\mathbf{\Sigma}}(n)}.\end{aligned} \tag{36}$$

Let $\hat{P}(\mathbf{\Sigma}^{-1}(n-1)) \sim W_K(\boldsymbol{o}_{n-1}, \hat{\mathbf{\Sigma}}^{-1}(n-1)/\boldsymbol{o}_{n-1})$ be the pdf of the inverse noise covariance matrix corresponding to the $(n-1)$th OFDM symbol. A heuristic dynamics [21] is adopted to perform the one-step prediction of the inverse noise covariance matrix by assuming $\tilde{\boldsymbol{o}}_n = \boldsymbol{o}_{n-1}$ and $\tilde{\mathbf{\Sigma}}^{-1}(n) = \hat{\mathbf{\Sigma}}^{-1}(n-1), n = 1, \cdots, N_s$.

With the aid of the above approach, we have completed the update-prediction procedure for the effective noise covariance matrix. It can be shown from (34) that the proposed approach takes into account the residual spatial interference covariance matrix, the MSE matrices of estimation caused by the imperfect channel estimation, signal detection as well as their combined effects. As will be shown in our simulations, the effective noise estimation plays an important role to mitigate the effect of error propagation, and it can also be regarded as a reliability indicator, from which we can estimate the BEP of the detector at each subcarrier.

C. Tracking of Channel Impulse Response

Similar to (32), the distribution of CIR $\mathbf{h}(n)$ is updated based on (6), while keeping the distributions $\hat{P}(\mathbf{X}(n))$ and $\hat{P}(\mathbf{\Sigma}^{-1}(n))$ fixed, we have

$$\begin{aligned}\ln \hat{P}\left(\mathbf{h}(n)\right) &= E_{\hat{P}(\mathbf{X}(n))} E_{\hat{P}(\mathbf{\Sigma}^{-1}(n))} \left[\ln P\left(\mathbf{X}(n), \mathbf{h}(n), \mathbf{\Sigma}^{-1}(n) \left| \mathbf{Y}_0^n \right.\right)\right] \\ &= \ln P(\mathbf{h}(n) \left| \mathbf{Y}_0^{n-1} \right.) + E_{\hat{P}(\mathbf{X}(n))} E_{\hat{P}(\mathbf{\Sigma}^{-1}(n))} \left[\ln P(\mathbf{Y}(n) \left| \mathbf{X}(n), \mathbf{h}(n), \mathbf{\Sigma}^{-1}(n) \right.)\right].\end{aligned} \tag{37}$$

Using the relationship $\mathbf{H}_r(n)\mathbf{x}(n) = \mathbf{X}(n)\mathbf{W}\mathbf{h}_r(n)$ and discarding the terms irrelevant to $\mathbf{h}_r(n)$, the expectation of the LLF in (37) can be derived as

$$\begin{aligned}& E_{\hat{P}(\mathbf{X}(n))} E_{\hat{P}(\mathbf{\Sigma}^{-1}(n))} \left[\ln P\left(\mathbf{Y}(n) \left| \mathbf{X}(n), \mathbf{h}(n)\right., \mathbf{\Sigma}^{-1}(n)\right)\right] \\ & \propto - E_{\hat{P}(\mathbf{X}(n))} \left\{ Tr \left[\sum_{r=1}^{N_r} \hat{\mathbf{\Sigma}}^{-1}(n) \left(\mathbf{y}_r(n) - \mathbf{H}_r(n)\mathbf{x}(n)\right) \left(\mathbf{y}_r(n) - \mathbf{H}_r(n)\mathbf{x}(n)\right)^H \right] \right\} \\ & \propto - \sum_{r=1}^{N_r} \left[\left(\mathbf{y}_r(n) - \hat{\mathbf{X}}(n)\mathbf{W}\mathbf{h}_r(n)\right)^H \hat{\mathbf{\Sigma}}^{-1}(n) \left(\mathbf{y}_r(n) - \hat{\mathbf{X}}(n)\mathbf{W}\mathbf{h}_r(n)\right) \right. \\ & \qquad \left. + \mathbf{h}_r{}^H(n)\mathbf{W}^H \left(\mathbf{I}_{N_t} \otimes \hat{\mathbf{\Sigma}}^{-1}(n)\right) \mathbf{D}(n)\mathbf{W}\mathbf{h}_r(n) \right],\end{aligned} \tag{38}$$

where $\hat{\mathbf{X}}(n) = E_{\hat{P}(\mathbf{X}(n))}[\mathbf{X}(n)]$, the element of which can be obtained from (27); $\mathbf{D}(n)$ is the covariance matrix of $\mathbf{x}(n)$ and can be expressed as

$$\begin{aligned}\mathbf{D}(n) &= E_{\hat{P}(\mathbf{X}(n))}\{[\mathbf{x}(n) - \hat{\mathbf{x}}(n)][\mathbf{x}(n) - \hat{\mathbf{x}}(n)]^H\} \\ &= diag([\varepsilon_1^2(n,0), \cdots, \varepsilon_1^2(n,K-1), \cdots, \varepsilon_{N_t}^2(n,0) \cdots \varepsilon_{N_t}^2(n,K-1)]),\end{aligned} \tag{39}$$

where $\varepsilon_t^2(n,k)$ is given by (28). From (38) one can see that the uncertainties of data detection are considered in the channel estimation under the VB framework. As exact derivation of (37) is intractable, we adopt the Gaussian distribution to approximate $P(\mathbf{h}(n) \left| \mathbf{Y}_0^{n-1} \right.)$ [12] in (37), yielding

$$\ln P(\mathbf{h}(n) \left| \mathbf{Y}_0^{n-1} \right.) \propto - \sum_{r=1}^{N_r} [\mathbf{h}_r(n) - \mathbf{h}_r(n \left| n-1 \right.)]^H \mathbf{\Xi}_r^{-1}(n \left| n-1 \right.)[\mathbf{h}_r(n) - \mathbf{h}_r(n \left| n-1 \right.)], \tag{40}$$

where $\mathbf{h}_r(n \left| n-1 \right.)$ is the conditional mean of $\mathbf{h}_r(n)$ given the observations up to the $(n-1)$th OFDM symbol; $\mathbf{\Xi}_r(n \left| n-1 \right.) = E\left[(\mathbf{h}_r(n) - \mathbf{h}_r(n \left| n-1 \right.))(\mathbf{h}_r(n) - \mathbf{h}_r(n \left| n-1 \right.))^H\right]$ is the corresponding MSE matrix, which represents the initial channel estimation error induced by the inaccurate prediction. After substituting (38)-(40) to (37), we can show that $\hat{P}(\mathbf{h}(n))$ is also joint Gaussian distribution with the pdf $\hat{P}(\mathbf{h}(n)) = \prod_{r=1}^{N_r} CN\left(\hat{\mathbf{h}}_r(n), \Xi_r(n)\right)$. Therefore, $\hat{P}(\mathbf{h}(n))$ can be obtained by the VB soft-input Kalman filter (VB-Soft-KF) algorithm [15], which can be described as follows: for the r-th receiver antenna , $r = 1, \cdots, N_r$, compute

$$\begin{aligned}\Xi_r(n) = \Big[&\Xi_r^{-1}(n\,|n-1) + \mathbf{W}^H\hat{\mathbf{X}}(n)^H\hat{\boldsymbol{\Sigma}}^{-1}(n)\hat{\mathbf{X}}(n)\mathbf{W} \\ &+ \mathbf{W}^H(\mathbf{I}_{N_t}\otimes\hat{\boldsymbol{\Sigma}}^{-1}(n))\mathbf{D}(n)\mathbf{W}\Big]^{-1}\end{aligned} \tag{41}$$

$$\hat{\mathbf{h}}_r(n) = \Xi_r(n)\left[\Xi_r^{-1}(n\,|n-1)\mathbf{h}_r(n\,|n-1) + \mathbf{W}^H\hat{\mathbf{X}}(n)^H\hat{\boldsymbol{\Sigma}}^{-1}(n)\mathbf{y}_r(n)\right] \tag{42}$$

$$\mathbf{h}_r(n+1\,|n) = \mathbf{A}\hat{\mathbf{h}}_r(n) \tag{43}$$

$$\Xi_r(n+1\,|n) = \mathbf{A}\Xi_r(n)\mathbf{A}^H + \mathbf{R}_v. \tag{44}$$

When the time-varying channels change significantly between consecutive OFDM symbols, the channel prediction may be poor which degrades the performance of the signal detector and possibly leads to a large number of unreliable soft symbols. These unreliable soft symbols act as outliers in the VB-Soft-KF algorithm and may incur severe performance degradation. Next, we discuss how to improve the VB-Soft-KF algorithm using the robust statistical theory.

Similar to the derivation of (21) in Sec. III-A, it can be shown that the following equivalent observation model has the same LLF as (38) (up to some additive constant)

$$\mathbf{y}_{eq}^r(n) = \mathbf{X}_{eq}(n)\mathbf{h}_r(n) + \boldsymbol{\omega}_r^+(n), \tag{45}$$

where $\mathbf{X}_{eq}(n) = \left[\begin{array}{cc}(\hat{\mathbf{X}}(n)\mathbf{W})^T & (\sqrt{\mathbf{D}(n)}\mathbf{W})^T\end{array}\right]^T$, $\mathbf{y}_{eq}^r(n) = \left[\begin{array}{cc}\mathbf{y}_r(n)^T & \mathbf{0}_{(N_tK\times 1)}^T\end{array}\right]^T$, and $\boldsymbol{\omega}_r^+(n) \sim CN(\mathbf{0}, \hat{\boldsymbol{\Sigma}}_+(n))$ is a $(N_tK + K)$ -length zero-mean virtual noise vector with covariance matrix $\hat{\boldsymbol{\Sigma}}_+(n) = \mathbf{I}_{N_t+1}\otimes\hat{\boldsymbol{\Sigma}}(n)$. The VB-Soft-KF algorithm expressed by (41)-(44) can also be obtained by applying the standard KF [30] to the state-space model expressed by (45) and (2). When the channel is fast fading, the soft symbols with low reliabilities render the state-space model to be an inaccurate error-in-variable model [6] and generate outliers, which may make the KF diverge from the real channel trajectory.

Note that the channel state prediction $\mathbf{h}_r(n\,|n-1)$ can be seen as the observation of the true state $\mathbf{h}_r(n)$ with prediction error $\mathbf{e}(n\,|n-1)$. Hence, we can obtain the batch-form linear regression equation as

$$\left[\begin{array}{c}\mathbf{y}_{eq}^r(n) \\ \mathbf{h}_r(n\,|n-1)\end{array}\right] = \left[\begin{array}{c}\mathbf{X}_{eq}(n) \\ \mathbf{I}_{L\times L}\end{array}\right]\mathbf{h}_r(n) + \left[\begin{array}{c}\boldsymbol{\omega}_r^+(n) \\ \mathbf{e}(n\,|n-1)\end{array}\right], \tag{46}$$

which can be expressed in a compact form as

$$\widehat{\mathbf{y}}_r(n) = \widehat{\mathbf{X}}(n)\mathbf{h}_r(n) + \widehat{\mathbf{e}}(n). \tag{47}$$

The covariance matrix of $\widehat{\mathbf{e}}(n)$ is given by

$$\mathbf{\Gamma}_{\widehat{\mathbf{e}}} = \begin{bmatrix} \hat{\mathbf{\Sigma}}_+(n) & 0 \\ 0 & \Xi_r(n\,|n-1) \end{bmatrix} = \mathbf{S}\mathbf{S}^H, \tag{48}$$

where $\Xi_r(n\,|n-1)$ is the MSE matrix of channel prediction with error vector $\mathbf{e}(n\,|n-1)$, and $\mathbf{S}$ may be obtained by Cholesky decomposition of $\mathbf{\Gamma}_{\widehat{\mathbf{e}}}$. Consequently, upon multiplying both sides of (47) by $\mathbf{S}^{-1}$ to perform prewhitening, we obtain

$$\bar{\mathbf{y}}_r(n) = \bar{\mathbf{X}}(n)\mathbf{h}_r(n) + \bar{\mathbf{e}}(n), \tag{49}$$

where $\bar{\mathbf{y}}_r(n) = \mathbf{S}^{-1}\widehat{\mathbf{y}}_r(n)$, $\bar{\mathbf{X}}(n) = \mathbf{S}^{-1}\widehat{\mathbf{X}}(n)$, and $\bar{\mathbf{e}}(n) = \mathbf{S}^{-1}\widehat{\mathbf{e}}(n)$ with $E[\bar{\mathbf{e}}(n)\bar{\mathbf{e}}(n)^H] = \mathbf{I}$. Clearly, (49) is a standard linear least squares (LS) regression problem which yields the LS estimate of $\mathbf{h}_r(n)$ as

$$\hat{\mathbf{h}}_r(n) = \left[\bar{\mathbf{X}}(n)^H\bar{\mathbf{X}}(n)\right]^{-1}\bar{\mathbf{X}}(n)^H\bar{\mathbf{y}}_r(n), \tag{50}$$

with the MSE matrix

$$\Xi_r(n) = \left[\bar{\mathbf{X}}(n)^H\bar{\mathbf{X}}(n)\right]^{-1}. \tag{51}$$

It can be easily proved that (41)-(42) obtained by the VB-Soft-KF are equivalent to (50)-(51) obtained from the LS problem given by (49). Since the LS solution suffers from the outlier problem, it can be inferred that the VB-Soft-KF algorithm has the same problem to estimate $\mathbf{h}_r(n)$. In order to mitigate the effect of outliers caused by unreliable soft symbols, the Huber M estimator can be applied instead of the LS estimation [23]–[25], [31], which finds the solutions according to the following minimization problem

$$\hat{\mathbf{h}}_r(n) = \arg\min_{\mathbf{h}} \left\{ \sum_{i=1}^{\bar{m}} \rho\left(\tilde{r}_i\right) \right\}, \tag{52}$$

where $\tilde{r}_i = [\bar{\mathbf{y}}_r(n)]_{(i)} - \breve{\mathbf{x}}_i^{\,H}\mathbf{h}$ with $\breve{\mathbf{x}}_i$ representing the i-th column of $\bar{\mathbf{X}}^H(n)$, $\bar{m}$ is the dimension of $\bar{\mathbf{y}}_r(n)$, $\rho(\cdot)$ is the Huber minimax penalty function given by

$$\rho(x) = \begin{cases} |x|^2/2, & \text{if } |x| \le \beta \\ \beta\,|x| - \beta^2/2, & \text{if } |x| > \beta \end{cases}. \tag{53}$$

When the absolute value of the residual is smaller than β, $\rho(\cdot)$ is L_2 norm. By contrast, when the absolute value of the residual is larger than β, $\rho(\cdot)$ is chosen as the L_1 norm to reduce the effect of outliers, and hence bounds the influence of low reliable data. The threshold value β results in a tradeoff between efficiency and robustness of the estimator. In this paper, we choose β to ensure the asymptotic efficiency of this estimator under normal assumptions is higher than 95% [7]. Upon solving (52) by setting the partial derivatives of the objective function on the right-hand-side of (52) with respect to $\mathbf{h}$ to zero, we obtain

$$\sum_{i=1}^{\bar{m}} -\psi(\tilde{r}_i)\breve{\mathbf{x}}_i = \mathbf{0}, \tag{54}$$

where $\psi(\cdot) = \rho'(\cdot)$ is the influence function [22], which is given by

$$\psi(x) = \begin{cases} x, & \text{if } |x| \leq \beta \\ \beta x / |x|, & \text{if } |x| > \beta \end{cases}. \tag{55}$$

(54) can be solved by using the iteratively reweighted least-squares (IRLS) approach [25]. By multiplying and dividing $\psi(\tilde{r}_i)$ in (54) by $\tilde{r}_i$, and defining a diagonal weight matrix $\mathbf{\Lambda}$ with the ith diagonal element $[\mathbf{\Lambda}]_{(i.i)} = \psi(\tilde{r}_i)/\tilde{r}_i$, (54) can be solved to update $\hat{\mathbf{h}}_r(n)$ as

$$\hat{\mathbf{h}}_r(n) = (\bar{\mathbf{X}}^H(n)\mathbf{\Lambda}\bar{\mathbf{X}}(n))^{-1}\bar{\mathbf{X}}^H(n)\mathbf{\Lambda}\bar{\mathbf{y}}_r(n), \tag{56}$$

In summary, the proposed VB-Robust-KF algorithm can be stated as follows: for the r-th receiver antenna, $r = 1, \cdots, N_r$

1) Form the equivalent observation equation (45) using $\hat{\mathbf{X}}(n)$, $\mathbf{D}(n)$ and $\hat{\mathbf{\Sigma}}^{-1}(n)$.
2) Form the equivalent standard linear LS regression equation (49) by prewhitening the batch-form linear regression equation (47), as shown in (46)-(49).
3) Initialize the trial value $\mathbf{h} = \mathbf{h}_r(n\,|n-1)$.
4) Compute the Huber weight matrix $\mathbf{\Lambda}$ based on (55).
5) Update $\hat{\mathbf{h}}_r(n)$ using (56), then set the trial value $\mathbf{h} = \hat{\mathbf{h}}_r(n)$ and go to 4).
6) Output $\hat{\mathbf{h}}_r(n)$ when the maximum number of IRLS iterations is reached. Correspondingly, the covariance matrix $\Xi_r(n)$ is approximated by (41), which is reasonable due to the fact that $\rho(\cdot)$ resembles the quadratic function [23].
7) Predict $\mathbf{h}_r(n+1\,|n)$ and $\Xi_r(n+1\,|n)$ at time $n+1$ using (43) and (44), respectively.

The updated channel estimate $\{\hat{\mathbf{h}}_r(n)\}_{r=1}^{N_r}$ and the MSE matrices $\{\Xi_r(n)\}_{r=1}^{N_r}$ are forwarded to the data detection and noise estimator in the next VB iteration.

D. Receiver Structure, Initialization and Complexity

The structure of the proposed Improved-VB receiver is illustrated in Fig. 2. There are two types of iterations in the proposed receiver. Turbo iterations are performed between the MLMMSE-IC detector and the SISO decoder, which exchange extrinsic LLRs to improve the detection and decoding performance. Within each turbo iteration, the VB assisted joint signal detection, noise covariance estimation and channel tracking are iterated several times for every OFDM symbol.

A good initialization is essential for successful implementation of the VB algorithm. In the proposed Improved-VB receiver, the phase-shift orthogonal training sequences are transmitted in the pilot slot $n = 0$, which enable the data-aided LS channel estimator to obtain the initial $\{\hat{\mathbf{h}}_r(0)\}_{r=1}^{Nr}$ [32]. With the initial estimates of channels and the known training sequence, the estimate of the inverse noise covariance matrix $\hat{\mathbf{\Sigma}}^{-1}(0)$ can be obtained by the estimator proposed in [11]. Here, we ignore channel estimation errors and assume that the initial noise variances on all subcarriers are the same. Then the distribution of $\mathbf{\Sigma^{-1}}(0)$ can be approximated as $\hat{P}(\mathbf{\Sigma^{-1}}(0)) = W_K(\boldsymbol{o}_0, \hat{\mathbf{\Sigma}}^{-1}(0)/\boldsymbol{o}_0)$ with the degree of freedom $\boldsymbol{o}_0 = K+N_r+1$. Given the initial estimated $\hat{\mathbf{\Sigma}}^{-1}(0)$, the MSE matrices $\{\Xi_r(0)\}_{r=1}^{N_r}$ for the LS channel estimation can be obtained, and we can thus approximate the distribution for $\mathbf{h}(0)$ as $\hat{P}\left(\mathbf{h}(0)\right) = \prod_{r=1}^{N_r} CN\left(\hat{\mathbf{h}}_r(0), \Xi_r(0)\right)$. For the first OFDM symbol, we set $\{\mathbf{h}_r(1\,|0)\}_{r=1}^{Nr} = \{\hat{\mathbf{h}}_r(0)\}_{r=1}^{Nr}$, $\{\Xi_r(1\,|0)\}_{r=1}^{Nr} = \{\Xi_r(0)\}_{r=1}^{Nr}$, $\vec{\boldsymbol{o}}_1 = \boldsymbol{o}_0$, and $\tilde{\mathbf{\Sigma}}(1)^{-1} = \hat{\mathbf{\Sigma}}(0)^{-1}$. For every OFDM symbol, $\hat{P}(\mathbf{h}(n))$ and $\hat{P}(\mathbf{\Sigma}^{-1}(n))$ are initialized to $P(\mathbf{h}(n)\left|\mathbf{Y}_0^{n-1}\right.)$ and $P(\mathbf{\Sigma}^{-1}(n)\left|\mathbf{Y}_0^{n-1}\right.)$, respectively, to start the VB iterations at each turbo iteration.

Considering the number of complex multiplications as a complexity metric, the inversion of an $n \times n$ matrix requires $\mathcal{O}(n^3)$ operations, and the product of an $m \times r$ matrix with a $r \times n$ matrix requires $\mathcal{O}(mrn)$ operations. The computational complexity for each update of the noise covariance estimation is of the order $\mathcal{O}(N_r N_t K^2 L)$. We assume that α IRLS iterations are performed in the proposed VB-Robust-KF algorithm for each update of channel estimation. The computational complexity for the VB-Soft-KF and the VB-Robust-KF is of the order $\mathcal{O}(N_r N_t^2 K L^2)$ and $\mathcal{O}(\alpha N_r N_t^3 K L^2)$, respectively. Note that in order to avoid numerical instabilities and decrease the complexity, we approximate the time-domain channel covariance matrix $\{\Xi_r(n)\}$ as a diagonal matrix which has been proved to be a good approximation with negligible performance loss. For the MLMMSE-IC detection at each subcarrier, the computational complexity of MMSE filter is of the order $\mathcal{O}(N_r N_t^3)$, the computation of conditional mean and variance poses a computational load roughly equal to $\mathcal{O}(N_t 2^{m_c+3})$.

IV. SIMULATION RESULTE AND DISCUSSION

In this section, the performance of the proposed algorithm is evaluated with simulations. In our studies, we assume a MIMO-OFDM system with two transmit antennas and two receive antennas. The OFDM modulation uses 64 subcarriers and a CP of length 8. A rate 1/2 convolutional coder with generator polynomial $(7,5)_8$ is used at the transmitter. Both QPSK and 16QAM modulations are considered. Each data frame includes 21 OFDM symbols, the first OFDM symbol contains the training sequences for initial channel estimation. The CIR of each transmit/receive antenna pair is assumed to have $L = 7$ paths, which follow an exponentially decaying power delay profile. The multiple paths are independent from each other and each path gain evolves according to the Jakes model [33]. In our simulation, the time-varying path coefficients are generated by the filter based method [34], where the white complex Gaussian noise is passed through the spectral filter. Two normalized doppler frequencies are considered, i.e., $f_d T_s = 0.02$ and $f_d T_s = 0.04$, where f_d is the maximum doppler frequency. Three turbo iterations are used in our simulations, since there is no apparent performance gain after 3 iterations according to our simulation results. The VB iterative procedure is terminated when the normalized difference between the CIRs' expectation of two consecutive VB iterations does not exceed a predefined threshold, such as 0.02.

In Fig. 3, we compare the performance of the iterative receivers employing different channel estimation methods when assuming QPSK modulation and 3 turbo iterations. Here the proposed MLMMSE-IC detector and effective noise covariance estimator are used for signal detection and noise estimation, respectively. Huber-LS refers to the robust LS channel tracking algorithm presented in [7]. It can be observed that the receiver employing the proposed VB-Robust-KF algorithm has a better performance than the receivers employing the VB-soft-KF algorithm and the Huber-LS algorithm at both normalized doppler frequencies, i.e., $f_d T_s = 0.02$ and $f_d T_s = 0.04$. When $f_d T_s = 0.02$, the VB-Robust-KF scheme is about 2dB better than the VB-soft-KF algorithm and about 0.5dB better than the Huber-LS algorithm at BER$= 10^{-4}$. When operated at a higher doppler rate of $f_d T_s = 0.04$, the VB-Robust-KF method achieves a

1.2dB gain compared to the Huber-LS algorithm at BER$= 10^{-3}$, while the VB-Soft-KF scheme exhibits an error floor. This follows from the fact that the VB-Soft-KF algorithm suffers from performance degradation in the presence of outliers (low-reliability soft symbols); while the VB-Robust-KF and Huber-LS estimators can downweight outliers by using the Huber cost function. The VB-Robust-KF method has better BER performance than the Huber-LS algorithm since the latter can neither use the statistical information of channel nor output channel covariance matrix to cope with the channel uncertainty. Fig. 4 shows the normalized mean-square error (NMSE) [35] performance of the different channel estimation algorithms. It can be observed that when $f_d T_s = 0.04$, the NMSE curve of the VB-Soft-KF algorithm descends slower as the SNR values increase, due to the detrimental effect of error propagation. The VB-Robust-KF algorithm achieves a lower NMSE than the VB-Soft-KF and Huber-LS algorithms. The better NMSE performance of the VB-Robust-KF results in a better BER performance, as shown in Fig. 3.

In Fig. 5 and Fig. 6, we provide simulation results to demonstrate the performance gains achieved by the proposed MLMMSE-IC detector over the conventional LMMSE-IC detector [19]. Here the proposed VB-Robust-KF estimator and effective noise covariance estimator are used for channel and noise estimation, respectively. Fig. 5 shows the BER performance of the proposed MLMMSE-IC detector and the conventional LMMSE-IC detector when assuming QPSK modulation and $f_d T_s = 0.04$. One can see that the proposed MLMMSE-IC detector performs better than the conventional LMMSE-IC detector at each turbo iteration. At a BER of 10^{-3}, the MLMMSE-IC algorithm yields a performance gain of about 1dB compared to the LMMSE-IC algorithm at the 3rd turbo iteration. Fig. 6 compares the BER performance of two detectors when assuming 16QAM modulation and $f_d T_s = 0.02$. The superiority of the MLMMSE-IC detector over the conventional LMMSE-IC detector can clearly been seen. At a BER of 10^{-3}, the MLMMSE-IC algorithm is about 2dB better than the LMMSE-IC algorithm at the 3rd turbo iteration. The better performance of MLMMSE-IC detector is anticipated since the uncertainties of channel estimation are incorporated in the detector design and the LLR clipping value is used to mitigate the error propagation in the first turbo iteration by constraining the possible over-optimistic LLRs to an appropriate level.

In Fig. 7, we compare the performance of the MIMO-OFDM systems employing different noise covariance estimation methods. Here we assume $f_d T_s = 0.02$ and 16QAM modulation.

The VB-Robust-KF algorithm is used for channel tracking and the MLMMSE-IC method is used for signal detection. The HEM noise estimator refers to the half-biased EM noise variance estimator proposed in [32] which is used to estimate the noise variance at every subcarrier. It is shown that the BER performance of the receiver that updates the noise covariance matrix according to the tracking algorithm proposed in Sec. III-B is better than all the other schemes. The receiver assuming ideal noise variance shows inferior performance than the one using the proposed tracking method. This is due to the fact that when the interference cancellation (IC) based signal detection is employed and the channel estimator works in a pilotless decision-directed mode to track the fast fading channels, the VB receiver needs to take into account the residual interference and the additional noise caused by the channel estimation error and signal detection error to combat error propagation; however, ideal noise variance only contains the thermal noise. When SNR increases, the ideal noise variance can not accommodate the large residual spatial interference which is caused by nonideal channel estimation and signal detection, and thus become too optimistic in presence of outliers. We can also observe from Fig. 7 that the receiver which estimates the noise variance assuming the reciprocal variance is chi-square distributed [11] shows severe performance degradation since the chi-square distributed noise variance estimator restricts the noise variance to be the same for all the subcarriers, even those subcarriers with large residuals, while the proposed Wishart distributed noise covariance estimator can estimate the effective noise at different subcarriers which can be regarded as a reliability indicator at each subcarrier. Therefore, Wishart distributed noise covariance model has much more flexibility to interpret outliers at different subcarriers and thus results in better performance than the chi-squared distributed noise variance model.

For an overall assessment, we compare the BER performance of the proposed Improved-VB receiver and other related receivers [7], [15]. The VB receiver is presented in [15], where list sphere decoder is used for signal detection, the forward-backward VB-soft-KF is used for channel estimation and the noise variance is assumed to be known. We set the list size to be 16 and 256 for QPSK and 16QAM modulation, respectively. The EM-HuberLS receiver is extended from the algorithm presented in [7] where a robust Huber-LS algorithm is used for channel estimation and the noise variance is estimated using the preamble data. For a fair comparison, the LMMSE-IC algorithm [19] is adopted for signal detection in the EM-HuberLS receiver.

In Fig. 8, the BER vs. SNR performance is presented for three iterative receivers, namely, the

Improved-VB, VB, EM-HuberLS receivers when assuming QPSK modulation and $f_dT_s = 0.04$. We observe from the figure that the performance of the Improved-VB receiver is significantly improved through turbo iterations. One can also see that the Improved-VB scheme achieves much better performance than the EM-HuberLS receiver. When SNR $<$13dB, the performance of the proposed Improved-VB receiver is comparable to that of the VB receiver. We reckon that this behavior stems from the combined effect of the following three factors. Firstly, the near-optimal LSD detector adopted by the VB receiver has better performance than the LMMSE-IC type detector at relative low SNR values. Secondly, the detection of the QPSK modulation signal with constant modulus is not sensitive to the amplitude degradation induced by the inaccurate channel estimation. Thus, QPSK systems are less prone to the error propagation problem than the systems which employ higher order constellations, such as 16QAM. Thirdly, at relative low SNR values, the thermal noise composes a significant part of the effective noise, which leads to a relatively minor difference between the ideal noise variance adopted by the VB receiver and the effective noise variance used by the Improved-VB receiver. When SNR is smaller than 16dB, the Improved-VB scheme outperforms the VB scheme by about 0.5dB at BER= 10^{-3} after there turbo iterations. When SNR is larger than 16dB, the BER performance of the VB receiver becomes worse as SNR increase. This follows from the fact that the VB-soft-KF channel estimation algorithm lacks robustness, and the ideal noise variance cannot accommodate the additional noise caused by the imperfect channel estimation and data detection, which become dominant at high SNRs. Fig. 9 shows the BER performance of the MIMO-OFDM systems with different receiver algorithms when assuming $f_dT_s = 0.02$ and 16QAM modulation. As shown in Fig. 9, the performance advantage of the Improved-VB method over other methods becomes more obvious when a multiamplitude, higher-order constellation (16QAM) is employed. In particular, it is observed that increasing the number of turbo iterations does not bring noticeable performance improvements for the VB receiver even at high SNRs. In this case, the use of a multiamplitude, higher-order modulation makes the soft detection more sensitive to channel estimation errors. Consequently, serious error propagation is introduced, rendering the turbo receiver ineffective. As depicted by both Fig. 8 and Fig. 9, although the EM-HuberLS receiver employs the robust Huber-LS channel estimation algorithm, it gives poor performance due to the error propagation problem inherent in the IC based LMMSE-IC detector, while the preamble aided noise variance estimation can not accommodate the residual interference and channel uncertainty to avoid the performance degradation. In the proposed Improved-VB receiver, this problem is solved by employing the improved MLMMSE-IC detector and the effective noise estimator.

In addition to the convolutional code employed in our simulations, we also tested a rate $Rc = 1/2$ turbo code which is a parallel concatenated convolutional code with constraint length of 3 and generation polynomial $g = (5, 7)_8$. The turbo decoder consists of two constituent log-maximum a posteriori decoders. The simulation results are similar to those with the convolutional code, they are thus omitted in order to conserve space.

V. Conclusions

In this paper, a novel Improved-VB iterative receiver has been presented for MIMO-OFDM systems to combat error propagation effect in fast fading channels. In our proposed receiver, a modified LMMSE-IC detector is developed based on the VB theory which considers the channel uncertainty and adopts an adaptive LLR clipping scheme according to the reliability of detection at each subcarrier. A noise covariance estimator has been derived to track the covariance matrix of the effective noise. Furthermore, a novel VB-Robust-KF algorithm based on the Huber M estimator has been proposed for channel tracking. It downweights the low reliability soft symbols to reduce error propagation, therefore considerably improves the performance of the traditional VB-Soft-KF algorithm. Simulation results show that the proposed receiver outperforms the other receivers considered, and the advantages of our proposed receiver become more explicit as the channel fading rate increases and/or the multiamplitude, higher-order modulation is adopted. The error propagation mitigation in the low SNR region will be studied in our future work. To combat the error propagation problem over fast fading channels at low SNRs, we reckon that following methods may be adopted. Firstly, we may construct a more precise channel state model to describe the channel statistics. Secondly, some other advanced signal processing algorithms can be proposed to achieve the signal detection and channel estimation. Thirdly, pilots can be transmitted more frequently or equivalently the length of frame needs to be shortened, during which the channel estimator works in the decision-directed mode. In this case, the performance improvement is achieved at the expense of additional pilot overhead.

Appendix A

Derivations for Estimating the Symbol Distribution

The expectation term in (9) with respect to $\hat{P}(\mathbf{h}(n))$ and $\hat{P}\left(\mathbf{\Sigma}^{-1}(n)\right)$ can be written as

$$
\begin{aligned}
& E_{\hat{P}(\mathbf{h}(n))} E_{\hat{P}(\mathbf{\Sigma}^{-1}(n))}\left[\ln P(\mathbf{Y}(n)\,|\mathbf{X}(n),\mathbf{h}(n)\;,\mathbf{\Sigma}^{-1}(n))\right] \\
=& E_{\hat{P}(\mathbf{h}(n))}\left\{-\sum_{r=1}^{N_r} Tr\left[\hat{\mathbf{\Sigma}}^{-1}(n)\left(\mathbf{y}_r(n)-\mathbf{X}(n)\mathbf{W}\mathbf{h}_r(n)\right)\left(\mathbf{y}_r(n)-\mathbf{X}(n)\mathbf{W}\mathbf{h}_r(n)\right)^H\right]\right\} \\
=& \sum_{k=0}^{K-1} -\frac{1}{\hat{\sigma}^2(n,k)}\sum_{r=1}^{Nr}\Bigg[y_r(n,k)y_r^*(n,k)-\sum_{t=1}^{Nt}x_t(n,k)\hat{H}_{r,t}(n,k)y_r^*(n,k) \\
& \qquad -\sum_{t=1}^{Nt}x_t^*(n,k)\hat{H}_{r,t}(n,k)^*y_r(n,k)+\sum_{t=1}^{Nt}\sum_{t'=1}^{Nt}x_t(n,k)\hat{H}_{r,t}(n,k)\hat{H}_{r,t'}(n,k)^*x_{t'}^*(n,k) \\
& \qquad +\sum_{t=1}^{Nt}\sum_{t'=1}^{Nt}x_t(n,k)\left[\ddot{\mathbf{\Xi}}_{(t,t')}^{(r)}(n)\right]_{(k,k)}x_{t'}^*(n,k)\Bigg] \\
=& \sum_{k=0}^{K-1} E_{LLF}(\mathbf{x}(n,k)).
\end{aligned}
\tag{57}
$$

Since we assume that the transmitted symbols on different subcarriers are uncorrelated, the first term in (9) can be written as

$$
\ln P^a(\mathbf{X}(n)) = \sum_{k=0}^{K-1} \ln P^a\left(\mathbf{x}(n,k)\right). \tag{58}
$$

Consequently, using (57) and (58), we can obtain (11).

Appendix B
Derivations of (33)

Using (3) and (4), the expectation of the log-likelihood function in (32) can be expressed as

$$\begin{aligned}
&E_{\hat{P}(\mathbf{h}(n))}E_{\hat{P}(\mathbf{X}(n))}\left[\ln P(\mathbf{Y}(n)\left|\mathbf{X}(n),\mathbf{h}(n),\mathbf{\Sigma}(n)^{-1}\right.)\right]\\
=&N_r\ln\left(\left|\mathbf{\Sigma}^{-1}(n)\right|\right)-E_{\hat{P}(\mathbf{h}(n))}\left\{Tr\left[\sum_{r=1}^{Nr}\mathbf{\Sigma}^{-1}(n)\Bigg(\mathbf{y}_r(n)\mathbf{y}_r{}^H(n)-\sum_{t=1}^{Nt}\hat{\mathbf{X}}_t(n)\mathbf{F}\mathbf{h}_{r,t}(n)\mathbf{y}_r^H(n)\right.\right.\\
&-\mathbf{y}_r(n)(\sum_{t=1}^{Nt}\hat{\mathbf{X}}_t(n)\mathbf{F}\mathbf{h}_{r,t}(n))^H+\sum_{t=1}^{Nt}\sum_{t'=1}^{Nt}\hat{\mathbf{X}}_t(n)\mathbf{F}\mathbf{h}_{r,t}(n)\mathbf{h}_{r,t'}^H(n)\mathbf{F}^H\hat{\mathbf{X}}_{t'}(n)^H\\
&\left.\left.+\sum_{t=1}^{Nt}D_{iag}[\mathbf{F}\mathbf{h}_{r,t}(n)\mathbf{h}_{r,t}^H(n)\mathbf{F}^H]\mathbf{D}_t(n)\Bigg)\right]\right\}.
\end{aligned} \tag{59}$$

Consequently, substituting $E_{\hat{P}(\mathbf{h}(n))}\left(\mathbf{h}_{r,t}(n)\mathbf{h}_{r,t'}^H(n)\right)=\hat{\mathbf{h}}_{r,t}(n)\hat{\mathbf{h}}_{r,t'}^H(n)+\mathbf{\Xi}_{(r,t);(r,t')}(n)$ and $E_{\hat{P}(\mathbf{h}(n))}(\mathbf{h}_{r,t}(n))=\hat{\mathbf{h}}_{r,t}(n)$ to (59), we obtain (33).

References

[1] R. v. Nee and R. Prasad, OFDM Wireless Multimedia Communications. London, U.K.: Artech House, 2000.

[2] A. J. Paulraj, D. A. Gore, R. U. Nabar, and H. Bolcskei, "An overview of MIMO communications-A key to gigabit wireless," *Proc. IEEE,* vol. 92, no. 2, pp. 198-218, Feb. 2004.

[3] K. J. Kim, J. Yue, R. A. Iltis, and J. D. Gibson, "A QRD-M/kalman filter-based detection and channel estimation algorithm for MIMO-OFDM systems," *IEEE Trans. on Wireless Commun.,* vol. 4, no. 2, pp. 710-721, Mar. 2005.

[4] T. Kashima, K. Fukawa, and H. Suzuki, "Adaptive MAP receiver via the EM algorithm and message passings for MIMO-OFDM mobile communications," *IEEE J. Select. Areas Commun.,* vol. 24, no. 3, pp. 437-447, Mar. 2006.

[5] T. Y. Al-Naffouri, "An EM-based forward-backward kalman filter for the estimation of time-variant channels in OFDM," *IEEE Trans. on Signal Processing,* vol. 55, no. 7, pp. 3924-3930, Jul. 2007.

[6] I. Nevat and J. Yuan, "Joint channel tracking and decoding for BICM-OFDM systems using consistency tests and adaptive detection selection," *IEEE Trans. Veh. Technol.,* vol. 58, no. 8, pp. 4316-4328, Oct. 2009.

[7] S. Kalyani and K. Giridhar, "Robust statistics based expectation-maximization algorithm for channel tracking in OFDM systems," in *IEEE Int. Conf. Commun.,* 2007, pp. 3051-3056.

[8] S. Kalyani and K. Giridhar, "Mitigation of error propagation in decision directed OFDM channel tracking using generalized M estimators," *IEEE Trans. Signal Process.,* vol. 55, no. 5, pp. 1659-1672, May 2007.

[9] M. Beal, *Variational Algorithms for Approximate Bayesian Inference,* Ph.D. thesis, Gatsby Computational Neuroscience Unit, London's Global University, UK, 2003.

[10] D. D. Lin and T. J. Lim, "A variational inference framework for soft-in soft-out detection in multiple-access channels," *IEEE Trans. Inf. Theory,* vol. 55, no. 5, pp. 2345-2364, May 2009.

[11] B. Hu, I. Land, L. Rasmussen, R. Piton, and B. H. Fleury, "A divergence minimization approach to joint multiuser decoding for coded CDMA," *IEEE J. Sel. Areas Commun.,* vol. 26, no. 3, pp. 432-445, Apr. 2008.

[12] S. Stefanatos and A. K. Katsaggelos, "Joint data detection and channel tracking for OFDM systems with phase noise," *IEEE Trans. Signal Process.,* vol. 56, no. 9, pp. 4230-4243, Sep. 2008.

[13] F. Li, Z. Xu, and S. Zhu, "Variational-inference-based data detection for OFDM systems with imperfect channel estimation," *IEEE Trans. Veh. Technol.,* vol. 62, no. 3, pp. 1394-1399, Mar. 2013.

[14] C. N. Manchón, G. E. Kirkelund, B. Fleury, P. Mogensen, L. Deneire, T. B. Sørensen, and C. Rom, "Interference cancellation based on divergence minimization for MIMO-OFDM receivers," in *IEEE Glob. Telecommun. Conf.,* Honolulu, USA, Nov. 30-Dec. 4, 2009, pp. 1-6.

[15] X. Zhang, D. Wang, C. Xiong, and J. Wei, "Joint symbol detection and channel tracking for MIMO-OFDM systems via the variational bayes EM algorithm," in *Proc. IEEE Int. Symp. Pers., Indoor, Mobile Radio Commun.,* Istanbul, Turkey, Sep. 2010, pp. 46-51.

[16] C. Xiong, D. Wang, X. Zhang, J. Wei, and C. Tang, "Recursive Channel Estimation Algorithms for Iterative Receiver in MIMO-OFDM Systems," in *Proc. IEEE Wireless Communications Networking Conf.,* Budapest, Hungary, Apr. 2009, pp.1-5.

[17] C. Xiong, H. Wang, X. Zhang, J. Wei, and C. Tang, "Low complexity variational bayes iterative receiver for MIMO-OFDM systems," in *Proc. IEEE Int. Conf. Commun.,* Dresden, Germany, Jun. 2009, pp.1-5.

[18] C. Xiong, X. Wang, D. Wang, and J. Wei, "Low complexity semi-blind bayesian iterative receiver for MIMO-OFDM systems " in *Proc. IEEE Global Telecommun. Conf.,* Honolulu, USA, Nov. 30-Dec. 4, 2009, pp.1-6.

[19] X. Wang and H. V. Poor, "Iterative (turbo) soft interference cancellation and decoding for coded CDMA," *IEEE Trans. Commun.,* vol. 47, no. 7, pp. 1046-1061, Jul. 1999.

[20] B. Lu and X. Wang, "Iterative receivers for multiuser space-time coding systems," *IEEE J. Sel. Areas Commun.,* vol. 18, no. 11, pp. 2322-2335, Nov. 2000.

[21] S. Särkkä and A. Nummenmaa, "Recursive noise adaptive kalman filtering by variational bayesian approximations," *IEEE Trans. Automat. Control,* vol. 54, no. 3, pp. 596-600, Mar. 2009.

[22] P. J. Huber, Robust Statistics. New York: Wiley, 1981.

[23] Z. M. Durovic and B. D. Kovacevic, "Robust estimation with unknown noise statistics," *IEEE Trans. Autom. Control,* vol. 44, no. 6, pp. 1292-1296, Jun. 1999.

[24] M. A. Gandhi and L. Mili, "Robust kalman filter based on a generalized maximum-likelihood-type estimator," *IEEE Trans. Signal Process.,* vol. 58, no. 5, pp. 2509-2520, May 2010.

[25] S. C. Chan, Z. G. Zhang, and K. W. Tse, "A new robust Kalman filter algorithm under outliers and system uncertainties," in *Proc. IEEE Int. Symp. Circuits Systems (ISCAS)*, 2005, pp. 4317-4320.

[26] R. M. Neal and G. E. Hinton, "A view of the EM algorithm that justifies incremental, sparse, and other variants," in *Learning in Graphical Models*, M. I. Jordan, Ed. Dordrecht: Kluwer Academic Publishers, 1998.

[27] Y. Jiang, M. K. Varanasi, and J. Li, "Performance Analysis of ZF and MMSE Equalizers for MIMO Systems: An In-Depth Study of the High SNR Regime," *IEEE Trans. Inf. Theory,* vol. 57, no. 4, pp. 2008-2026, Apr. 2011.

[28] J. G. Proakis, *Digital Communications*, 3rd ed. New York: McGraw-Hill , 1995.

[29] I. Land and P. A. Hoeher, "Log-likelihood values and Monte Carlo simulation-some fundamental results," in *Proc. Int. Symposium on Turbo Codes and Related Topics,* 2000.

[30] B. D. O. Anderson and J. B. Moore, *Optimal Filtering*. Englewood Cliffs, NJ: Prentice-Hall, 1979.

[31] T. Perälä and R. Pichë, "Robust extended kalman filtering in hybrid positioning applications," in *Proc. 4th Workshop on Positioning, Navigation and Communication (WPNC '07)*, Hannover, Germany, 22-22 Mar. 2007, pp. 55-63.

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

[33] W. C. Jakes, *Microwave Mobile Communications*. New York: John Wiley and Sons, 1974.

[34] T. S. Rappaport, *Wireless Communications Principles and Practice*, New York: Prentice Hall, 1996.

[35] Y. G. Li, "Simplified channel estimation for OFDM systems with multiple transmit antennas," *IEEE Trans. on Wireless Commun.,* vol. 1, no. 1, pp. 67-75, Jan. 2002.

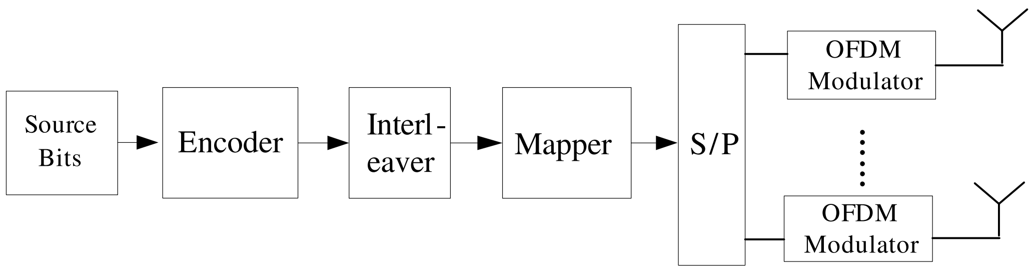

Fig. 1. Block diagram of the transmitter.

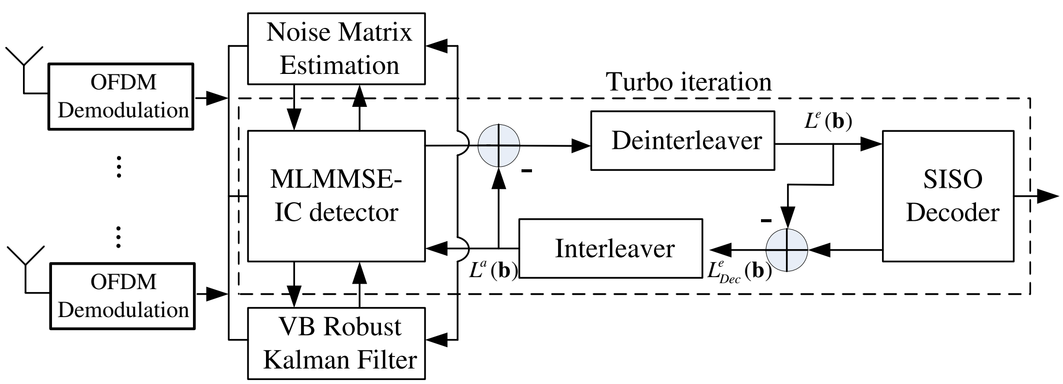

Fig. 2. Block diagram of the proposed Improved-VB receiver scheme.

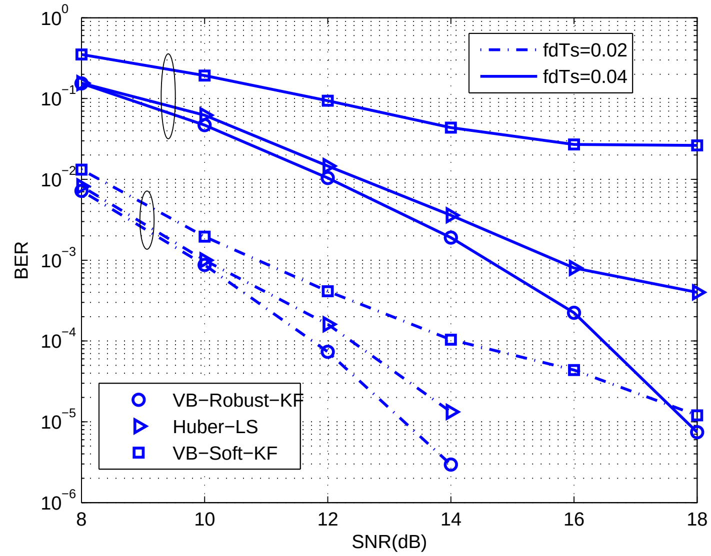

Fig. 3. BER performance comparison of the MIMO-OFDM systems employing different channel estimation methods with QPSK modulation.

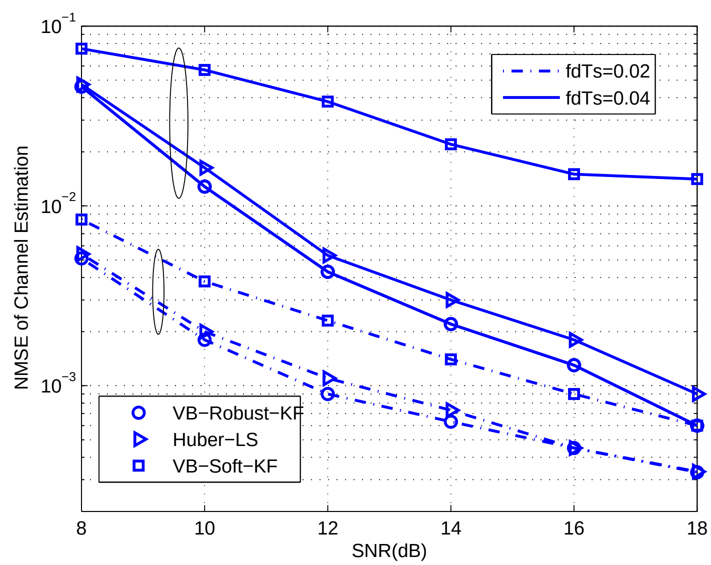

Fig. 4. Performance comparison of different channel estimators with QPSK modulation.

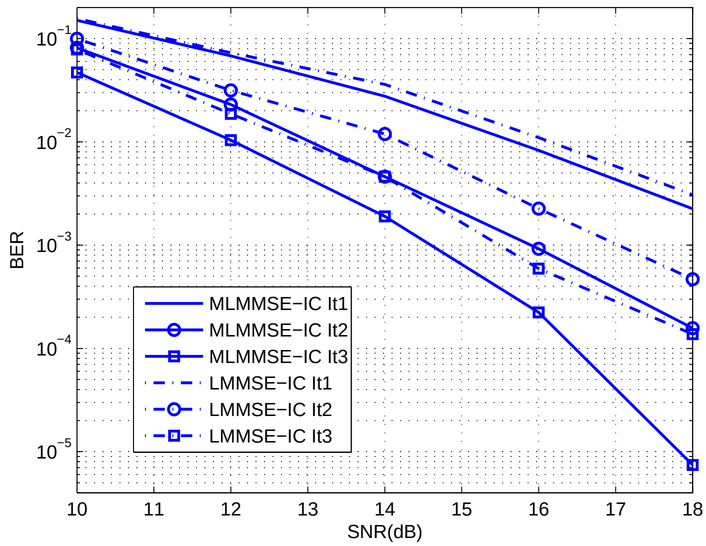

Fig. 5. BER performance comparison of different signal detectors with $f_d T_s = 0.04$ and QPSK modulation.

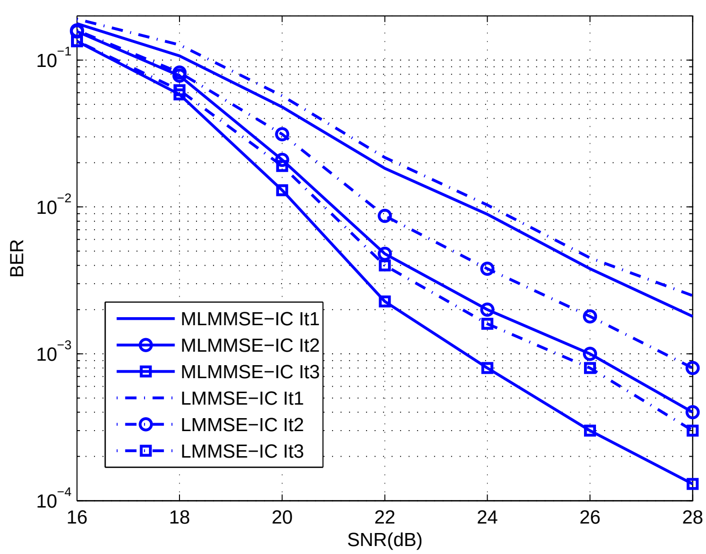

Fig. 6. BER performance comparison of different signal detectors with $f_d T_s = 0.02$ and 16QAM modulation.

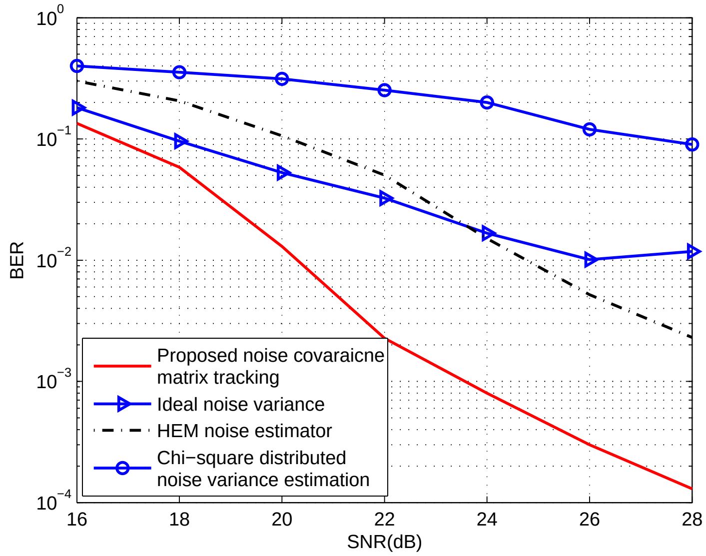

Fig. 7. Comparison of BER performance of the MIMO-OFDM systems employing different noise covariance estimation with $f_d T_s = 0.02$ and 16QAM modulation.

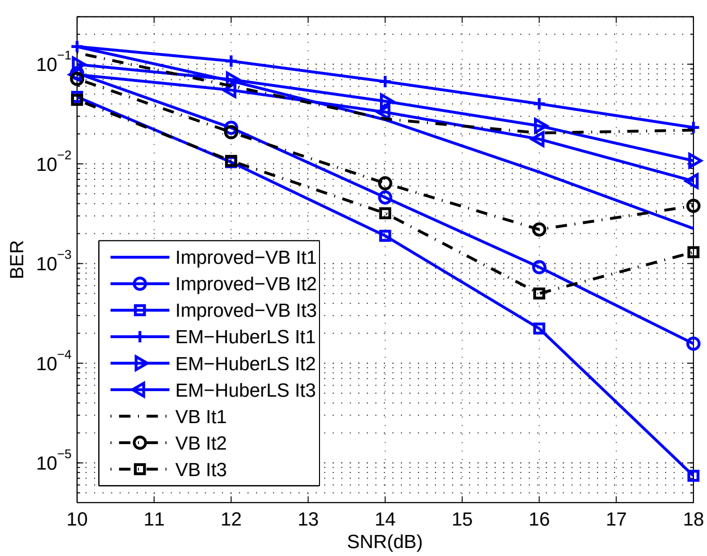

Fig. 8. BER performance comparison of different receivers with $f_d T_s = 0.04$ and QPSK modulation.

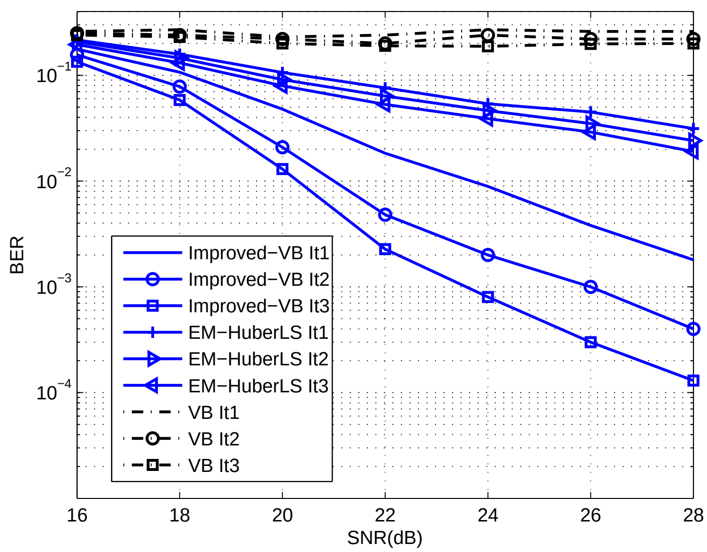

Fig. 9. BER performance comparison of different receivers with $f_d T_s = 0.02$ and 16QAM modulation.