

Semiblind Iterative Data Detection for OFDM Systems with CFO and Doubly Selective Channels

Lanlan He, Shaodan Ma, Yik-Chung Wu, and Tung-Sang Ng

Abstract—Data detection for OFDM systems over unknown doubly selective channels (DSCs) and carrier frequency offset (CFO) is investigated. A semiblind iterative detection algorithm is developed based on the expectation-maximization (EM) algorithm. It iteratively estimates the CFO, channel and recovers the unknown data using only limited number of pilot subcarriers in one OFDM symbol. In addition, efficient initial CFO and channel estimates are also derived based on approximated maximum likelihood (ML) and minimum mean square error (MMSE) criteria respectively. Simulation results show that the proposed data detection algorithm converges in a few iterations and moreover, its performance is close to the ideal case with perfect CFO and channel state information.

Index Terms—Carrier frequency offset (CFO), orthogonal frequency division multiplexing (OFDM), doubly selective channel (DSC), data detection, expectation-maximization (EM).

I. INTRODUCTION

ORTHOAGONAL frequency division multiplexing (OFDM) has been widely recognized as an efficient transmission technique for wireless communications. In OFDM systems, each subcarrier has a bandwidth narrow enough to experience flat fading, which makes the signal robust against a frequency selective fading channel. However, OFDM systems are known to be sensitive to carrier frequency offset (CFO), and frequency synchronization becomes a critical part in the design of OFDM receivers. In time-invariant frequency-selective channels (at least for one OFDM symbol), the problem of CFO, channel estimation and data detection has been addressed in [1]–[3].

Recently, there is an increasing demand for OFDM systems operating in high mobility environment, such as Digital Multimedia Broadcasting (DMB), DVB-H (Digital Video Broadcast-Handheld), Media Forward Link Only (MediaFLO) and Wireless Metropolitan Area Networks (WiMAX) [4]–[6]. For broadband OFDM systems, high speed movement of mobile terminals causes Doppler spread and results in multipath time-varying channels, i.e., doubly selective channels (DSCs). Due to DSCs, channel responses vary sample by sample, and the number of channel parameters in one OFDM symbol increases significantly, which makes the channel estimation not trivial. In this case, both the CFO and the doubly-selectivity in the channel destroy the orthogonal property

among subcarriers and induce intercarrier interference (ICI), which also complicates the data detection. Therefore, data detection for OFDM systems over unknown DSC and CFO is very challenging.

Over DSCs, data detection for systems without CFO has been discussed in [7]–[12]. In [7], [8], data detection is carried out by assuming perfectly known channel state information (CSI). However, in practice, the CSI becomes outdated very quickly and perfect CSI is generally not available. In [9]–[12], data detection is considered with unknown channel. In [9], data detection with channel estimation is developed for known symbol padding transmission systems where short sequences of known symbols acting as guard bands are inserted between successive blocks of data symbols. In [10], superimposed training is adopted for the joint channel estimation and data detection for single carrier transmission systems. However, the extension of [9], [10] to OFDM systems is by no means straightforward. For OFDM systems, data detection with channel estimation is studied through inserting training symbols periodically [11], [12]. In a slowly time-varying channel, the time correlation of the channel coefficients among OFDM symbols does not decay rapidly, and accurate channel estimates on those data symbols can be obtained. However, when the channel is varying very fast, the time correlation of the channel coefficients among OFDM symbols becomes weak, which would result in poor channel estimation on those data symbols and subsequently degrade the data detection performance severely.

In this paper, data detection for OFDM systems over unknown DSCs and CFO is addressed. Based on the expectation-maximization (EM) algorithm [13], a semiblind iterative data detection algorithm is proposed. With only limited number of pilot subcarriers in one OFDM symbol, the proposed algorithm iteratively estimates the CFO, channel and recovers the unknown data. The tentatively recovered data could then be exploited to refine the CFO and channel estimations in the next iteration. In addition, efficient initial CFO, channel and data estimators are also derived. Simulation results show that, with the efficient initialization developed in this paper, the proposed data detection algorithm is robust to Doppler spread and its performance is close to the ideal case with perfect CFO and channel state information.

Notice that EM algorithms are widely used for iterative receivers, such as [1]–[3]. However, they are all considering channels which are assumed to be constant within one OFDM symbol. In this situation, the number of unknown channel parameters is equal to the channel length, and is much smaller than that over DSCs. Together with the fact that only CFO

Paper approved by M. Juntti, the Editor for MIMO and Multiple-Access of the IEEE Communications Society. Manuscript received November 6, 2009; revised March 1, 2010.

The authors are with the Department of Electrical and Electronic Engineering, The University of Hong Kong, Hong Kong (e-mail: {llhe, sdma, ycwu, tsng}@eee.hku.hk).

The work was supported in part by the GRF, Project No. HKU 7154/08E and the HKU Seed Funding Programme, Project No. 200811159094.

Digital Object Identifier 10.1109/TCOMM.2010.092810.090682

causes ICI, derivation of EM algorithms is relatively routine. Furthermore, initialization to the EM algorithm is straightforward, as CFO and channel estimates obtained in the previous symbol can be utilized. Due to significant differences between time-invariant frequency-selective channels and DSCs, the design of the EM-based iterative receiver over DSCs is much more difficult.

The rest of the paper is organized as follows. The system model for OFDM systems over DSCs with CFO is introduced in Section II. The data detection problem is addressed in Section III. The initialization of the proposed algorithm is discussed in Section IV. Simulation results are provided in Section V to demonstrate the effectiveness of the proposed algorithm. Conclusions are drawn in Section VI.

Notation: Boldface uppercase and lowercase letters are used for matrices and vectors respectively. Superscripts H and T denote Hermitian and transpose respectively. The symbol \mathbf{I}_N denotes an $N \times N$ identity matrix, with \mathbf{e}_l being the l^{th} column of \mathbf{I}_N . The symbol $\mathbf{1}_N$ denotes an all one column vector with length N , while $\text{diag}\{\mathbf{x}\}$ stands for the diagonal matrix with vector \mathbf{x} on its diagonal. The $(m, n)^{\text{th}}$ entry of a matrix \mathbf{X} is denoted by $[\mathbf{X}]_{m,n}$. The symbol \otimes denotes the Kronecker product and \odot denotes the Hadamard product. $\mathbb{E}\{\cdot\}$ denotes the expectation. $\text{Tr}\{\mathbf{X}\}$ and $|\mathbf{X}|$ are the trace and the determinant of a square matrix \mathbf{X} respectively. $\Re\{\cdot\}$ and $\Im\{\cdot\}$ are the real and imaginary parts of the element in the brackets respectively. The matrix \mathbf{F} is the FFT matrix with $[\mathbf{F}]_{m,n} = \frac{1}{\sqrt{N}}e^{-j2\pi mn/N}$. $\lceil a \rceil$ rounds a to the nearest integer greater than or equal to a .

II. SYSTEM MODEL

In an OFDM system, the source data in frequency domain $\mathbf{x} = [x(0), \dots, x(N-1)]^T$ is modulated onto N parallel subcarriers to obtain the time domain signal $\mathbf{s} = \mathbf{F}^H \mathbf{x}$. In general, the elements of \mathbf{x} can be categorized into

$$x(k) = \begin{cases} x_p(k) & \forall k \in I_p \\ x_d(k) & \forall k \in I_d \\ 0 & \forall k \in I_v \end{cases} \quad (1)$$

where I_p is the index set of subcarriers allocated for pilot symbols with N_p elements, I_d is the index set of data subcarriers with N_d elements and I_v is the index set of subcarriers reserved for virtual subcarriers with N_v elements. Notice that $N_p + N_d + N_v = N$. The data $x_d(k)$ is independently drawn from a complex symmetrical signal constellation. Using (1), the time domain signal \mathbf{s} can be written as

$$\mathbf{s} = \mathbf{F}_p^H \mathbf{x}_p + \mathbf{F}_d^H \mathbf{x}_d \quad (2)$$

where \mathbf{F}_p collects those rows of \mathbf{F} corresponding to pilot subcarriers, while \mathbf{F}_d collects those rows of \mathbf{F} corresponding to data subcarriers, and \mathbf{x}_p and \mathbf{x}_d denote pilots and data vectors respectively.

A cyclic prefix (CP) with length longer than the delay spread of the channel is inserted at the beginning of each OFDM symbol to prevent intersymbol interference (ISI). The signal is then transmitted through a multi-path time-varying channel which has L independent taps with average power of the l^{th} tap denoted by σ_l^2 . In the case both transmitter and receiver are moving, accurate mathematical model

to describe this non-stationary channel is not available yet [14], [15]. The scenario with moving scattering is believed to be more complicated. Here, we only consider a moving receiver [10]–[12], where the channel follows Rayleigh fading and the auto-correlation of the l^{th} channel tap is given by $\mathbb{E}\{h_l(mT_s)h_l(nT_s)\} = \sigma_l^2 J_0(2\pi f_d(m-n)T_s)$ [16], with $J_0(\cdot)$ representing the zero-order Bessel function of the first kind, f_d representing the maximum Doppler shift, and T_s being the sample interval.

At the receiver side, assuming perfect timing synchronization is achieved, the n^{th} sample of the received signal is given by

$$y(n) = e^{j2\pi\epsilon n/N} \sum_{l=0}^{L-1} h_l(n)s(n-l) + w(n) \quad (3)$$

where ϵ represents CFO normalized by subcarrier spacing, $w(n)$ is the zero mean additive white Gaussian noise (AWGN) with power σ_w^2 . For notational simplicity, $h_l(n)$ is used to denote $h_l(nT_s)$. After discarding the CP, the received signal $\mathbf{y} = [y(0), y(1), \dots, y(N-1)]^T$ for a whole OFDM symbol can be expressed in a vector form as

$$\mathbf{y} = \mathbf{\Phi}(\epsilon)\mathbf{H}[\mathbf{h}]\mathbf{s} + \mathbf{w} \quad (4)$$

where $\mathbf{\Phi}(\epsilon) = \text{diag}\{1, e^{j2\pi\epsilon/N}, \dots, e^{j2\pi(N-1)\epsilon/N}\}$, $\mathbf{w} = [w(0), w(1), \dots, w(N-1)]^T$ and

$$\mathbf{H}[\mathbf{h}] = \begin{bmatrix} h_0(0) & \mathbf{0} & h_{L-1}(0) & \cdots & h_1(0) \\ h_1(1) & h_0(1) & \mathbf{0} & \cdots & h_{L-1}(1) \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \mathbf{0} & h_{L-1}(N-1) & \cdots & \cdots & h_0(N-1) \end{bmatrix}. \quad (5)$$

The channel vector \mathbf{h} is defined as

$$\mathbf{h} = [\mathbf{h}_0^T, \dots, \mathbf{h}_{L-1}^T]^T \quad (6)$$

with $\mathbf{h}_l = [h_l(0), \dots, h_l(N-1)]^T$ being the channel coefficients of the l^{th} tap for the whole OFDM symbol. Notice that, over DSCs, the channel varies sample by sample, and therefore the number of unknown channel parameter is $N \times L$.

Before the detection algorithm is presented, we note the following property, which can be proved by direct computation.

Property 1: If $\mathbf{H}[\mathbf{a}]$ is in the structure of (5), for any vector \mathbf{b} with compatible dimension with $\mathbf{H}[\mathbf{a}]$, we have

$$\mathbf{H}[\mathbf{a}]\mathbf{b} \equiv \mathbf{D}[\mathbf{b}]\mathbf{a} \quad (7)$$

where $\mathbf{D}[\mathbf{b}] = [\text{diag}\{\mathbf{\Xi}_0\mathbf{b}\}, \dots, \text{diag}\{\mathbf{\Xi}_{L-1}\mathbf{b}\}]$ with $\mathbf{\Xi}_l = [\mathbf{e}_{l+1}, \dots, \mathbf{e}_N, \mathbf{e}_1, \dots, \mathbf{e}_l]$.

With *Property 1*, an equivalent model for \mathbf{y} is given by

$$\mathbf{y} = \mathbf{\Phi}(\epsilon)\mathbf{D}[\mathbf{s}]\mathbf{h} + \mathbf{w}. \quad (8)$$

Since each channel tap is independent Gaussian distributed, the probability density function (pdf) of \mathbf{h} follows

$$p(\mathbf{h}) = \frac{1}{(\pi)^{NL} |\mathbf{R}_h|} \exp(-\mathbf{h}^H \mathbf{R}_h^{-1} \mathbf{h}) \quad (9)$$

where the correlation matrix \mathbf{R}_h is given by $\mathbf{R}_h = \mathbf{R}_L \otimes \mathbf{J}$ with $\mathbf{R}_L = \text{diag}\{\sigma_0^2, \dots, \sigma_{L-1}^2\}$ and $[\mathbf{J}]_{k,m} = J_0(2\pi f_d(m-k)T_s)$.

III. DATA DETECTION WITH DSCS AND CFO

In this section, data detection for OFDM systems with unknown channel and CFO is addressed. Since $(\varepsilon, \mathbf{h}, \mathbf{x}_d)$ are all unknown parameters, strictly speaking, they should be estimated together. With the noise being Gaussian distributed, based on (2) and (4), the likelihood function for $(\varepsilon, \mathbf{h}, \mathbf{x}_d)$ is given by

$$p(\mathbf{y}|\varepsilon, \mathbf{h}, \mathbf{x}_d) = \frac{1}{(\pi\sigma_w^2)^N} \times \exp\left(-\frac{1}{\sigma_w^2}\|\mathbf{y} - \Phi(\varepsilon)\mathbf{H}[\mathbf{h}](\mathbf{F}_p^H \mathbf{x}_p + \mathbf{F}_d^H \mathbf{x}_d)\|^2\right). \quad (10)$$

However, since the number of unknown channel parameters is $N \times L$, the joint maximum likelihood (ML) CFO, channel estimation and data detection problem over DSC based on (10) is an ill-posed one. Though the unknown channel can be directly integrated out to obtain $p(\mathbf{y}|\varepsilon, \mathbf{x}_d)$, multidimensional search is still required, since $p(\mathbf{y}|\varepsilon, \mathbf{x}_d)$ depends on data in a highly non-linear way. To overcome these problems, expectation-maximization (EM) algorithm [13] is adopted to solve the data detection problem iteratively.

In the following, an iterative data detection method is derived based on EM algorithm. For the derivation, the received signal \mathbf{y} is referred as incomplete data and we take (\mathbf{y}, \mathbf{h}) as the complete data. The proposed algorithm consists of two steps: an expectation step (E-step) and a maximization step (M-step). In the E-step, \mathbf{h} is treated as a random variable, and integrated out by taking expectation with respect to its conditional pdf given \mathbf{y} and the previous estimates of ε and \mathbf{x}_d . After integration, a new objective function depending only on ε and \mathbf{x}_d is obtained, and in the M-step, ε and \mathbf{x}_d are estimated by maximizing this objective function. The two steps at the i^{th} iteration are [13]:

E-step: Compute $Q(\varepsilon, \mathbf{x}_d|\hat{\varepsilon}^{i-1}, \hat{\mathbf{x}}_d^{i-1}) = \mathbb{E}\{\log p(\mathbf{y}, \mathbf{h}|\varepsilon, \mathbf{x}_d)|\mathbf{y}, \hat{\varepsilon}^{i-1}, \hat{\mathbf{x}}_d^{i-1}\}$;

M-step: Solve $(\hat{\varepsilon}^i, \hat{\mathbf{x}}_d^i) = \arg \max Q(\varepsilon, \mathbf{x}_d|\hat{\varepsilon}^{i-1}, \hat{\mathbf{x}}_d^{i-1})$.

The expectation in the E-step is taken with respect to the conditional pdf $p(\mathbf{h}|\mathbf{y}, \hat{\varepsilon}^{i-1}, \hat{\mathbf{x}}_d^{i-1})$, while the symbols $\hat{\varepsilon}^{i-1}$ and $\hat{\mathbf{x}}_d^{i-1}$ are the $(i-1)^{\text{th}}$ estimate of the CFO and data respectively. The derivations of the E-step and M-step are detailed as follows.

E-step:

Using Bayes' rule, we have

$$p(\mathbf{y}, \mathbf{h}|\varepsilon, \mathbf{x}_d) = p(\mathbf{y}|\varepsilon, \mathbf{h}, \mathbf{x}_d)p(\mathbf{h}) \quad (11)$$

where the pdfs $p(\mathbf{h})$ and $p(\mathbf{y}|\varepsilon, \mathbf{h}, \mathbf{x}_d)$ are specified in (9) and (10) respectively. Notice that $p(\mathbf{h})$ is independent of ε and \mathbf{x}_d . Putting (9) and (10) into (11), and dropping those terms irrelevant to ε and \mathbf{x}_d , we have

$$\log p(\mathbf{y}, \mathbf{h}|\varepsilon, \mathbf{x}_d) \propto 2\Re\{\mathbf{y}^H \Phi(\varepsilon)\mathbf{H}[\mathbf{h}](\mathbf{F}_p^H \mathbf{x}_p + \mathbf{F}_d^H \mathbf{x}_d)\} - (\mathbf{F}_p^H \mathbf{x}_p + \mathbf{F}_d^H \mathbf{x}_d)^H \mathbf{H}^H[\mathbf{h}]\mathbf{H}[\mathbf{h}](\mathbf{F}_p^H \mathbf{x}_p + \mathbf{F}_d^H \mathbf{x}_d). \quad (12)$$

According to *Property 1* and with $\mathbf{s} = \mathbf{F}_p^H \mathbf{x}_p + \mathbf{F}_d^H \mathbf{x}_d$, (12) is equivalent to

$$\log p(\mathbf{y}, \mathbf{h}|\varepsilon, \mathbf{x}_d) \propto 2\Re\{\mathbf{y}^H \Phi(\varepsilon)\mathbf{D}[\mathbf{s}]\mathbf{h}\} - \mathbf{h}^H \mathbf{D}^H[\mathbf{s}]\mathbf{D}[\mathbf{s}]\mathbf{h}. \quad (13)$$

The function $Q(\varepsilon, \mathbf{x}_d|\hat{\varepsilon}^{i-1}, \hat{\mathbf{x}}_d^{i-1})$ to be maximized is then computed as

$$Q(\varepsilon, \mathbf{x}_d|\hat{\varepsilon}^{i-1}, \hat{\mathbf{x}}_d^{i-1}) = 2\Re\{\mathbf{y}^H \Phi(\varepsilon)\mathbf{D}[\mathbf{s}]\underbrace{\mathbb{E}\{\mathbf{h}|\mathbf{y}, \hat{\varepsilon}^{i-1}, \hat{\mathbf{x}}_d^{i-1}\}}_{\triangleq \hat{\mathbf{h}}^i}\} - \text{Tr}\left\{\mathbf{D}^H[\mathbf{s}]\mathbf{D}[\mathbf{s}]\left(\underbrace{\mathbb{E}\{(\mathbf{h} - \hat{\mathbf{h}}^i)(\mathbf{h} - \hat{\mathbf{h}}^i)^H|\mathbf{y}, \hat{\varepsilon}^{i-1}, \hat{\mathbf{x}}_d^{i-1}\}}_{\triangleq \mathbf{C}_h^i}\right) + \underbrace{\mathbb{E}\{\mathbf{h}|\mathbf{y}, \hat{\varepsilon}^{i-1}, \hat{\mathbf{x}}_d^{i-1}\}}_{=\hat{\mathbf{h}}^i} \underbrace{\mathbb{E}\{\mathbf{h}^H|\mathbf{y}, \hat{\varepsilon}^{i-1}, \hat{\mathbf{x}}_d^{i-1}\}}_{=(\hat{\mathbf{h}}^i)^H}\right\}. \quad (14)$$

Here, $\mathbb{E}\{\mathbf{h}|\mathbf{y}, \hat{\varepsilon}^{i-1}, \hat{\mathbf{x}}_d^{i-1}\}$ is the conditional mean of \mathbf{h} , which will be denoted by $\hat{\mathbf{h}}^i$. Furthermore, $\mathbb{E}\{(\mathbf{h} - \hat{\mathbf{h}}^i)(\mathbf{h} - \hat{\mathbf{h}}^i)^H|\mathbf{y}, \hat{\varepsilon}^{i-1}, \hat{\mathbf{x}}_d^{i-1}\}$ is the conditional covariance matrix of \mathbf{h} , which will be denoted by \mathbf{C}_h^i .

It is shown in Appendix A that the conditional mean and covariance matrix are given by

$$\hat{\mathbf{h}}^i = \mathbf{R}_h \mathbf{D}^H[\hat{\mathbf{s}}^{i-1}](\sigma_w^2 \mathbf{I}_N + \mathbf{D}[\hat{\mathbf{s}}^{i-1}]\mathbf{R}_h \mathbf{D}^H[\hat{\mathbf{s}}^{i-1}])^{-1} \times \Phi^H(\hat{\varepsilon}^{i-1})\mathbf{y}. \quad (15)$$

and

$$\mathbf{C}_h^i = \mathbf{R}_h - \mathbf{R}_h \mathbf{D}^H[\hat{\mathbf{s}}^{i-1}] \times (\sigma_w^2 \mathbf{I}_N + \mathbf{D}[\hat{\mathbf{s}}^{i-1}]\mathbf{R}_h \mathbf{D}^H[\hat{\mathbf{s}}^{i-1}])^{-1} \mathbf{D}[\hat{\mathbf{s}}^{i-1}]\mathbf{R}_h \quad (16)$$

with $\hat{\mathbf{s}}^{i-1} = \mathbf{F}_d^H \hat{\mathbf{x}}_d^{i-1} + \mathbf{F}_p^H \mathbf{x}_p$. Moreover, $\hat{\mathbf{h}}^i$ is in fact the MAP channel estimator as a byproduct of the E-step [13]. Putting (15) and (16) into (14), the function $Q(\varepsilon, \mathbf{x}_d|\hat{\varepsilon}^{i-1}, \hat{\mathbf{x}}_d^{i-1})$ becomes

$$Q(\varepsilon, \mathbf{x}_d|\hat{\varepsilon}^{i-1}, \hat{\mathbf{x}}_d^{i-1}) = 2\Re\{\mathbf{y}^H \Phi(\varepsilon)\mathbf{D}[\mathbf{s}]\hat{\mathbf{h}}^i\} - \text{Tr}\{\mathbf{D}^H[\mathbf{s}]\mathbf{D}[\mathbf{s}](\mathbf{C}_h^i + \hat{\mathbf{h}}^i(\hat{\mathbf{h}}^i)^H)\}. \quad (17)$$

M-step:

In the M-step, we need to maximize $Q(\varepsilon, \mathbf{x}_d|\hat{\varepsilon}^{i-1}, \hat{\mathbf{x}}_d^{i-1})$ in (17) with respect to ε and \mathbf{x}_d . However, noticing that $\mathbf{D}[\mathbf{s}]$ depends on \mathbf{x}_d in an implicit way, direct maximization of (17) with respect to \mathbf{x}_d is difficult. In the following, we derive an alternative expression for $Q(\varepsilon, \mathbf{x}_d|\hat{\varepsilon}^{i-1}, \hat{\mathbf{x}}_d^{i-1})$ from which a closed-form solution for \mathbf{x}_d can be obtained.

Since \mathbf{C}_h^i is a $NL \times NL$ Hermitian matrix, based on eigen-decomposition, we have $\mathbf{C}_h^i = \sum_{m=1}^r \beta_{m,i} \mathbf{u}_{m,i} \mathbf{u}_{m,i}^H$ where r is the rank of \mathbf{C}_h^i and $\beta_{m,i}$ is the m^{th} eigenvalue of \mathbf{C}_h^i with $\mathbf{u}_{m,i}$ being the corresponding eigenvector. On the other hand, according to *Property 1*, we have $\mathbf{D}[\mathbf{s}]\hat{\mathbf{h}}^i = \mathbf{H}[\hat{\mathbf{h}}^i]\mathbf{s}$ and $\mathbf{D}[\mathbf{s}]\mathbf{u}_{m,i} = \mathbf{H}[\mathbf{u}_{m,i}]\mathbf{s}$. Therefore, $Q(\varepsilon, \mathbf{x}_d|\hat{\varepsilon}^{i-1}, \hat{\mathbf{x}}_d^{i-1})$ in (17) is equivalent to

$$Q(\varepsilon, \mathbf{x}_d|\hat{\varepsilon}^{i-1}, \hat{\mathbf{x}}_d^{i-1}) = 2\Re\{\mathbf{y}^H \Phi(\varepsilon)\mathbf{H}[\hat{\mathbf{h}}^i]\mathbf{s}\} - \sum_{m=1}^r \beta_{m,i} \mathbf{s}^H \mathbf{H}^H[\mathbf{u}_{m,i}]\mathbf{H}[\mathbf{u}_{m,i}]\mathbf{s} - \mathbf{s}^H \mathbf{H}^H[\hat{\mathbf{h}}^i]\mathbf{H}[\hat{\mathbf{h}}^i]\mathbf{s}. \quad (18)$$

Putting (2) into (18) and dropping those terms irrelevant to ε and \mathbf{x}_d , we have (19), as shown at the top of the next page. Although (19) is in quadratic form of \mathbf{x}_d , strictly speaking, its maximization is still a multidimensional search problem due to the existence of discrete data. To overcome this problem, we first relax \mathbf{x}_d to be continuous. By setting the first derivative

$$\begin{aligned}
Q(\varepsilon, \mathbf{x}_d | \hat{\varepsilon}^{i-1}, \hat{\mathbf{x}}_d^{i-1}) &= 2\Re\{\mathbf{y}^H \Phi(\varepsilon) \mathbf{H}[\hat{\mathbf{h}}^i] (\mathbf{F}_p^H \mathbf{x}_p + \mathbf{F}_d^H \mathbf{x}_d)\} - 2\Re\{\underbrace{\mathbf{x}_p^H \mathbf{F}_p \left(\sum_{m=1}^r \beta_{m,i} \mathbf{H}^H[\mathbf{u}_{m,i}] \mathbf{H}[\mathbf{u}_{m,i}] + \mathbf{H}^H[\hat{\mathbf{h}}^i] \mathbf{H}[\hat{\mathbf{h}}^i] \right) \mathbf{F}_d^H \mathbf{x}_d}_{\triangleq \mathbf{P}_i}\} \\
&\quad - \underbrace{\mathbf{x}_d^H \mathbf{F}_d \left(\sum_{m=1}^r \beta_{m,i} \mathbf{H}^H[\mathbf{u}_{m,i}] \mathbf{H}[\mathbf{u}_{m,i}] + \mathbf{H}^H[\hat{\mathbf{h}}^i] \mathbf{H}[\hat{\mathbf{h}}^i] \right) \mathbf{F}_d^H \mathbf{x}_d}_{=\mathbf{P}_i}.
\end{aligned} \tag{19}$$

of (19) with respect to \mathbf{x}_d to zero, the i^{th} relaxed data estimate which depends on ε is then given by

$$\mathbf{x}_d^i(\varepsilon) = (\mathbf{F}_d \mathbf{P}_i \mathbf{F}_d^H)^{-1} \mathbf{F}_d (\mathbf{H}^H[\hat{\mathbf{h}}^i] \Phi^H(\varepsilon) \mathbf{y} - \mathbf{P}_i \mathbf{F}_p^H \mathbf{x}_p). \tag{20}$$

This corresponds to linear data detector similar to those in [7][11][12]. Then constellation mapping is carried out to obtain the discrete data estimate as $\hat{\mathbf{x}}_d^i(\varepsilon) = \text{Qant}(\mathbf{x}_d^i(\varepsilon))$, where $\text{Qant}(\cdot)$ denotes quantization on the element in the brackets. Putting $\hat{\mathbf{x}}_d^i(\varepsilon)$ back to (19), the CFO estimate follows

$$\begin{aligned}
\hat{\varepsilon}^i &= \arg \max_{\varepsilon} 2\Re\{\mathbf{y}^H \Phi(\varepsilon) \mathbf{H}[\hat{\mathbf{h}}^i] (\mathbf{F}_p^H \mathbf{x}_p + \mathbf{F}_d^H \hat{\mathbf{x}}_d^i(\varepsilon))\} \\
&\quad - 2\Re\{\mathbf{x}_p^H \mathbf{F}_p \mathbf{P}_i \mathbf{F}_d^H \hat{\mathbf{x}}_d^i(\varepsilon)\} - (\hat{\mathbf{x}}_d^i(\varepsilon))^H \mathbf{F}_d \mathbf{P}_i \mathbf{F}_d^H \hat{\mathbf{x}}_d^i(\varepsilon),
\end{aligned} \tag{21}$$

and only one dimension search is required to get the i^{th} estimate of ε .

In summary, the proposed algorithm iterates between the E-step and M-step as follows:

E-step (update of \mathbf{h}):

$$\begin{aligned}
\hat{\mathbf{h}}^i &= \mathbf{R}_h \mathbf{D}^H[\hat{\mathbf{s}}^{i-1}] (\sigma_w^2 \mathbf{I}_N + \mathbf{D}[\hat{\mathbf{s}}^{i-1}] \mathbf{R}_h \mathbf{D}^H[\hat{\mathbf{s}}^{i-1}])^{-1} \\
&\quad \times \Phi^H(\hat{\varepsilon}^{i-1}) \mathbf{y}
\end{aligned} \tag{22}$$

with $\hat{\mathbf{s}}^{i-1} = \mathbf{F}_d^H \hat{\mathbf{x}}_d^{i-1} + \mathbf{F}_p^H \mathbf{x}_p$.

M-step (update of ε and \mathbf{x}_d):

$$\begin{aligned}
\hat{\varepsilon}^i &= \arg \max_{\varepsilon} 2\Re\{\mathbf{y}^H \Phi(\varepsilon) \mathbf{H}[\hat{\mathbf{h}}^i] (\mathbf{F}_p^H \mathbf{x}_p + \mathbf{F}_d^H \hat{\mathbf{x}}_d^i(\varepsilon))\} \\
&\quad - 2\Re\{\mathbf{x}_p^H \mathbf{F}_p \mathbf{P}_i \mathbf{F}_d^H \hat{\mathbf{x}}_d^i(\varepsilon)\} - (\hat{\mathbf{x}}_d^i(\varepsilon))^H \mathbf{F}_d \mathbf{P}_i \mathbf{F}_d^H \hat{\mathbf{x}}_d^i(\varepsilon).
\end{aligned} \tag{23}$$

and

$$\hat{\mathbf{x}}_d^i = \text{Qant}\{(\mathbf{F}_d \mathbf{P}_i \mathbf{F}_d^H)^{-1} \mathbf{F}_d (\mathbf{H}^H[\hat{\mathbf{h}}^i] \Phi^H(\hat{\varepsilon}^i) \mathbf{y} - \mathbf{P}_i \mathbf{F}_p^H \mathbf{x}_p)\} \tag{24}$$

where $\mathbf{P}_i = \sum_{m=1}^r \beta_{m,i} \mathbf{H}^H[\mathbf{u}_{m,i}] \mathbf{H}[\mathbf{u}_{m,i}] + \mathbf{H}^H[\hat{\mathbf{h}}^i] \mathbf{H}[\hat{\mathbf{h}}^i]$ with $\beta_{m,i}$ being the m^{th} eigenvalue of \mathbf{C}_h^i given in (16) and $\mathbf{u}_{m,i}$ being the corresponding eigenvector.

Remark 1: In the M-step, the rank of \mathbf{C}_h^i should be specified for the eigen-decomposition. According to (16), another equivalent form of \mathbf{C}_h^i is given by

$$\begin{aligned}
\mathbf{C}_h^i &= \mathbf{R}_h (\mathbf{I}_{NL} - \mathbf{D}^H[\hat{\mathbf{s}}^{i-1}]) \\
&\quad \times (\sigma_w^2 \mathbf{I}_N + \mathbf{D}[\hat{\mathbf{s}}^{i-1}] \mathbf{R}_h \mathbf{D}^H[\hat{\mathbf{s}}^{i-1}])^{-1} \mathbf{D}[\hat{\mathbf{s}}^{i-1}] \mathbf{R}_h.
\end{aligned} \tag{25}$$

It follows that $r = \text{rank}(\mathbf{C}_h^i) \leq \text{rank}(\mathbf{R}_h)$. Since the magnitude of the eigenvalues of \mathbf{R}_h is known to drop rapidly after $L(2\lceil N f_d T_s \rceil + 1)$ largest values [16], the magnitude of the eigenvalues of \mathbf{C}_h^i should also follow this trend and we choose $r = L(2\lceil N f_d T_s \rceil + 1)$ for the implementation.

Remark 2: The complexity of the proposed algorithm is dominated by one $N \times N$ matrix inversion on $\sigma_w^2 \mathbf{I}_N + \mathbf{D}[\hat{\mathbf{s}}^{i-1}] \mathbf{R}_h \mathbf{D}^H[\hat{\mathbf{s}}^{i-1}]$ shown in (22), a partial eigen-decomposition on $NL \times NL$ matrix \mathbf{C}_h^i to obtain \mathbf{P}_i , and one $N_d \times N_d$ matrix inversion on $\mathbf{F}_d \mathbf{P}_i \mathbf{F}_d^H$ shown in (24). For a $M \times M$ matrix, the complexity of inversion is $O(M^3)$, and the complexity of partial eigen-decomposition is $O(M^2)$ [17]. Therefore, the complexity of our proposed algorithm is $O(N^3 + N_d^3 + (NL)^2)$.

IV. INITIALIZATION

A good initialization is essential to EM algorithm, and therefore, the problem now becomes how to obtain the initial estimates of CFO and data. Strictly speaking, if the channel and data can be integrated out, the CFO estimator is given by

$$\begin{aligned}
\hat{\varepsilon} &= \arg \max p(\mathbf{y}|\varepsilon) \\
&= \arg \max \int p(\mathbf{y}|\varepsilon, \mathbf{h}, \mathbf{x}_d) p(\mathbf{h}) p(\mathbf{x}_d) d\mathbf{h} d\mathbf{x}_d.
\end{aligned} \tag{26}$$

However, the integration is difficult to compute due to the discrete nature of the data \mathbf{x}_d . Notice that the received signal is given by $y(n) = e^{j2\pi\varepsilon n/N} \sum_{l=0}^{L-1} h_l(n) s(n-l) + w(n)$. Since each tap of the channel is i.i.d., when the channel length L is large enough, according to the central limit theory, $\sum_{l=0}^{L-1} h_l(n) s(n-l)$ is approximately Gaussian. Together with $w(n)$ being additive white Gaussian noise and independent of the channel and data, $p(\mathbf{y}|\varepsilon)$ is approximately Gaussian and thus can be properly characterized by its mean $\mathbb{E}\{\mathbf{y}|\varepsilon\}$ and correlation matrix $\mathbb{E}\{\mathbf{y}\mathbf{y}^H|\varepsilon\}$.

From (2) and (4), the mean can be computed as

$$\begin{aligned}
\mathbb{E}\{\mathbf{y}|\varepsilon\} &= \mathbb{E}_{\mathbf{h}, \mathbf{x}_d, \mathbf{w}}\{\Phi(\varepsilon) \mathbf{H}[\mathbf{h}] \mathbf{s} + \mathbf{w}|\varepsilon\} \\
&= \mathbb{E}_{\mathbf{h}}\{\Phi(\varepsilon) \mathbf{D}[\mathbf{F}_p^H \mathbf{x}_p] \mathbf{h}|\varepsilon\} = \mathbf{0}
\end{aligned} \tag{27}$$

where the second equality comes from *Property 1* after expectation with respect to \mathbf{x}_d and \mathbf{w} , and the third equality is based on expectation with respect to \mathbf{h} .

On the other hand, after lengthy but straightforward computation (shown in Appendix B), the covariance matrix, i.e., the correlation matrix in this case, is given by

$$\begin{aligned}
\mathbb{E}\{\mathbf{y}\mathbf{y}^H|\varepsilon\} &= \Phi(\varepsilon) (\mathbf{D}[\mathbf{F}_p^H \mathbf{x}_p] \mathbf{R}_h \mathbf{D}^H[\mathbf{F}_p^H \mathbf{x}_p] \\
&\quad + \mathbf{J} \odot (\mathbf{F}_d^H \mathbf{\Lambda}_d \mathbf{F}_d) + \sigma_w^2 \mathbf{I}_N) \Phi^H(\varepsilon)
\end{aligned} \tag{28}$$

where $\mathbf{\Lambda}_d = \mathbb{E}\{\mathbf{x}_d \mathbf{x}_d^H\}$ is a diagonal matrix. Therefore, based on the approximated Gaussian distribution $p(\mathbf{y}|\varepsilon)$ with zero

mean and covariance matrix given by (28), the initial CFO estimator can be derived as

$$\begin{aligned} \hat{\varepsilon}^0 &= \arg \min_{\varepsilon} \mathbf{y}^H \Phi(\varepsilon) (\mathbf{D}[\mathbf{F}_p^H \mathbf{x}_p] \mathbf{R}_h \mathbf{D}^H [\mathbf{F}_p^H \mathbf{x}_p] \\ &+ \mathbf{J} \odot (\mathbf{F}_d^H \mathbf{\Lambda}_d \mathbf{F}_d) + \sigma_w^2 \mathbf{I}_N)^{-1} \Phi^H(\varepsilon) \mathbf{y}. \end{aligned} \quad (29)$$

After CFO estimation, initial channel estimation is now considered. The system model in (4) can be rewritten as

$$\mathbf{y} = \Phi(\varepsilon) \mathbf{D}[\mathbf{F}_p^H \mathbf{x}_p] \mathbf{h} + \Phi(\varepsilon) \mathbf{H}[\mathbf{h}] \mathbf{F}_d^H \mathbf{x}_d + \mathbf{w}. \quad (30)$$

By treating the term containing \mathbf{x}_d as interference, the MMSE channel estimate is then given by

$$\begin{aligned} \hat{\mathbf{h}}^0 &= \mathbf{R}_h \mathbf{D}^H [\mathbf{F}_p^H \mathbf{x}_p] (\mathbf{D}[\mathbf{F}_p^H \mathbf{x}_p] \mathbf{R}_h \mathbf{D}^H [\mathbf{F}_p^H \mathbf{x}_p] \\ &+ \mathbf{J} \odot (\mathbf{F}_d^H \mathbf{\Lambda}_d \mathbf{F}_d) + \sigma_w^2 \mathbf{I}_N)^{-1} \Phi^H(\hat{\varepsilon}^0) \mathbf{y}. \end{aligned} \quad (31)$$

Based on MMSE criterion and with the initial CFO and channel estimates given in (29) and (31) respectively, the data detection directly follows from (30) as

$$\begin{aligned} \hat{\mathbf{x}}_d^0 &= \text{Qant}\{(\mathbf{F}_d \mathbf{H}^H [\hat{\mathbf{h}}^0] \mathbf{H} [\hat{\mathbf{h}}^0] \mathbf{F}_d^H + \sigma_w^2 \mathbf{\Lambda}_d^{-1})^{-1} \\ &\times \mathbf{F}_d \mathbf{H}^H [\hat{\mathbf{h}}^0] (\Phi^H(\hat{\varepsilon}^0) \mathbf{y} - \mathbf{D}[\mathbf{F}_p^H \mathbf{x}_p] \hat{\mathbf{h}}^0)\}. \end{aligned} \quad (32)$$

Remark 3: In the extreme case that all data are training, \mathbf{x}_d becomes known and $p(\mathbf{y}|\varepsilon)$ is definitely Gaussian. The initial CFO and channel estimator in (29) and (31) reduce to the optimal estimators proposed in [18].

Remark 4: In time-invariant case, channel is constant during each OFDM symbol, i.e., $f_d = 0$ and then \mathbf{J} is an all-one matrix. Accordingly, (28) reduces to

$$\begin{aligned} \mathbb{E}\{\mathbf{y}\mathbf{y}^H|\varepsilon\} &= \Phi(\varepsilon) (\mathbf{D}[\mathbf{F}_p^H \mathbf{x}_p] \mathbf{R}_h \mathbf{D}^H [\mathbf{F}_p^H \mathbf{x}_p] \\ &+ \mathbf{F}_d^H \mathbf{\Lambda}_d \mathbf{F}_d + \sigma_w^2 \mathbf{I}_N) \Phi^H(\varepsilon). \end{aligned} \quad (33)$$

Moreover, it can be easily verified that $\mathbf{h} = (\mathbf{I}_L \otimes \mathbf{1}_N) \mathbf{h}_s$ with $\mathbf{h}_s = [h_0, \dots, h_{L-1}]^T$ being the channel impulse response (CIR) in time-invariant case. It follows that

$$\mathbf{R}_h = (\mathbf{I}_L \otimes \mathbf{1}_N) \mathbf{R}_L (\mathbf{I}_L \otimes \mathbf{1}_N)^H. \quad (34)$$

Putting (34) into (33), and together with the fact that

$$\begin{aligned} \mathbf{D}[\mathbf{F}_p^H \mathbf{x}_p] (\mathbf{I}_L \otimes \mathbf{1}_N) &= [\mathbf{\Xi}_0 \mathbf{F}_p^H \mathbf{x}_p, \dots, \mathbf{\Xi}_{L-1} \mathbf{F}_p^H \mathbf{x}_p] \\ &= \mathbf{F}_p^H \text{diag}\{\mathbf{x}_p\} \mathbf{F}_L \end{aligned} \quad (35)$$

where \mathbf{F}_L contains the first L columns of the FFT matrix, we have the covariance matrix

$$\begin{aligned} \mathbb{E}\{\mathbf{y}\mathbf{y}^H|\varepsilon\} &= \Phi(\varepsilon) (\mathbf{F}_p^H \text{diag}\{\mathbf{x}_p\} \mathbf{F}_L \mathbf{R}_L (\mathbf{F}_p^H \text{diag}\{\mathbf{x}_p\} \mathbf{F}_L)^H \\ &+ \mathbf{F}_d^H \mathbf{\Lambda}_d \mathbf{F}_d + \sigma_w^2 \mathbf{I}_N) \Phi^H(\varepsilon) \end{aligned} \quad (36)$$

and the CFO estimator in (29) becomes

$$\begin{aligned} \hat{\varepsilon} &= \arg \min_{\varepsilon} \mathbf{y}^H \Phi(\varepsilon) (\mathbf{F}_p^H \text{diag}\{\mathbf{x}_p\} \mathbf{F}_L \mathbf{R}_L (\mathbf{F}_p^H \text{diag}\{\mathbf{x}_p\} \mathbf{F}_L)^H \\ &+ \mathbf{F}_d^H \mathbf{\Lambda}_d \mathbf{F}_d + \sigma_w^2 \mathbf{I}_N)^{-1} \Phi^H(\varepsilon) \mathbf{y}. \end{aligned} \quad (37)$$

This estimator is equivalent to the MLE1 in [19] which has been derived for OFDM systems over time-invariant frequency-selective channels.

V. SIMULATION RESULTS AND DISCUSSIONS

In this section, the performance of the proposed algorithm for data detection is demonstrated by Monte Carlo simulations, where each point is obtained by averaging the results over $M = 10^4$ runs. Each OFDM symbol has 128 subcarriers ($N=128$) and the length of CP is 8. Carrier frequency is $f_c = 2$ GHz and the sample interval is $T_s = 2\mu s$. Normalized maximal Doppler shift is set as $Nf_d T_s = 0.05$ (corresponding to the speed of the mobile terminal at 105.5 km/hr) and $Nf_d T_s = 0.15$ (corresponding to the speed at 316.4 km/hr) for illustration. The channel has five taps ($L = 5$) with an exponential power delay profile (PDP) $\sigma_l^2 = \exp(-\beta l) \frac{1 - \exp(-\beta)}{1 - \exp(-\beta L)}$, $l = 0, \dots, L-1$ and $\beta = 0.2$. Each tap is independent, assumed to experience the same f_d and generated according to the Jakes' model. The CFO is uniformly distributed within the range $[-0.5, 0.5]$.

The pilot structure in [20] is adopted. More specifically, fourteen pilot clusters are used and each cluster occupies three pilots, with only one non-zero pilot in the middle of the cluster. This means that roughly 32.81% of the subcarriers are occupied by pilots. The non-zero pilots are generated as standard complex Gaussian random variables and the data are modulated by 16 QAM.

Over DSCs, the impacts of CFO and channel are similar, i.e., both the CFO and the channel induce ICI. From data detection point of view, according to (24), it can be seen that, only the estimate of the combined CFO and channel matrix $\Phi(\hat{\varepsilon}^i) \mathbf{H}[\hat{\mathbf{h}}^i]$ is required. Ignoring those zero entries in $\Phi(\hat{\varepsilon}^i) \mathbf{H}[\hat{\mathbf{h}}^i]$, the combined CFO and channel matrix is completely characterized by the combined CFO and channel vector $(\mathbf{I}_L \otimes \Phi(\hat{\varepsilon}^i)) \hat{\mathbf{h}}^i$. To evaluate the performance of the proposed algorithm, the MSE of the combined CFO and channel estimation is compared with the Bayesian Cramer Rao Bound (BCRB) given in Appendix C. Notice that in the derived BCRB, we consider all subcarriers are used for pilot transmission, since the performance of the proposed algorithm with scattered pilots is expected to approach the all pilot case. The simulated normalized MSE (NMSE) of combined CFO and channel estimation at the i^{th} iteration is defined as

$$NMSE = \frac{\sum_{n=0}^{M-1} \|(\mathbf{I}_L \otimes \Phi(\hat{\varepsilon}^i)) \hat{\mathbf{h}}^i - (\mathbf{I}_L \otimes \Phi(\varepsilon)) \mathbf{h}\|^2}{\sum_{n=0}^{M-1} \|(\mathbf{I}_L \otimes \Phi(\varepsilon)) \mathbf{h}\|^2}. \quad (38)$$

A. Convergence of the proposed algorithm

Fig. 1 shows the NMSE convergence performance of the combined CFO and channel estimation with signal-to-noise ratios (SNRs) equal to 20dB and 30dB. It can be seen that the NMSEs of the combined CFO and channel estimation improve significantly in the first iteration and they converge to stable values quickly after about three iterations. Furthermore, after convergence, the performance of the combined CFO and channel estimation only shows slight degradation when the normalized maximum Doppler shift increases from 0.05 to 0.15, i.e., the velocity increases from 105.5km/hr to 316.4km/hr.

Fig. 2 shows the symbol error rates (SERs) of data detection versus the number of iterations for SNR=20dB and 30dB.

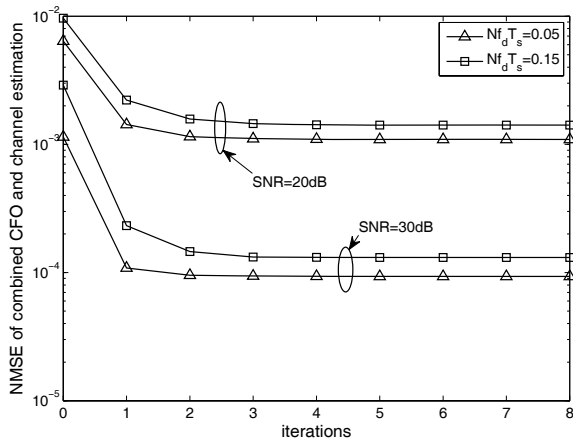


Fig. 1. Convergence of the combined CFO and channel estimates.

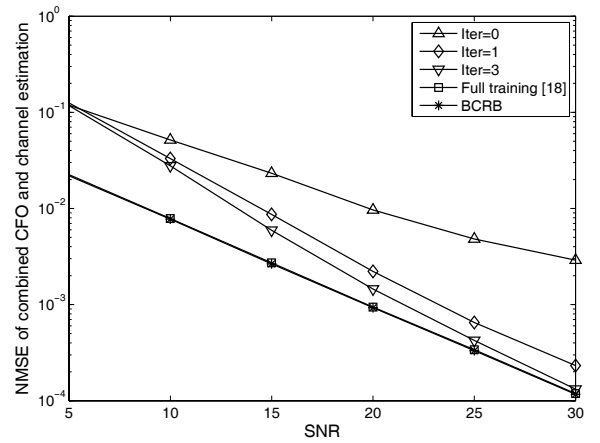
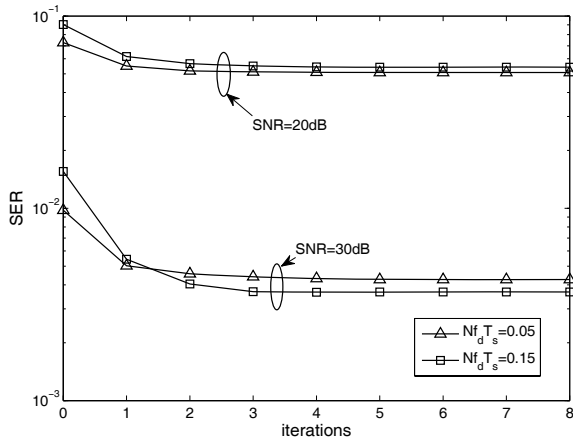
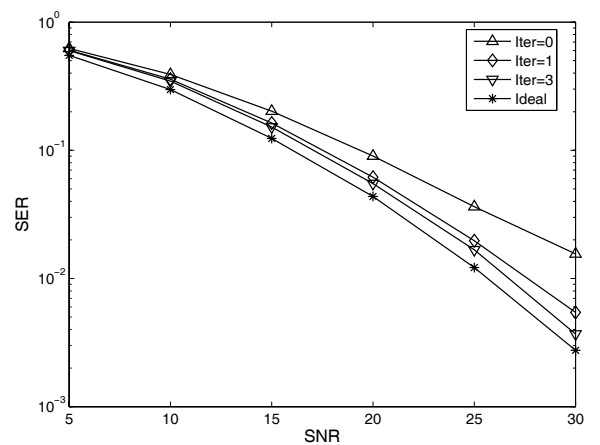
Fig. 3. NMSE of combined CFO and channel estimation for $Nf_d T_s = 0.15$.

Fig. 2. Convergence of data detection.

Fig. 4. Performance of data detection for $Nf_d T_s = 0.15$.

Similarly, the SERs converge quickly after three iterations for both cases. After convergence, the performance of the proposed data estimator over different maximum Doppler shifts is almost the same at the same SNR level, which demonstrates the robustness of the proposed data detection against Doppler spreads. Notice that, for SNR=30dB, the proposed data detection performs slightly better at $Nf_d T_s = 0.15$ than $Nf_d T_s = 0.05$. This slight improvement comes from Doppler diversity [16], which increases as the velocity becomes larger.

B. Performance of the proposed algorithm

In Fig. 3, the NMSE of the combined CFO and channel estimates versus different SNRs are depicted for the case $Nf_d T_s = 0.15$. The performance of the proposed initial CFO and channel estimators is marked as 'Iter=0'. The BCRB derived with full training and the performance of the CFO and channel estimation with full training [18] are also shown for comparison. It can be seen that the performance of combined CFO and channel estimation improves significantly in the first iteration. Furthermore, after three iterations, the NMSE of the proposed algorithm almost touches the derived BCRB (after normalization) at high SNRs. Fig. 4 shows the BER performance of the proposed algorithm as a function of SNR

with $Nf_d T_s = 0.15$. The proposed initial data estimator is labeled as 'Iter=0'. The ideal case which assumes perfect CFO and channel state information is also depicted as a reference. As can be seen, though the density of pilots is low (roughly 32.81% pilots inserted), after three iterations, the SER performance is very close to the ideal case. Similar results are also observed for the case $Nf_d T_s = 0.05$ and figures are not presented here due to space limitation.

Besides that, comparison has also been performed between our proposed algorithm and those derived for channel estimation only given in [20] and [21], where basis expansion models (BEMs) are adopted to approximate the DSCs. Specifically, the proposed algorithm is compared with the LS estimator using GCE-BEM, the LMMSE estimator using KL-BEM in [21], and the LMMSE estimator using CE-BEM in [20]. The number of BEM coefficients used for each channel tap are three, five, three for GCE-BEM, KL-BEM and CE-BEM respectively. For a fair comparison, CFO is not considered here, i.e., $\varepsilon = 0$. Results are shown in Fig. 5. It can be seen that, after convergence, the performance of the proposed algorithm is much better than those in [20] and [21]. This is because in the proposed algorithm, data is tentatively recovered to aid the channel estimation via EM algorithm, while it is treated

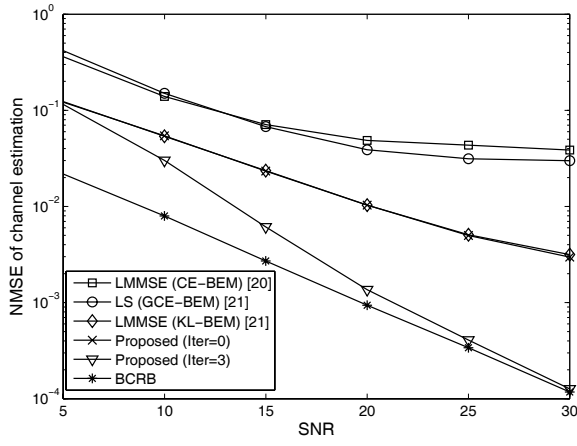


Fig. 5. Performance comparison of the proposed algorithm with [20] and [21] ($Nf_d T_s = 0.15$).

as interference in [20] and [21].

C. Effect of channel statistics mismatch

In the proposed algorithm, the required channel statistics depends on the maximum Doppler shift f_d and PDP. Generally, f_d and PDP are obtained by estimation or measurement, and mismatch between estimated and true channel statistics may exist. The mismatch effect on the performance of the proposed algorithm is now illustrated. Take the maximum Doppler shift $f_d = 585.94$ Hz as an example, which corresponds to the case $Nf_d T_s = 0.15$. The normalized mismatch of the maximum Doppler shift e_{f_d} , which is defined as $e_{f_d} = \frac{\hat{f}_d - f_d}{f_d}$, varies from -0.5 to 0.5 to test its effect on the proposed algorithm. Simulation results given in Fig. 6 show that the mismatch of maximum Doppler shift ($e_{f_d} \neq 0$) results in only slight performance degradation on the proposed algorithm. With respect to the impact of PDP mismatch, following the idea of Whitworth [22], PDP mismatch is characterized by the mismatch in the knowledge of β , since β is the only essential parameter in the PDP once the channel length L is fixed. Fig. 7 shows the performance of the proposed algorithm when the normalized mismatch of β , i.e., $e_\beta = \frac{\hat{\beta} - \beta}{\beta}$, varies within $[-0.5, 0.5]$. It can be seen that the performance degradation caused by the PDP mismatch is almost negligible, which is consistent with the conclusion given in [22]. These results demonstrate that the proposed method is robust against mismatch in the channel statistics.

VI. CONCLUSIONS

Semiblind iterative data detection for OFDM systems over unknown doubly selective channels and CFO was developed based on the expectation-maximization algorithm. For initialization, approximated ML CFO estimator and MMSE-based channel and data estimators were derived respectively. Simulation results demonstrated that the proposed EM algorithm is robust to Doppler spread. After convergence, the performance of data detection is close to the ideal case with perfect CFO and channel state information. Moreover, the NMSE of the combined CFO and channel estimates almost touch the derived BCRB for full training at high SNRs.

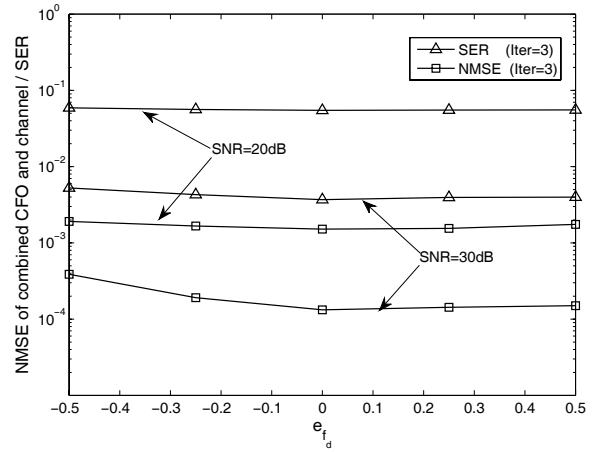


Fig. 6. The effect of Doppler shift mismatch ($Nf_d T_s = 0.15$).

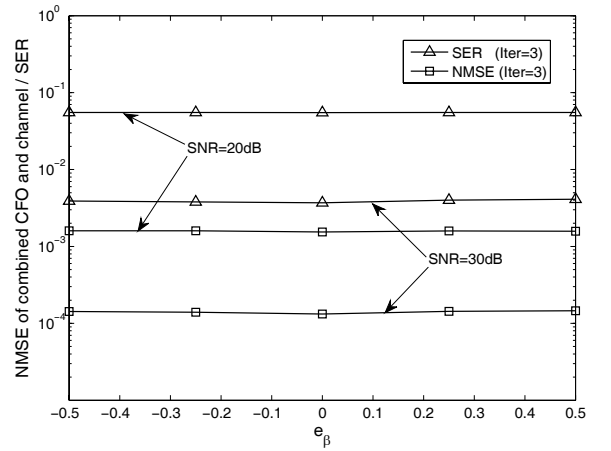


Fig. 7. The effect of power delay profile (PDP) mismatch ($Nf_d T_s = 0.15$).

APPENDIX A DERIVATION OF (15) AND (16)

Using Bayes' rule, the conditional pdf can be expressed as

$$p(\mathbf{h}|\mathbf{y}, \hat{\varepsilon}^{i-1}, \hat{\mathbf{x}}_d^{i-1}) = \frac{p(\mathbf{y}|\varepsilon, \mathbf{h}, \mathbf{x}_d)p(\mathbf{h})}{p(\mathbf{y}|\varepsilon, \mathbf{x}_d)} \Big|_{\varepsilon=\hat{\varepsilon}^{i-1}, \mathbf{x}_d=\hat{\mathbf{x}}_d^{i-1}} \quad (39)$$

where $p(\mathbf{h})$ and $p(\mathbf{y}|\varepsilon, \mathbf{h}, \mathbf{x}_d)$ are specified in (9) and (10) respectively. Putting (9) and (10) into (39), using the fact that $\mathbf{D}[\mathbf{s}|\mathbf{h}] = \mathbf{H}[\mathbf{h}](\mathbf{F}_p^H \mathbf{x}_p + \mathbf{F}_d^H \mathbf{x}_d)$ and completing the square for \mathbf{h} , $p(\mathbf{h}|\mathbf{y}, \hat{\varepsilon}^{i-1}, \hat{\mathbf{x}}_d^{i-1})$ becomes

$$\begin{aligned} p(\mathbf{h}|\mathbf{y}, \hat{\varepsilon}^{i-1}, \hat{\mathbf{x}}_d^{i-1}) &= \frac{1}{p(\mathbf{y}|\hat{\varepsilon}^{i-1}, \hat{\mathbf{x}}_d^{i-1})} \\ &\times \frac{1}{(\pi)^{NL} |\mathbf{C}_h^i|} \exp\{-\mathbf{h} - \hat{\mathbf{h}}^i)^H (\mathbf{C}_h^i)^{-1} (\mathbf{h} - \hat{\mathbf{h}}^i)\} \\ &\times \frac{|\mathbf{C}_h^i|}{(\pi \sigma_w^2)^N |\mathbf{R}_h|} \exp\left\{-\frac{1}{\sigma_w^2} (\mathbf{y}^H \mathbf{y} - \sigma_w^2 (\hat{\mathbf{h}}^i)^H (\mathbf{C}_h^i)^{-1} \hat{\mathbf{h}}^i)\right\} \end{aligned} \quad (40)$$

where $\hat{\mathbf{h}}^i \triangleq (\mathbf{D}^H[\hat{\mathbf{s}}^{i-1}]\mathbf{D}[\hat{\mathbf{s}}^{i-1}] + \sigma_w^2 \mathbf{R}_h^{-1})^{-1} \mathbf{D}^H[\hat{\mathbf{s}}^{i-1}] \Phi^H(\hat{\varepsilon}^{i-1})\mathbf{y}$ and $\mathbf{C}_h^i \triangleq \sigma_w^2 (\mathbf{D}^H[\hat{\mathbf{s}}^{i-1}]\mathbf{D}[\hat{\mathbf{s}}^{i-1}] + \sigma_w^2 \mathbf{R}_h^{-1})^{-1}$ with $\hat{\mathbf{s}}^{i-1} = \mathbf{F}_d^H \hat{\mathbf{x}}_d^{i-1} + \mathbf{F}_p^H \mathbf{x}_p$.

In $\hat{\mathbf{h}}^i$ and \mathbf{C}_h^i , the matrix inverse of \mathbf{R}_h is required, but it does not exist. To avoid this, using the matrix inversion

lemma,

$$\hat{\mathbf{h}}^i = \mathbf{R}_h \mathbf{D}^H [\hat{\mathbf{s}}^{i-1}] (\sigma_w^2 \mathbf{I}_N + \mathbf{D} [\hat{\mathbf{s}}^{i-1}] \mathbf{R}_h \mathbf{D}^H [\hat{\mathbf{s}}^{i-1}])^{-1} \times \Phi^H(\hat{\varepsilon}^{i-1}) \mathbf{y}, \quad (41)$$

$$\mathbf{C}_h^i = \mathbf{R}_h - \mathbf{R}_h \mathbf{D}^H [\hat{\mathbf{s}}^{i-1}] \times (\sigma_w^2 \mathbf{I}_N + \mathbf{D} [\hat{\mathbf{s}}^{i-1}] \mathbf{R}_h \mathbf{D}^H [\hat{\mathbf{s}}^{i-1}])^{-1} \mathbf{D} [\hat{\mathbf{s}}^{i-1}] \mathbf{R}_h. \quad (42)$$

Obviously from (40), the pdf $p(\mathbf{y}|\hat{\varepsilon}^{i-1}, \hat{\mathbf{x}}_d^{i-1})$ needs to be computed. Based on (9) and (10), and after completing square for \mathbf{h} ,

$$p(\mathbf{y}|\hat{\varepsilon}^{i-1}, \hat{\mathbf{x}}_d^{i-1}) = \int p(\mathbf{y}|\hat{\varepsilon}^{i-1}, \hat{\mathbf{x}}_d^{i-1}, \mathbf{h}) p(\mathbf{h}) d\mathbf{h} \\ = \frac{|\mathbf{C}_h^i|}{(\pi \sigma_w^2)^N |\mathbf{R}_h|} \exp\left\{-\frac{1}{\sigma_w^2} (\mathbf{y}^H \mathbf{y} - \sigma_w^2 (\hat{\mathbf{h}}^i)^H (\mathbf{C}_h^i)^{-1} \hat{\mathbf{h}}^i)\right\}. \quad (43)$$

Putting (43) into (40),

$$p(\mathbf{h}|\mathbf{y}, \hat{\varepsilon}^{i-1}, \hat{\mathbf{x}}_d^{i-1}) = \frac{1}{(\pi)^{NL} |\mathbf{C}_h^i|} \times \exp\left\{-(\mathbf{h} - \hat{\mathbf{h}}^i)^H (\mathbf{C}_h^i)^{-1} (\mathbf{h} - \hat{\mathbf{h}}^i)\right\}. \quad (44)$$

The pdf $p(\mathbf{h}|\mathbf{y}, \hat{\varepsilon}^{i-1}, \hat{\mathbf{x}}_d^{i-1})$ is a Gaussian distribution. Therefore, $\hat{\mathbf{h}}^i$ and \mathbf{C}_h^i given in (41) and (42) respectively are in fact its conditional mean and covariance, and thus (15) and (16) are proved.

APPENDIX B DERIVATION OF (28)

Notice that $\mathbb{E}\{\mathbf{y}|\varepsilon\} = \mathbf{0}$, the covariance matrix of \mathbf{y} given ε is equivalent to its correlation matrix and is given by

$$\mathbb{E}\{\mathbf{y}\mathbf{y}^H|\varepsilon\} = \mathbb{E}_{\mathbf{h}, \mathbf{x}_d, \mathbf{w}}\{(\Phi(\varepsilon)\mathbf{H}[\mathbf{h}]\mathbf{s} + \mathbf{w})(\Phi(\varepsilon)\mathbf{H}[\mathbf{h}]\mathbf{s} + \mathbf{w})^H\} \\ = \mathbb{E}_{\mathbf{h}}\{\Phi(\varepsilon)\mathbf{H}[\mathbf{h}](\mathbf{F}_p^H \mathbf{x}_p \mathbf{x}_p^H \mathbf{F}_p + \mathbf{F}_d^H \Lambda_d \mathbf{F}_d) \\ \times \mathbf{H}^H[\mathbf{h}]\Phi^H(\varepsilon) + \sigma_w^2 \mathbf{I}_N|\varepsilon\} \quad (45)$$

with the second equality due to expectation with respect to \mathbf{x}_d and \mathbf{w} . Notice that Λ_d is diagonal and the FFT matrix is orthogonal, we then have $\mathbf{F}_d^H \Lambda_d \mathbf{F}_d = \sum_{m \in I_d} \sigma_{x_m}^2 \mathbf{f}_m \mathbf{f}_m^H$ with $\sigma_{x_m}^2$ being the m^{th} diagonal element of Λ_d and \mathbf{f}_m being the corresponding row of \mathbf{F}_d . Therefore, with *Property 1*, $\mathbf{H}[\mathbf{h}]\mathbf{F}_d^H \Lambda_d \mathbf{F}_d \mathbf{H}^H[\mathbf{h}] = \sum_{m \in I_d} \sigma_{x_m}^2 \mathbf{D}[\mathbf{f}_m] \mathbf{h} \mathbf{h}^H \mathbf{D}^H[\mathbf{f}_m]$. Putting this result into (45) and taking expectation with respect to \mathbf{h} , $\mathbb{E}\{\mathbf{y}\mathbf{y}^H|\varepsilon\}$ becomes

$$\mathbb{E}\{\mathbf{y}\mathbf{y}^H|\varepsilon\} = \Phi(\varepsilon) (\mathbf{D}[\mathbf{F}_p^H \mathbf{x}_p] \mathbf{R}_h \mathbf{D}^H [\mathbf{F}_p^H \mathbf{x}_p] \\ + \sum_{m \in I_d} \sigma_{x_m}^2 \mathbf{D}[\mathbf{f}_m] \mathbf{R}_h \mathbf{D}^H [\mathbf{f}_m] + \sigma_w^2 \mathbf{I}_N) \Phi^H(\varepsilon), \quad (46)$$

where we also used the fact that $\mathbf{H}[\mathbf{h}]\mathbf{F}_p^H \mathbf{x}_p = \mathbf{D}[\mathbf{F}_p^H \mathbf{x}_p] \mathbf{h}$.

According to the definition of $\mathbf{D}[\mathbf{f}_m]$ and \mathbf{R}_h , we have $\mathbf{D}[\mathbf{f}_m] \mathbf{R}_h \mathbf{D}^H [\mathbf{f}_m] = \sum_{l=0}^{L-1} \sigma_l^2 \text{diag}\{\Xi_l \mathbf{f}_m\} \mathbf{J} \text{diag}^H\{\Xi_l \mathbf{f}_m\}$. Notice that $[\Xi_l \mathbf{f}_m]_n = e^{j2\pi m(n-l)/N}$, it then follows that $[\text{diag}\{\Xi_l \mathbf{f}_m\} \mathbf{J} \text{diag}^H\{\Xi_l \mathbf{f}_m\}]_{n,k} = [\mathbf{J}]_{n,k} e^{j2\pi m(n-k)/N}$ which does not depend on l . Therefore,

$$\left[\sum_{m \in I_d} \sigma_{x_m}^2 \mathbf{D}[\mathbf{f}_m] \mathbf{R}_h \mathbf{D}^H [\mathbf{f}_m] \right]_{n,k} = [\mathbf{J}]_{n,k} \cdot [\mathbf{F}_d^H \Lambda_d \mathbf{F}_d]_{n,k} \quad (47)$$

where the fact that $\sum_{l=0}^{L-1} \sigma_l^2 = 1$ has been used. It follows that

$$\sum_{m \in I_d} \sigma_{x_m}^2 \mathbf{D}[\mathbf{f}_m] \mathbf{R}_h \mathbf{D}^H [\mathbf{f}_m] = \mathbf{J} \odot (\mathbf{F}_d^H \Lambda_d \mathbf{F}_d). \quad (48)$$

Putting (48) into (46), we finally have (28).

APPENDIX C BCRB

First we rewrite the system model as

$$\mathbf{y} = \mathbf{D}[\mathbf{s}] (\mathbf{I}_L \otimes \Phi(\varepsilon)) \mathbf{h} + \mathbf{w}. \quad (49)$$

The pdf of the combined channel vector $\bar{\mathbf{h}} \triangleq (\mathbf{I}_L \otimes \Phi(\varepsilon)) \mathbf{h}$ can be easily shown to be

$$p(\bar{\mathbf{h}}) = \frac{1}{(\pi)^{NL} |\mathbf{R}_h|} \times \exp(-\bar{\mathbf{h}}^H (\mathbf{I}_L \otimes \Phi(\varepsilon)) \mathbf{R}_h^{-1} (\mathbf{I}_L \otimes \Phi(\varepsilon))^H \bar{\mathbf{h}}). \quad (50)$$

With \mathbf{s} being known, the likelihood function of \mathbf{y} given $\bar{\mathbf{h}}$ is

$$p(\mathbf{y}|\bar{\mathbf{h}}) = \frac{1}{(\pi \sigma_w^2)^N} \exp(-\frac{1}{\sigma_w^2} \|\mathbf{y} - \mathbf{D}[\mathbf{s}] \bar{\mathbf{h}}\|^2). \quad (51)$$

Defining $\boldsymbol{\theta} = [\Re\{\bar{\mathbf{h}}\}^T \Im\{\bar{\mathbf{h}}\}^T]^T$, the BCRB for the combined CFO and channel estimate can be expressed as [23]

$$BCRB = \text{Tr}\{(\mathbf{J}_F + \mathbf{J}_P)^{-1}\} \quad (52)$$

where \mathbf{J}_F is given by [24]

$$\mathbf{J}_F = \mathbb{E}_{\bar{\mathbf{h}}} \left\{ \mathbb{E}_{\mathbf{y}|\bar{\mathbf{h}}} \left\{ \frac{\partial^2 \log p(\mathbf{y}|\bar{\mathbf{h}})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} \right\} \right\} \\ = \frac{2}{\sigma_w^2} \begin{bmatrix} \Re\{\mathbf{D}^H(\mathbf{s}) \mathbf{D}[\mathbf{s}]\} & -\Im\{\mathbf{D}^H(\mathbf{s}) \mathbf{D}[\mathbf{s}]\} \\ \Im\{\mathbf{D}^H(\mathbf{s}) \mathbf{D}[\mathbf{s}]\} & \Re\{\mathbf{D}^H(\mathbf{s}) \mathbf{D}[\mathbf{s}]\} \end{bmatrix} \quad (53)$$

and \mathbf{J}_P is the prior information matrix defined as [23]

$$\mathbf{J}_P = -\mathbb{E}_{\bar{\mathbf{h}}} \left\{ \frac{\partial^2 \log p(\bar{\mathbf{h}})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} \right\} \\ = 2 \begin{bmatrix} (\mathbf{I}_L \otimes \Phi(\varepsilon)) \mathbf{R}_h^{-1} (\mathbf{I}_L \otimes \Phi(\varepsilon))^H & \mathbf{0} \\ \mathbf{0} & (\mathbf{I}_L \otimes \Phi(\varepsilon)) \mathbf{R}_h^{-1} (\mathbf{I}_L \otimes \Phi(\varepsilon))^H \end{bmatrix} \quad (54)$$

with the second equality due to (50).

After inverting $\mathbf{J}_F + \mathbf{J}_P$, combining together the real part and imaginary part of $\bar{\mathbf{h}}$ and eliminating ε , we can obtain the bound as

$$BCRB = \sigma_w^2 \text{Tr}\{(\mathbf{D}^H[\mathbf{s}] \mathbf{D}[\mathbf{s}] + \sigma_w^2 \mathbf{R}_h^{-1})^{-1}\}. \quad (55)$$

REFERENCES

- [1] Y. Sun, Z. Xiong, and X. Wang, "EM-based iterative receiver design with carrier-frequency offset estimation for MIMO OFDM systems," *IEEE Trans. Commun.*, vol. 53, no. 4, pp. 581-586, Apr. 2005.
- [2] M. Guenach, H. Wymeersch, H. Steendam, and M. Moeneclaey, "Code-aided ML joint synchronization and channel estimation for downlink MC-CDMA," *IEEE J. Sel. Areas Commun.*, vol. 24, no. 6, pp. 1105-1114, June 2006.
- [3] M. Pun, M. Morelli, and C. Kuo, "Iterative detection and frequency synchronization for OFDMA uplink transmissions," *IEEE Trans. Commun.*, vol. 6, no. 2, pp. 629-639, Feb. 2007.
- [4] J. Gozálviz, "Moving beyond 3G," *IEEE Veh. Technol. Mag.*, vol. 1, no. 2, pp. 48-55, June 2006.
- [5] C. Yeun, "Mobile TV technologies," in *Proc. IEEE International Conf. Inf. Commun. Technol. (ICIT)*, pp. 2-9, Cairo, Egypt, Dec. 2007.
- [6] J. Zhou, Z. Ou, M. Rautiainen, T. Koskela, and M. Ylianttila, "Digital television for mobile devices," *IEEE Multimedia*, vol. 16, no. 1, pp. 60-71, Jan.-Mar. 2009.

- [7] I. Barhumi, G. Leus, and M. Moonen, "Equalization for OFDM over doubly selective channels," *IEEE Trans. Signal Process.*, vol. 54, no. 4, pp. 1445-1457, Apr. 2006.
- [8] Z. Tang and G. Leus, "A novel receiver architecture for single-carrier transmission over time-varying channels," *IEEE J. Sel. Areas Commun.*, vol. 26, no. 2, pp. 366-377, Feb. 2008.
- [9] O. Rousseaoux, G. Leus, and M. Moonen, "Estimation and equalization of doubly selective channels using known symbol sadding," *IEEE Trans. Signal Process.*, vol. 54, no. 3, pp. 979-990, Mar. 2006.
- [10] S. He and J. K. Tugnait, "On doubly selective channel estimation using superimposed training and discrete prolate spheroidal sequences," *IEEE Trans. Signal Process.*, vol. 56, no. 7, pp. 3214-3228, July 2008.
- [11] Y.-S. Choi, P. J. Voltz, and F. A. Cassara, "On channel estimation and detection for multicarrier signals in fast and selective Rayleigh fading channels," *IEEE Trans. Commun.*, vol. 49, no.8, pp. 1375-1387, Aug. 2001.
- [12] X. Cai and G. B. Giannakis, "Bounding performance and suppressing intercarrier interference in wireless mobile OFDM," *IEEE Trans. Commun.*, vol. 51, no. 12, pp. 2047-1056, Dec. 2003.
- [13] J. Choi, *Adaptive and Iterative Signal Processing in Communications*. Cambridge, UK: Cambridge University Press, 2006.
- [14] A. F. Molisch, F. Tufvesson, J. Karedal, and C. Mecklenbrauker, "Propagation aspects of vehicle-to-vehicle communications—an overview," in *Proc. IEEE Radio Wireless Symp. (RWS)*, pp. 179-182, Jan. 2009.
- [15] I. Sen and D. W. Matolak, "Vehicle-vehicle channel models for the 5-GHz band," *IEEE Trans. Intelligent Transportation Syst.*, vol. 9, no. 2, pp. 235-245, June 2008.
- [16] X. Ma and G. B. Giannakis, "Maximum-diversity transmissions over doubly selective wireless channels," *IEEE Trans. Inf. Theory.*, vol. 49, no. 7, pp. 1832-1840, July 2003.
- [17] S. Shlien, "A method for computing the partial singular value decomposition," *IEEE Trans. Pattern Analysis Machine Intelligence*, vol. 4, pp. 671-676, Nov. 1982.
- [18] N. Ricklin and J. Zeidler, "Data-aided joint estimation of carrier frequency offset and frequency-selective time-varying channel," in *Proc. IEEE International Conf. Commun. (ICC)*, pp. 5098-5102, May 2008.
- [19] T. Cui and C. Tellambura, "Maximum-likelihood carrier frequency offset estimation for OFDM systems over frequency-selective fading channels," in *Proc. IEEE International Conf. Commun. (ICC)*, vol. 4, pp. 2506-2510, May 2005.
- [20] A. Kannu and P. Schniter, "Design and analysis of MMSE pilot-aided cyclic-prefixed block transmissions for doubly selective channels," *IEEE Trans. Signal Process.*, vol. 56, no. 3, pp. 1148-1160, Mar. 2008.
- [21] Z. Tang, G. Leus, R. C. Cannizzaro, and P. Banelli, "Pilot-assisted time-varying channel estimation for OFDM systems," *IEEE Trans. Signal Process.*, vol. 55, no. 5, pp. 2226-2238, May 2007.
- [22] T. Whitworth, M. Ghogho, and D. McLernon, "Optimized training and basis expansion model parameters for doubly-selective channel estimation," *IEEE Trans. Wireless Commun.*, vol.8, no.3, pp. 1490-1498, Mar. 2009.
- [23] H. Tree and K. Bell, *Bayesian Bounds for Parameter Estimation and Nonlinear Filtering/Tracking*. Wiley-IEEE Press, 2007.
- [24] P. Stoica and O. Besson, "Training sequence design for frequency offset and frequency-selective channel estimation," *IEEE Trans. Commun.*, vol. 51, no. 11, pp. 1910-1917, Nov. 2003.

Lanlan He received the B.Sc.(Eng.) degree in 2004 and the M.Eng. degree in 2006 from Huazhong University of Science and Technology (HUST), Wuhan, China, both in control and system engineering. She is currently pursuing the Ph.D degree at the Department of Electrical and Electronic Engineering, the University of Hong Kong. Her current research interests include synchronization, channel estimation and data detection for OFDM systems over doubly selective channels.

Shaodan Ma received the B.Sc.(Eng.) and the M.Eng. degrees from NanKai University, Tianjin, China, in 1999 and 2002, respectively, all in electrical engineering, and the Ph.D. degree in electrical and electronic engineering from the University of Hong Kong, Hong Kong, in 2006. Since 2006, she has been with the Department of Electrical and Electronic Engineering, the University of Hong Kong, as a post-doctoral fellow. From March 2010, she is with Princeton University as a visiting researcher. Her research interests include MIMO systems, OFDM technique, cooperative networks, cognitive radio systems and related digital signal processing.

Yik-Chung Wu obtained the B.Eng. (EEE) degree in 1998 and the M.Phil. degree in 2001 from The University of Hong Kong (HKU). After graduation from Master degree, he was a research assistant in the same university. He started his Ph.D. degree in 2002 at the Texas A&M University, USA and obtained the Ph.D. degree in 2005. During his study at Texas A&M University, he was fully supported by the prestigious Croucher Foundation scholarship. From Aug. 2005 to Aug. 2006, he was with the Thomson Corporate Research, Princeton, NJ, as a Member of Technical Staff. Since Sep. 2006, he has been with the University of Hong Kong as an Assistant Professor. He was a TPC member for IEEE VTC Fall 2005, Globecom 2006, 2008, ICC 2007 and 2008. He is currently serving as an associate editor for the IEEE COMMUNICATIONS LETTERS. Yik-Chung's research interests are in general area of signal processing and communication systems, and in particular receiver algorithm design, synchronization techniques, channel estimation and equalization.

Tung-Sang Ng received the B.Sc.(Eng.) degree from The University of Hong Kong in 1972, the M.Eng.Sc. and Ph.D. degrees from the University of Newcastle, Australia, in 1974 and 1977, respectively, all in electrical engineering.

He worked for BHP Steel International and The University of Wollongong, Australia after graduation for 14 years before returned to The University of Hong Kong in 1991, taking up the position of Professor and Chair of Electronic Engineering. He was Head of Department of Electrical and Electronic Engineering from 2000 to 2003 and Dean of Engineering from 2003 to 2007. His current research interests include wireless communication systems, spread spectrum techniques, CDMA and digital signal processing. He has published over 300 international journal and conference papers.

He was the General Chair of ISCAS'97 and the VP-Region 10 of IEEE CAS Society in 1999 & 2000. He was an Executive Committee Member and a Board Member of the IEE Informatics Divisional Board (1999-2001) and was an ordinary member of IEE Council (1999-2001). He was awarded the Honorary Doctor of Engineering Degree by the University of Newcastle, Australia in 1997, the Senior Croucher Foundation Fellowship in 1999, the IEEE Third Millennium medal in 2000 and the Outstanding Researcher Award by The University of Hong Kong in 2003. He is a Fellow of IEEE, IET and HKIE.