Blind Carrier Frequency Offset Estimation for Interleaved OFDMA Uplink

Weile Zhang, Feifei Gao, Qinye Yin, and Arumugam Nallanathan, *Senior Member, IEEE*

Abstract—In this paper, we develop two novel blind carrier frequency offset (CFO) estimators for interleaved orthogonal frequency division multiple access (OFDMA) uplink transmission in a multi-antenna system. The first estimator is the subspace-based one and could blindly estimate multiple CFOs from a rank reduction approach. The second estimator is based on the maximum likelihood (ML) approach and improves the performance as compared to the first one. The higher computational complexity of the ML estimator is alleviated by the alternating projection (AP) method. Both the proposed estimators support fully loaded data transmissions, i.e., all subcarriers being occupied, which provides higher bandwidth efficiency as compared to the existing schemes. The numerical results are then provided to corroborate the proposed studies.

Index Terms—Carrier frequency offsets (CFO), orthogonal frequency division multiple access (OFDMA), subspace, maximum likelihood (ML).

I. INTRODUCTION

AS a promising technique for next-generation multiuser broadband wireless communications, orthogonal frequency division multiple access (OFDMA) has recently received a considerable amount of interest [1], [2]. It is well known that OFDMA inherits from orthogonal frequency division multiplexing (OFDM) the weakness of being sensitive to carrier frequency offset (CFO) that is generated by the frequency mismatch between the transceiver oscillators or the Doppler effect.

W. Zhang and Q. Yin are with the Ministry of Education Key Lab for Intelligent Networks and Network Security, Xi'an Jiaotong University, Xi'an, Shaanxi, 710049, P. R. China (e-mail: wlzhang1984@gmail.com, qyyin@mail.xjtu.edu.cn). W. Zhang is also with the Department of Automation, Tsinghua University as a guest student.

F. Gao is with Tsinghua National Laboratory for Information Science and Technology, Beijing, China and is with National Mobile Communications Research Laboratory, Southeast University, Nanjing, China, and is also with the School of Engineering and Science, Jacobs University, Bremen, Germany (e-mail: feifeigao@ieee.org).

A. Nallanathan is with the Department of Electronic Engineering, King's College London, London WC2R 2LS, United Kingdom (Email: nallanathan@ieee.org).

For OFDMA downlink transmissions where a single CFO exists between each transceiver pair, CFO estimation algorithms for OFDM or multi-input multi-output (MIMO) OFDM are directly applicable, see for examples [3]–[5]. However for OFDMA uplink, multiple CFOs co-exist at the receiver, making the CFO estimation more challenging. The effect of imperfect CFO estimation in OFDMA uplink was investigated in [6]. Using the frequency domain embedded pilot symbols, Sun *et al.* [7] proposed an iterative CFO estimation approach for tile structure based OFDMA transmission [8]. When training sequences are transmitted from each user, CFOs can be estimated from the maximum likelihood (ML) approach but with very high complexity. The approach in [9], [10] reduced the complexity by applying the divide-and-update frequency estimator (DUFE). Another interesting alternative was reported in [11] where the important sampling method [12] was applied and the CFOs were estimated by maximizing the mean likelihood function.

In addition to data-aided estimation, blind methods were also developed to improve the bandwidth efficiency. The CFOs can be computed by looking for the position of null subcarriers [13] or based on the optimization of a kurtosis-type cost function [14]. A frequency estimation scheme for uplink interleaved OFDMA that exploits the periodic structure of the signals from each user has been presented in [15], where the subspace estimation theory was adopted and the designed scheme is similar to the multiple signal classification (MUSIC) technique [16]. Based on the observation of [15], several advancements have been proposed later. For instance, Lee *et al.* [17] suggested using the estimation of signal parameters via rotational invariance technique (ESPRIT) [18], which provides better performance as well as lower computational complexity. Fan *et al.* [19] replaced MUSIC by a two-stage process where an initial coarse search is conducted before the precise searching step. Another CFO estimation scheme for the interleaved OFDMA/space division multiple access (SDMA) uplink systems was developed in [20] to support spatially separated users with a multi-antenna receiver. Zhu *et al.* [21] described a CFO estimation method for single carrier interleaved FDMA systems, aiming to improve the transmission efficiency.

Note that [15] and [17], originally derived for single-antenna receiver, need null subcarriers or a longer cyclic prefix (CP) to build the noise space. The other variations [19]–[21] mentioned above also have the same requirement. Recently, Hsieh *et al.* [22] presented a blind ML CFO estimation approach for interleaved OFDMA, where the series expansion of the correlation matrix was expressed and the multiple CFOs were

retrieved from a root-finding method. Although at most half subchannels can be used for data transmissions, [22] achieves better performance as compared to the subspace based approach [17]. Moreover, by deploying uniform linear array (ULA) at the receiver, several schemes have been reported to support full carrier-load transmission [23], [24]. However, these schemes are only valid under the assumption that the ULA at the receiver is elevated and the transmitted signal of each user must be restricted to a single direction of arrival (DOA).

In this paper, we develop two novel blind CFO estimation methods for interleaved OFDMA uplink transmission by equipping multi-antenna at the receiver. Based on the subspace theory and from the rank reduction approach, the multiple CFOs can be individually derived in a blind way. To improve the performance, we further propose a ML estimator whose higher computational complexity can be alleviated from the alternating-projection (AP) method [25]. We show that when the number of antenna elements at the receiver is larger than the number of multipaths from each user, both the proposed estimators support full carrier-load transmission, which provides higher bandwidth efficiency as compared to the existing methods [15], [17], [19]–[22]. Another byproduct of our proposed estimators is that, the channel responses of all users can be obtained simultaneously with little ambiguity, which differs from all the existing methods.

The rest of this paper is organized as follows. Section II formulates the problem. The proposed subspace based estimator and ML estimator are developed in sections III and IV, respectively. Simulation results are given in Section V and conclusions are drawn in Section VI.

Notations: Superscripts $(\cdot)^*$, $(\cdot)^T$, $(\cdot)^H$, $[\cdot]^\dagger$ and $E[\cdot]$ represent conjugate, transpose, Hermitian, pseudo inverse and expectation, respectively; $\mathbf{j} = \sqrt{-1}$ is the imaginary unit; $\|\mathbf{X}\|$ denotes the Frobenius norm of $\mathbf{X}$, and diag$(\cdot)$ is a diagonal matrix with main diagonal $(\cdot)$; blkdiag$(\cdot)$ represents the block-diagonal matrix operator; $\mathbb{C}^{m \times n}$ defines the vector space of all $m \times n$ complex matrices; $\mathbf{I}_N$ is the $N \times N$ identity matrix; $\mathbf{0}$ represents an all-zero matrix with appropriate dimension; $\otimes$ stands for the Kronecker product; Tr$\{\cdot\}$ denotes the trace operation; Matlab matrix representations are adopted, for example, $\mathbf{X}(r_1 : r_2, c_1 : c_2)$ denotes the submatrix of $\mathbf{X}$ with the rows from r_1 to r_2 and the columns from c_1 to c_2.

II. PROBLEM FORMULATION

Consider a multiuser OFDMA system with K users, N subcarriers, and M antennas at the receiver, as shown in Fig. 1. All subcarriers are sequentially indexed with $\{i\}$, $i = 0, 1, \cdots, N-1$, and are equally divided into Q subchannels, each having $P = N/Q$ subcarriers. The qth subchannel consists of subcarriers with the index set $\{q, Q+q, \cdots, (P-1)Q+q\}$, $q = 0, 1, \cdots, Q-1$. Each subchannel will be exclusively assigned to one user and, thus, no subchannel can be shared by more than one user. Without loss of generality, we consider only the case of $K = Q$ to simplify the illustration, whereas the designed scheme can be straightforwardly extended to more general cases.
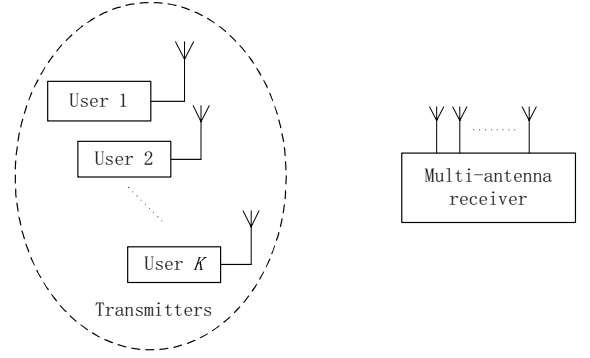

Fig. 1. System model with a multi-antenna receiver and K single-antenna users.

Assume the kth user occupies the $q^{(k)}$th subchannel and denote $\mathbf{s}_g^{(k)} = \left[s_{0,g}^{(k)}, s_{1,g}^{(k)}, \cdots, s_{P-1,g}^{(k)} \right]^T$ as the P information symbols of the kth user in the gth OFDMA block. The overall OFDMA block of length N can be represented by $\mathbf{x}_g^{(k)} = [x_{0,g}^{(k)}, x_{1,g}^{(k)}, \cdots, x_{N-1,g}^{(k)}]^T$ with

$$x_{i,g}^{(k)} = \begin{cases} s_{p,g}^{(k)}, & i = pK + q^{(k)}, \\ 0, & \text{otherwise.} \end{cases} \quad (1)$$

The transmitted time-domain signal by the kth user can then be expressed as

$$y_g^{(k)}(n) = \frac{1}{\sqrt{N}} \sum_{i=0}^{N-1} x_{i,g}^{(k)} \mathrm{e}^{\mathbf{j} \frac{2\pi i}{N} n}, \quad (2)$$

$n = 0, 1, \cdots, N-1$. The channel impulse response from the kth user to the mth antenna element can be modeled as:

$$h_m^{(k)}(\tau) = \sum_{l=1}^{L_h} h_{l,m}^{(k)} \delta(\tau - \tau_l^{(k)}), \quad (3)$$

where L_h is the number of multipaths, while $h_{l,m}^{(k)}$ and $\tau_l^{(k)}$ are the complex amplitude and delay for the lth multipath, respectively. We assume $h_{l,m}^{(k)}$'s are independent and identically distributed (i.i.d) complex Gaussian variables with zero mean and power $E[|h_{l,m}^{(k)}|^2] = 1/L_h$ such that the total power is normalized, i.e., $E\left[\sum_{l=1}^{L_h} |h_{l,m}^{(k)}|^2 \right] = 1$. Moreover, we assume that the delays are rounded to the nearest sampling position and $\tau_l^{(k)}$'s are of integer values. Then, in noise-free environment, the received signal at the mth receiving antenna element from the kth user, after CP removal, can be expressed as

$$\gamma_{m,g}^{(k)}(n) = \frac{1}{\sqrt{N}} \sum_{i=0}^{N-1} \left(\sum_{l=1}^{L_h} h_{l,m}^{(k)} \mathrm{e}^{-\mathbf{j} \frac{2\pi i}{N} \tau_l^{(k)}} \right) x_{i,g}^{(k)} \mathrm{e}^{\mathbf{j} \frac{2\pi i}{N} n}, \quad (4)$$

$n = 0, 1, \cdots, N-1$, where the expression in the bracket stands for the frequency-domain channel response at the ith subcarrier from the kth user to the mth antenna.

Denote the normalized CFO of the kth user as $\xi^{(k)} = \Delta f^{(k)} / \Delta f$, where Δf is the subcarrier spacing and $\Delta f^{(k)}$ is the real CFO of the kth user. We assume $\xi^{(k)} \in (-0.5, 0.5)$

in this paper. With the presence of CFOs, the signal model (4) changes to

$$\gamma_{m,g}^{(k)}(n) = \frac{\chi_g^{(k)}}{\sqrt{N}} \sum_{i=0}^{N-1} \left(\sum_{l=1}^{L_h} h_{l,m}^{(k)} e^{-j\frac{2\pi i}{N}\tau_l^{(k)}} \right) x_{i,g}^{(k)} e^{j\frac{2\pi(i+\xi^{(k)})}{N}n},$$
(5)

where $\chi_g^{(k)} = e^{j\frac{2\pi(g-1)(N+G)\xi^{(k)}}{N}}$ denotes the phase shift of the kth user accumulated from the previous $g-1$ blocks and G denotes the length of CP. Bearing in mind (1), we further rewrite (5) into

$$\gamma_{m,g}^{(k)}(n) = \sum_{p=0}^{P-1} e^{j\frac{2\pi}{N}(pK+q^{(k)}+\xi^{(k)})n} \sum_{l=1}^{L_h} h_{l,m}^{(k)} \tilde{s}_{l,p,g}^{(k)},$$
(6)

where

$$\tilde{s}_{l,p,g}^{(k)} = \frac{\chi_g^{(k)}}{\sqrt{N}} e^{-j\frac{2\pi(pK+q^{(k)})\tau_l^{(k)}}{N}} s_{p,g}^{(k)}.$$
(7)

The overall received signal from all K users at the mth antenna is then

$$\gamma_{m,g}(n) = \sum_{k=1}^{K} \gamma_{m,g}^{(k)}(n) = \sum_{k=1}^{K} \sum_{l=1}^{L_h} h_{l,m}^{(k)} z_{l,n,g}^{(k)},$$
(8)

where

$$z_{l,n,g}^{(k)} = \sum_{p=0}^{P-1} \tilde{s}_{l,p,g}^{(k)} e^{j\frac{2\pi}{N}(pK+q^{(k)}+\xi^{(k)})n}.$$
(9)

We define the channel response vector of the lth multipath component from the kth user as

$$\mathbf{h}_l^{(k)} = \left[h_{l,1}^{(k)}, h_{l,2}^{(k)}, \cdots, h_{l,M}^{(k)} \right]^T.$$
(10)

The corresponding $M \times L_h$ channel response matrix from the kth user is then obtained as

$$\mathbf{H}^{(k)} = \left[\mathbf{h}_1^{(k)}, \mathbf{h}_2^{(k)}, \cdots, \mathbf{h}_{L_h}^{(k)} \right].$$
(11)

Stacking the received signals from all M antennas, we obtain the following space-domain snapshot

$$\boldsymbol{\gamma}_{n,g} = \left[\gamma_{1,g}(n), \gamma_{2,g}(n), \cdots, \gamma_{M,g}(n) \right]^T = \mathbf{H}\mathbf{z}_{n,g},$$
(12)

where

$$\mathbf{H} = \left[\mathbf{H}^{(1)}, \mathbf{H}^{(2)}, \cdots, \mathbf{H}^{(K)} \right] \in \mathbb{C}^{M \times KL_h},$$

$$\mathbf{z}_{n,g} = \left[(\mathbf{z}_{n,g}^{(1)})^T, (\mathbf{z}_{n,g}^{(2)})^T, \cdots, (\mathbf{z}_{n,g}^{(K)})^T \right]^T \in \mathbb{C}^{KL_h \times 1},$$

$$\mathbf{z}_{n,g}^{(k)} = \left[z_{1,n,g}^{(k)}, z_{2,n,g}^{(k)}, \cdots, z_{L_h,n,g}^{(k)} \right]^T \in \mathbb{C}^{L_h \times 1}.$$

Observing from (9), for any $n = 0,1,\cdots,P-1$, $t = 0,1,\cdots,K-1$, we have

$$z_{l,n+tP,g}^{(k)} = \sum_{p=0}^{P-1} \tilde{s}_{l,p,g}^{(k)} e^{j\frac{2\pi}{N}(pK+q^{(k)}+\xi^{(k)})(n+tP)}$$

$$= e^{j\frac{2\pi t(q^{(k)}+\xi^{(k)})}{K}} z_{l,n,g}^{(k)}.$$
(13)

Hence, we can treat $\theta^{(k)} = \frac{q^{(k)}+\xi^{(k)}}{K}$ as the effective CFO of the kth user and obtain

$$z_{l,n+tP,g}^{(k)} = e^{j2\pi t\theta^{(k)}} z_{l,n,g}^{(k)}, \quad \mathbf{z}_{n+tP,g}^{(k)} = e^{j2\pi t\theta^{(k)}} \mathbf{z}_{n,g}^{(k)},$$
$$\mathbf{z}_{n+tP,g} = \mathbf{\Phi}^t \mathbf{z}_{n,g},$$
(14)

where

$$\mathbf{\Phi} = \text{blkdiag}\left(\mathbf{\Phi}^{(1)}, \mathbf{\Phi}^{(2)}, \cdots, \mathbf{\Phi}^{(K)} \right)$$
(15)

is a $KL_h \times KL_h$ diagonal matrix with $\mathbf{\Phi}^{(k)} = e^{j2\pi\theta^{(k)}}\mathbf{I}_{L_h}$. Combining (12) and (14) yields

$$\boldsymbol{\gamma}_{n+tP,g} = \mathbf{H}\mathbf{\Phi}^t\mathbf{z}_{n,g}.$$
(16)

With the presence of the noise, (16) can be rewritten as

$$\boldsymbol{\gamma}_{n+tP,g} = \mathbf{H}\mathbf{\Phi}^t\mathbf{z}_{n,g} + \mathbf{n}_{n+tP,g},$$
(17)

where $\mathbf{n}_{n+tP,g}$ is a length-M additive white Gaussian noise (AWGN) vector with covariance matrix $\sigma_n^2\mathbf{I}_M$ at the $(n+tP)$-th sample in the gth OFDMA block.

III. SUBSPACE ESTIMATOR

A. Properties of the Subspace

Stacking space-domain snapshot vectors from K equally spaced time samples (P samples apart) gives the following length-MK vector

$$\mathbf{a}_{p,g} = \begin{bmatrix} \boldsymbol{\gamma}_{p,g} \\ \boldsymbol{\gamma}_{p+P,g} \\ \vdots \\ \boldsymbol{\gamma}_{p+N-P,g} \end{bmatrix} = \begin{bmatrix} \mathbf{H} \\ \mathbf{H}\mathbf{\Phi} \\ \vdots \\ \mathbf{H}\mathbf{\Phi}^{K-1} \end{bmatrix} \mathbf{z}_{p,g} + \underbrace{\begin{bmatrix} \mathbf{n}_{p,g} \\ \mathbf{n}_{p+P,g} \\ \vdots \\ \mathbf{n}_{p+N-P,g} \end{bmatrix}}_{\bar{\mathbf{n}}_{p,g}},$$
(18)

where $\bar{\mathbf{n}}_{p,g}$ denotes the corresponding length-MK noise vector as defined above.

Defining

$$\mathbf{v}^{(k)} = \left[1, e^{j2\pi\theta^{(k)}}, \cdots, e^{j2\pi(K-1)\theta^{(k)}} \right]^T \in \mathbb{C}^{K \times 1},$$
$$\mathbf{V}^{(k)} = \mathbf{v}^{(k)} \otimes \mathbf{H}^{(k)} \in \mathbb{C}^{KM \times L_h},$$
$$\mathbf{V} = \left[\mathbf{V}^{(1)}, \mathbf{V}^{(2)}, \cdots, \mathbf{V}^{(K)} \right] \in \mathbb{C}^{KM \times KL_h},$$

we rewrite (18) as

$$\mathbf{a}_{p,g} = \mathbf{V}\mathbf{z}_{p,g} + \bar{\mathbf{n}}_{p,g},$$
(19)

whose covariance matrix can be computed as

$$\mathbf{R_a} = E\left[\mathbf{a}_{p,g}\mathbf{a}_{p,g}^H\right] = \mathbf{V}\mathbf{R}_{ZZ}\mathbf{V}^H + \sigma_n^2\mathbf{I}_{MK},$$
(20)

$p = 0,1,\cdots,P-1$, where $\mathbf{R}_{ZZ} = E[\mathbf{z}_{p,g}\mathbf{z}_{p,g}^H]$. We assume the fading channels are constant over successive L_s block durations, and thus, the covariance matrix can be approximated by $\hat{\mathbf{R}}_\mathbf{a} = \frac{1}{PL_s}\sum_{g=1}^{L_s}\sum_{p=0}^{P-1}\mathbf{a}_{p,g}\mathbf{a}_{p,g}^H$. When $\mathbf{V}$ is tall with full column rank and $\mathbf{R}_{ZZ}$ is nonsingular, the singular value decomposition (SVD) of $\mathbf{R_a}$ is

$$\mathbf{R_a} = [\mathbf{U_s}, \mathbf{U_n}]\boldsymbol{\Sigma}_\mathbf{a}[\mathbf{U_s}, \mathbf{U_n}]^H,$$
(21)

where $\mathbf{U_s} \in \mathbb{C}^{MK \times KL_h}$ and $\mathbf{U_n} \in \mathbb{C}^{MK \times K(M-L_h)}$ represent the signal and the noise space matrices, respectively.

Remark: From the observation in Appendix A, we need to assume $M > L_h$, i.e., the number of antenna elements at the receiver is larger than the number of multipaths from each user, in the following to guarantee the column full rank condition of $\mathbf{V}$. Meanwhile, we assume $L_h \leq P$ and $\tau_{L_h}^{(k)} < P$, and then the nonsingularity of $\mathbf{R}_{ZZ}$ will be guaranteed from the discussion in Appendix B.

We define the following parameterized Vandermonde vector $\mathbb{B}^{(k)}(\xi)$ with respect to ξ as

$$\mathbb{B}^{(k)}(\xi) = \left[1, e^{\mathbf{j}2\pi \frac{q^{(k)}+\xi}{K}}, \cdots, e^{\mathbf{j}2\pi \frac{(K-1)(q^{(k)}+\xi)}{K}}\right]^T \in \mathbb{C}^{K \times 1}, \tag{22}$$

where $\xi \in (-0.5, 0.5)$. Clearly, $\mathbb{B}^{(k)}(\xi)$ equals $\mathbf{v}^{(k)}$ when $\xi = \xi^{(k)}$, i.e., $\mathbb{B}^{(k)}(\xi^{(k)}) = \mathbf{v}^{(k)}$.

The following *Lemmas* are the key properties to design the subspace-based CFO estimator.

Lemma 1: For *any* non-zero length-M vector $\boldsymbol{x}$, there holds

$$(\mathbb{B}^{(k)}(\xi^{(k)}) \otimes \boldsymbol{x})^H \mathbf{U_n} \begin{cases} = \mathbf{0}, & \boldsymbol{x} \in \text{Span}(\mathbf{H}^{(k)}), \\ \neq \mathbf{0}, & \boldsymbol{x} \notin \text{Span}(\mathbf{H}^{(k)}). \end{cases} \tag{23}$$

Proof: See Appendix C. ∎

Lemma 2: Given *any* $M \times L_h$ matrix $\boldsymbol{X}$ with full column rank and $\xi \neq \xi^{(k)}$, there holds

$$(\mathbb{B}^{(k)}(\xi) \otimes \boldsymbol{X})^H \mathbf{U_n} \neq \mathbf{0}. \tag{24}$$

Proof: See Appendix D. ∎

B. Blind CFO Estimation

For any non-zero length-M vector $\boldsymbol{x}$, we have $\mathbb{B}^{(k)}(\xi) \otimes \boldsymbol{x} = (\mathbb{B}^{(k)}(\xi) \otimes \mathbf{I}_M)\boldsymbol{x}$ and

$$(\mathbb{B}^{(k)}(\xi) \otimes \boldsymbol{x})^H \mathbf{U_n} \mathbf{U_n^H} (\mathbb{B}^{(k)}(\xi) \otimes \boldsymbol{x}) = \boldsymbol{x}^H \mathbf{\Pi}^{(k)}(\xi)\boldsymbol{x}, \tag{25}$$

where $\mathbf{\Pi}^{(k)}(\xi) = (\mathbb{B}^{(k)}(\xi) \otimes \mathbf{I}_M)^H \mathbf{U_n} \mathbf{U_n^H} (\mathbb{B}^{(k)}(\xi) \otimes \mathbf{I}_M) \in \mathbb{C}^{M \times M}$. The following two important observations are made:

1) From *Lemma 1*, we know that the matrix $\mathbf{\Pi}^{(k)}(\xi)$ is singular when $\xi = \xi^{(k)}$. Meanwhile, $\mathbf{\Pi}^{(k)}(\xi^{(k)})$ has L_h zero eigenvalues, i.e., the rank of $\mathbf{\Pi}^{(k)}(\xi^{(k)})$ should be $M - L_h$.
2) From *Lemma 2*, we know when $\xi \neq \xi^{(k)}$, the number of zero eigenvalues of the matrix $\mathbf{\Pi}^{(k)}(\xi)$ should be less than L_h. The implication is that the rank of $\mathbf{\Pi}^{(k)}(\xi)$ should be larger than $M - L_h$, when $\xi \neq \xi^{(k)}$.

Hence, $\mathbf{\Pi}^{(k)}(\xi)$ drops its rank to $M - L_h$ if and only if $\xi = \xi^{(k)}$. Based on the above observations, we design the CFO estimation by one-dimensional (1-D) search as follows. For each trial ξ from $(-0.5, 0.5)$,

1) compute the M eigenvalues of the matrix $\mathbf{\Pi}^{(k)}(\xi)$, denoted by $\lambda_1^{(k)}(\xi), \lambda_2^{(k)}(\xi), \cdots, \lambda_M^{(k)}(\xi)$ in an ascending order;
2) calculate the summation of the smallest L_h eigenvalues of $\mathbf{\Pi}^{(k)}(\xi)$ as the cost of the current trial value, denoted by

$$g_k(\xi) = \sum\nolimits_{i=1}^{L_h} \lambda_i^{(k)}(\xi). \tag{26}$$

The CFO estimation for the kth user is the trial value that gives the minimum cost, i.e.,

$$\hat{\xi}^{(k)} = \arg\min_{\xi} g_k(\xi). \tag{27}$$

Another observation from *Lemma 1* is that, the matrix $\mathbf{H}^{(k)}$ constitutes the L_h dimensional orthogonal complement space of $\mathbf{\Pi}^{(k)}(\xi^{(k)})$, i.e., $\text{Span}(\mathbf{H}^{(k)}) = \left(\mathbf{\Pi}^{(k)}(\xi^{(k)})\right)^{\perp}$. Therefore, after deriving the CFO of the kth user from (27), $\left(\mathbf{\Pi}^{(k)}(\hat{\xi}^{(k)})\right)^{\perp}$ serves as the estimation of $\mathbf{H}^{(k)}$ with certain ambiguity, which is one byproduct of the proposed estimator. Note that this multi-dimensional ambiguity is well known for the blind channel estimation in multi-antenna systems and can be removed by transmitting several pilot symbols [26]. The opportunity here is that, each user has an individual ambiguity matrix rather than a huge ambiguity matrix over all K users. Therefore, the amount of the required training to remove such an ambiguity is much less.

Remark: From above discussions, we know that when the number of antenna elements at the receiver is larger than the number of multipaths from each user, i.e., $M > L_h$, our subspace method could estimate each user's CFO individually and support fully loaded transmissions. Considering that the antenna number at the receiver is limited in practice, e.g., $M = 8$, our estimator is applicable to the channels with a relatively small number of multipaths.

IV. THE ML ESTIMATOR

A. The Estimator

We denote the parameter set for the kth user by $\boldsymbol{\omega}^{(k)} = \{\xi^{(k)}, \mathbf{H}^{(k)}\}$, and the corresponding whole parameter set for the K users by $\boldsymbol{\omega} = \{\boldsymbol{\omega}^{(1)}, \boldsymbol{\omega}^{(2)}, \cdots, \boldsymbol{\omega}^{(K)}\}$. Rewrite (19) as the following parameterized equation

$$\mathbf{a}_{p,g} = \mathbf{V}(\boldsymbol{\omega})\mathbf{z}_{p,g} + \bar{\mathbf{n}}_{p,g}, \tag{28}$$

where

$$\mathbf{V}(\boldsymbol{\omega}) = \left[\mathbf{V}^{(1)}(\boldsymbol{\omega}^{(1)}), \mathbf{V}^{(2)}(\boldsymbol{\omega}^{(2)}), \cdots, \mathbf{V}^{(K)}(\boldsymbol{\omega}^{(K)})\right],$$

$$\mathbf{V}^{(k)}(\boldsymbol{\omega}^{(k)}) = \mathbf{v}^{(k)}(\xi^{(k)}) \otimes \mathbf{H}^{(k)},$$

$$\mathbf{v}^{(k)}(\xi^{(k)}) = \left[1, e^{\mathbf{j}2\pi \frac{q^{(k)}+\xi^{(k)}}{K}}, \cdots, e^{\mathbf{j}2\pi(K-1)\frac{q^{(k)}+\xi^{(k)}}{K}}\right]^T.$$

Placing the vectors $\mathbf{a}_{p,g}$, $p = 0, 1, \cdots, P-1$, next to each other, we obtain the following $MK \times P$ matrix

$$\mathbf{A}_g = \left[\mathbf{a}_{0,g}, \mathbf{a}_{1,g}, \cdots, \mathbf{a}_{P-1,g}\right] = \mathbf{V}(\boldsymbol{\omega})\mathbf{Z}_g + \bar{\mathbf{N}}_g, \tag{29}$$

where $\mathbf{Z}_g = [\mathbf{z}_{0,g}, \cdots, \mathbf{z}_{P-1,g}]$ and $\bar{\mathbf{N}}_g = [\bar{\mathbf{n}}_{0,g}, \cdots, \bar{\mathbf{n}}_{P-1,g}]$. We then place L_s continuously observed OFDMA blocks next to each other and obtain the following $MK \times PL_s$ matrix

$$\mathbf{A} = \left[\mathbf{A}_g, \mathbf{A}_{g+1}, \cdots, \mathbf{A}_{g+L_s-1}\right] = \mathbf{V}(\boldsymbol{\omega})\mathbf{Z} + \mathbf{N}, \tag{30}$$

where $\mathbf{Z}$ and $\mathbf{N}$ denote the corresponding signal and AWGN matrices, i.e., $\mathbf{Z} = [\mathbf{Z}_g, \mathbf{Z}_{g+1}, \cdots, \mathbf{Z}_{g+L_s-1}]$ and $\mathbf{N} = [\bar{\mathbf{N}}_g, \bar{\mathbf{N}}_{g+1}, \cdots, \bar{\mathbf{N}}_{g+L_s-1}]$. Following the observation in Appendix A, matrix $\mathbf{V}(\boldsymbol{\omega})$ will be tall and has full column rank when $M > L_h$.

$$\mathbf{V}\big(\tilde{\boldsymbol{\omega}}^{(k)}, \hat{\boldsymbol{\omega}}_{|k,i}\big) = \Big[\underbrace{\mathbf{V}^{(k)}(\tilde{\boldsymbol{\omega}}^{(k)})}_{\mathbf{C}(\tilde{\boldsymbol{\omega}}^{(k)})}, \underbrace{\mathbf{V}^{(1)}(\hat{\boldsymbol{\omega}}_{i+1}^{(1)}), \cdots, \mathbf{V}^{(k-1)}(\hat{\boldsymbol{\omega}}_{i+1}^{(k-1)}), \mathbf{V}^{(k+1)}(\hat{\boldsymbol{\omega}}_i^{(k+1)}), \cdots, \mathbf{V}^{(K)}(\hat{\boldsymbol{\omega}}_i^{(K)})}_{\mathbf{B}(\hat{\boldsymbol{\omega}}_{|k,i})}\Big], \qquad (36)$$

The ML parameter estimation of $\boldsymbol{\omega}, \mathbf{Z}$ can be obtained from (30) as

$$\{\hat{\boldsymbol{\omega}}, \hat{\mathbf{Z}}\} = \arg\min_{\tilde{\boldsymbol{\omega}}, \tilde{\mathbf{Z}}} \big\|\mathbf{A} - \mathbf{V}(\tilde{\boldsymbol{\omega}})\tilde{\mathbf{Z}}\big\|^2, \qquad (31)$$

where $\tilde{\mathbf{x}}$ stands for the trial value of $\mathbf{x}$. For a given $\mathbf{V}(\tilde{\boldsymbol{\omega}})$, by substituting the ML solution $\tilde{\mathbf{Z}} = [\mathbf{V}(\tilde{\boldsymbol{\omega}})]^\dagger \mathbf{A}$ into the above equation, the ML estimation becomes

$$\hat{\boldsymbol{\omega}} = \arg\min_{\tilde{\boldsymbol{\omega}}} \big\|\mathbf{A} - \mathbf{P_V}(\tilde{\boldsymbol{\omega}})\mathbf{A}\big\|^2 = \arg\max_{\tilde{\boldsymbol{\omega}}} \mathrm{Tr}\Big\{\mathbf{A}^H \mathbf{P_V}(\tilde{\boldsymbol{\omega}})\mathbf{A}\Big\}, \qquad (32)$$

where $\mathbf{P_V}(\tilde{\boldsymbol{\omega}}) = \mathbf{V}(\tilde{\boldsymbol{\omega}})[\mathbf{V}(\tilde{\boldsymbol{\omega}})]^\dagger$ is the projection operator onto the space spanned by the columns of the matrix $\mathbf{V}(\tilde{\boldsymbol{\omega}})$. Note that, the parameter $\tilde{\boldsymbol{\omega}}$ not only consists of the unknown CFO items of all users, but also the unknown channel responses. As a result, directly deriving the ML solution for (32) is computationally expensive. We then apply AP algorithm [25] to reduce the multi-dimensional searching to a series of 1-D searching.

B. Iterative Estimation via Alternating Projection

The estimation procedure consists of cycles and steps, where a cycle is made of K steps. Each step updates the CFO and the channel of a single user while keeping the other CFOs and channels constant at their most updated values. We follow the natural ordering $k = 1, 2, \cdots, K$ in updating the users' parameters. Denote $\hat{\boldsymbol{\omega}}_i^{(k)} = \{\hat{\xi}_i^{(k)}, \hat{\mathbf{H}}_i^{(k)}\}$ as the estimate of $\boldsymbol{\omega}^{(k)}$ at the ith cycle, and define

$$\hat{\boldsymbol{\omega}}_{|k,i} = \{\hat{\boldsymbol{\omega}}_{i+1}^{(1)}, \cdots, \hat{\boldsymbol{\omega}}_{i+1}^{(k-1)}, \hat{\boldsymbol{\omega}}_i^{(k+1)}, \cdots, \hat{\boldsymbol{\omega}}_i^{(K)}\}. \qquad (33)$$

At the kth step of the $(i + 1)$th cycle, the alternating projection algorithm updates the estimate of $\boldsymbol{\omega}^{(k)}$ by solving the following problem:

$$\hat{\boldsymbol{\omega}}_{i+1}^{(k)} = \arg\max_{\tilde{\boldsymbol{\omega}}^{(k)}} \mathrm{Tr}\Big\{\mathbf{A}^H \mathbf{P_V}\big(\tilde{\boldsymbol{\omega}}^{(k)}, \hat{\boldsymbol{\omega}}_{|k,i}\big)\mathbf{A}\Big\}, \qquad (34)$$

where $\mathbf{P_V}\big(\tilde{\boldsymbol{\omega}}^{(k)}, \hat{\boldsymbol{\omega}}_{|k,i}\big)$ is used to indicate the functional dependence of $\mathbf{P_V}$ on

$$\{\hat{\boldsymbol{\omega}}_{i+1}^{(1)}, \cdots, \hat{\boldsymbol{\omega}}_{i+1}^{(k-1)}, \tilde{\boldsymbol{\omega}}^{(k)}, \hat{\boldsymbol{\omega}}_i^{(k+1)}, \cdots, \hat{\boldsymbol{\omega}}_i^{(K)}\}. \qquad (35)$$

Similar to [25], we observe that most columns of $\mathbf{V}\big(\tilde{\boldsymbol{\omega}}^{(k)}, \hat{\boldsymbol{\omega}}_{|k,i}\big)$ are fixed when updating $\tilde{\boldsymbol{\omega}}^{(k)}$. Thus, we split $\mathbf{V}\big(\tilde{\boldsymbol{\omega}}^{(k)}, \hat{\boldsymbol{\omega}}_{|k,i}\big)$ into two parts in (36) on top of this page, where $\mathbf{C}(\tilde{\boldsymbol{\omega}}^{(k)}) \in \mathbb{C}^{MK \times L_h}$ and $\mathbf{B}(\hat{\boldsymbol{\omega}}_{|k,i}) \in \mathbb{C}^{MK \times (K-1)L_h}$ are defined as the corresponding items. From the well known projection formula, we derive

$$\mathbf{P_V}\big(\tilde{\boldsymbol{\omega}}^{(k)}, \hat{\boldsymbol{\omega}}_{|k,i}\big) = \mathbf{P_B}\big(\hat{\boldsymbol{\omega}}_{|k,i}\big) + \mathbf{P_{C_B}}\big(\tilde{\boldsymbol{\omega}}^{(k)}, \hat{\boldsymbol{\omega}}_{|k,i}\big) \qquad (37)$$

where $\mathbf{P_B}\big(\hat{\boldsymbol{\omega}}_{|k,i}\big)$ and $\mathbf{P_{C_B}}\big(\tilde{\boldsymbol{\omega}}^{(k)}, \hat{\boldsymbol{\omega}}_{|k,i}\big)$ denote the projectors defined:

$$\mathbf{P_B}\big(\hat{\boldsymbol{\omega}}_{|k,i}\big) = \mathbf{B}\big(\hat{\boldsymbol{\omega}}_{|k,i}\big)\big[\mathbf{B}\big(\hat{\boldsymbol{\omega}}_{|k,i}\big)\big]^\dagger,$$
$$\mathbf{P_{C_B}}\big(\tilde{\boldsymbol{\omega}}^{(k)}, \hat{\boldsymbol{\omega}}_{|k,i}\big) = \mathbf{P_B^\perp}\big(\hat{\boldsymbol{\omega}}_{|k,i}\big)\mathbf{C}(\tilde{\boldsymbol{\omega}}^{(k)})\big[\mathbf{P_B^\perp}\big(\hat{\boldsymbol{\omega}}_{|k,i}\big)\mathbf{C}(\tilde{\boldsymbol{\omega}}^{(k)})\big]^\dagger.$$

Afterwards, (34) becomes

$$\hat{\boldsymbol{\omega}}_{i+1}^{(k)} = \arg\max_{\tilde{\boldsymbol{\omega}}^{(k)}} \mathrm{Tr}\Big\{\mathbf{A}^H \mathbf{P_{C_B}}\big(\tilde{\boldsymbol{\omega}}^{(k)}, \hat{\boldsymbol{\omega}}_{|k,i}\big)\mathbf{A}\Big\}. \qquad (38)$$

Different from the conventional AP solution in [25], here (38) is still a multi-dimensional problem since $\tilde{\boldsymbol{\omega}}^{(k)} = \{\tilde{\xi}^{(k)}, \tilde{\mathbf{H}}^{(k)}\}$ is composed of both the unknown CFO and channel response matrix for the kth user. Directly solving (38) still needs multi-dimensional search with respect to both CFO $\tilde{\xi}^{(k)}$ and the channel matrix $\tilde{\mathbf{H}}^{(k)} \in \mathbb{C}^{M \times L_h}$, which requires very high computational complexity. However, we will show later that the problem in (38) can be solved by an efficient 1-D search.

We further rewrite (38) as the following equivalent minimization problem:

$$\hat{\boldsymbol{\omega}}_{i+1}^{(k)} = \arg\min_{\tilde{\boldsymbol{\omega}}^{(k)}, \tilde{\mathbf{X}}} \big\|\mathbf{P_B^\perp}\big(\hat{\boldsymbol{\omega}}_{|k,i}\big)\mathbf{C}(\tilde{\boldsymbol{\omega}}^{(k)})\tilde{\mathbf{X}} - \mathbf{A}\big\|^2, \qquad (39)$$

where $\tilde{\mathbf{X}} \in \mathbb{C}^{L_h \times PL_s}$ is a trial matrix that aims to minimize the above cost function. Note that for a given $\tilde{\boldsymbol{\omega}}^{(k)}$, by substituting the ML solution $\tilde{\mathbf{X}} = \big[\mathbf{P_B^\perp}\big(\hat{\boldsymbol{\omega}}_{|k,i}\big)\mathbf{C}(\tilde{\boldsymbol{\omega}}^{(k)})\big]^\dagger \mathbf{A}$ into (39), we can arrive at (38) after some simple manipulations. For a better understanding of the introduced trial matrix $\tilde{\mathbf{X}}$ in (39), let us rewrite $\mathbf{Z} = [(\mathbf{Z}^{(1)})^T, (\mathbf{Z}^{(2)})^T, \cdots, (\mathbf{Z}^{(K)})^T]$ with $\mathbf{Z}^{(k)} = [\mathbf{Z}_g^{(k)}, \mathbf{Z}_{g+1}^{(k)}, \cdots, \mathbf{Z}_{g+L_s-1}^{(k)}]$ and $\mathbf{Z}_g^{(k)} = [\mathbf{z}_{0,g}^{(k)}, \mathbf{z}_{1,g}^{(k)}, \cdots, \mathbf{z}_{P-1,g}^{(k)}]$. Denote $\mathbf{Z}_{\mathbf{B}}^{(k)}$ as the submatrix of $\mathbf{Z}$ formed by deleting the corresponding rows of $\mathbf{Z}^{(k)}$. Then, (34) can be rewritten as

$$\hat{\boldsymbol{\omega}}_{i+1}^{(k)} = \arg\min_{\tilde{\boldsymbol{\omega}}^{(k)}, \tilde{\mathbf{Z}}^{(k)}, \tilde{\mathbf{Z}}_{\mathbf{B}}^{(k)}} \big\|\mathbf{A} - \mathbf{C}(\tilde{\boldsymbol{\omega}}^{(k)})\tilde{\mathbf{Z}}^{(k)} - \mathbf{B}\big(\hat{\boldsymbol{\omega}}_{|k,i}\big)\tilde{\mathbf{Z}}_{\mathbf{B}}^{(k)}\big\|^2. \qquad (40)$$

By substituting the ML solution $\tilde{\mathbf{Z}}_{\mathbf{B}}^{(k)} = [\mathbf{B}(\hat{\boldsymbol{\omega}}_{|k,i})]^\dagger(\mathbf{A} - \mathbf{C}(\tilde{\boldsymbol{\omega}}^{(k)})\tilde{\mathbf{Z}}^{(k)})$ back into (40), we can rewrite (40) as

$$\begin{aligned}\hat{\boldsymbol{\omega}}_{i+1}^{(k)} &= \arg\min_{\tilde{\boldsymbol{\omega}}^{(k)}, \tilde{\mathbf{Z}}^{(k)}} \big\|\mathbf{P_B^\perp}\big(\hat{\boldsymbol{\omega}}_{|k,i}\big)(\mathbf{A} - \mathbf{C}(\tilde{\boldsymbol{\omega}}^{(k)})\tilde{\mathbf{Z}}^{(k)})\big\|^2 \\ &= \arg\min_{\tilde{\boldsymbol{\omega}}^{(k)}, \tilde{\mathbf{Z}}^{(k)}} \big\|\mathbf{P_B^\perp}\big(\hat{\boldsymbol{\omega}}_{|k,i}\big)\mathbf{A} - \mathbf{P_B^\perp}\big(\hat{\boldsymbol{\omega}}_{|k,i}\big)\mathbf{C}(\tilde{\boldsymbol{\omega}}^{(k)})\tilde{\mathbf{Z}}^{(k)}\big\|^2 \\ &\qquad + \big\|\mathbf{P_B}\big(\hat{\boldsymbol{\omega}}_{|k,i}\big)\mathbf{A}\big\|^2 \\ &= \arg\min_{\tilde{\boldsymbol{\omega}}^{(k)}, \tilde{\mathbf{Z}}^{(k)}} \big\|\mathbf{P_B^\perp}\big(\hat{\boldsymbol{\omega}}_{|k,i}\big)\mathbf{C}(\tilde{\boldsymbol{\omega}}^{(k)})\tilde{\mathbf{Z}}^{(k)} - \mathbf{A}\big\|^2. \qquad (41)\end{aligned}$$

From (39) and (41), we see that $\tilde{\mathbf{X}}$ is equivalent to $\tilde{\mathbf{Z}}^{(k)}$, which serves as the trial value of $\mathbf{Z}^{(k)}$.

Afterwards, by substituting $\mathbf{C}(\tilde{\boldsymbol{\omega}}^{(k)}) = \mathbf{V}^{(k)}(\tilde{\boldsymbol{\omega}}^{(k)}) = \mathbf{v}^{(k)}(\tilde{\xi}^{(k)}) \otimes \tilde{\mathbf{H}}^{(k)}$, we can rewrite (39) as

$$\hat{\boldsymbol{\omega}}_{i+1}^{(k)} = \arg \min_{\tilde{\xi}^{(k)}, \tilde{\mathbf{H}}^{(k)}, \tilde{\mathbf{X}}} \left\| \mathbf{P}_{\mathbf{B}}^{\perp}(\hat{\boldsymbol{\omega}}_{|k,i}) \left(\mathbf{v}^{(k)}(\tilde{\xi}^{(k)}) \otimes \tilde{\mathbf{H}}^{(k)} \right) \tilde{\mathbf{X}} - \mathbf{A} \right\|^2. \tag{42}$$

Based on the fact that $\mathbf{v}^{(k)}(\tilde{\xi}^{(k)}) \otimes \tilde{\mathbf{H}}^{(k)} = \left(\mathbf{v}^{(k)}(\tilde{\xi}^{(k)}) \otimes \mathbf{I}_M \right) \tilde{\mathbf{H}}^{(k)}$, (42) becomes

$$\hat{\boldsymbol{\omega}}_{i+1}^{(k)}$$
$$= \arg \min_{\tilde{\xi}^{(k)}, \tilde{\mathbf{H}}^{(k)}, \tilde{\mathbf{X}}} \left\| \underbrace{\mathbf{P}_{\mathbf{B}}^{\perp}(\hat{\boldsymbol{\omega}}_{|k,i}) \left(\mathbf{v}^{(k)}(\tilde{\xi}^{(k)}) \otimes \mathbf{I}_M \right)}_{\mathbf{\Pi}(\hat{\boldsymbol{\omega}}_{|k,i}, \tilde{\xi}^{(k)})} \tilde{\mathbf{H}}^{(k)} \tilde{\mathbf{X}} - \mathbf{A} \right\|^2$$
$$= \arg \min_{\tilde{\xi}^{(k)}, \tilde{\mathbf{H}}^{(k)}, \tilde{\mathbf{X}}} \left\| \mathbf{\Pi}(\hat{\boldsymbol{\omega}}_{|k,i}, \tilde{\xi}^{(k)}) \tilde{\mathbf{H}}^{(k)} \tilde{\mathbf{X}} - \mathbf{A} \right\|^2. \tag{43}$$

In Appendix E, we prove that $\mathbf{\Pi}(\hat{\boldsymbol{\omega}}_{|k,i}, \tilde{\xi}^{(k)}) \in \mathbb{C}^{MK \times M}$ has full column rank M. If we treat $\tilde{\mathbf{H}}^{(k)} \tilde{\mathbf{X}} \in \mathbb{C}^{M \times PL_s}$ as a single unknown matrix, the minimization in (43) resembles the conventional least square problem. However, the important difference is that the rank of matrix $\tilde{\mathbf{H}}^{(k)} \tilde{\mathbf{X}}$ is smaller than or equal to L_h.

Let the SVD of $\mathbf{\Pi}(\hat{\boldsymbol{\omega}}_{|k,i}, \tilde{\xi}^{(k)})$ be

$$\mathbf{\Pi}(\hat{\boldsymbol{\omega}}_{|k,i}, \tilde{\xi}^{(k)}) = \left[\mathbf{U}_{\mathbf{\Pi}}, \mathbf{U}_{\mathbf{\Pi},0} \right] \begin{bmatrix} \mathbf{\Sigma}_{\mathbf{\Pi}} \\ \mathbf{0} \end{bmatrix} \mathbf{V}_{\mathbf{\Pi}}^H, \tag{44}$$

where the diagonal entries of $\mathbf{\Sigma}_{\mathbf{\Pi}} \in \mathbb{C}^{M \times M}$ are the nonzero singular values, $\mathbf{V}_{\mathbf{\Pi}} \in \mathbb{C}^{M \times M}$ denotes the right singular matrix, $\mathbf{U}_{\mathbf{\Pi}} \in \mathbb{C}^{MK \times M}$ and $\mathbf{U}_{\mathbf{\Pi},0} \in \mathbb{C}^{MK \times M(K-1)}$ denote the left singular vector matrices that correspond to the M nonzero and the rest zero singular values, respectively.

It can be further obtained that

$$\left\| \mathbf{\Pi}(\hat{\boldsymbol{\omega}}_{|k,i}, \tilde{\xi}^{(k)}) \tilde{\mathbf{H}}^{(k)} \tilde{\mathbf{X}} - \mathbf{A} \right\|^2$$
$$= \left\| \begin{bmatrix} \mathbf{\Sigma}_{\mathbf{\Pi}} \\ \mathbf{0} \end{bmatrix} \mathbf{V}_{\mathbf{\Pi}}^H \tilde{\mathbf{H}}^{(k)} \tilde{\mathbf{X}} - \begin{bmatrix} \mathbf{U}_{\mathbf{\Pi}}^H \\ \mathbf{U}_{\mathbf{\Pi},0}^H \end{bmatrix} \mathbf{A} \right\|^2$$
$$= \left\| \mathbf{\Sigma}_{\mathbf{\Pi}} \mathbf{V}_{\mathbf{\Pi}}^H \tilde{\mathbf{H}}^{(k)} \tilde{\mathbf{X}} - \mathbf{U}_{\mathbf{\Pi}}^H \mathbf{A} \right\|^2 + \left\| \mathbf{U}_{\mathbf{\Pi},0}^H \mathbf{A} \right\|^2$$
$$= \left\| \mathbf{A} \right\|^2 + \left\| \mathbf{\Sigma}_{\mathbf{\Pi}} \mathbf{V}_{\mathbf{\Pi}}^H \tilde{\mathbf{H}}^{(k)} \tilde{\mathbf{X}} - \mathbf{U}_{\mathbf{\Pi}}^H \mathbf{A} \right\|^2 - \left\| \mathbf{U}_{\mathbf{\Pi}}^H \mathbf{A} \right\|^2. \tag{45}$$

As a result, we can rewrite (43) as

$$\hat{\boldsymbol{\omega}}_{i+1}^{(k)} = \arg \min_{\tilde{\xi}^{(k)}, \tilde{\mathbf{H}}^{(k)}, \tilde{\mathbf{X}}} \left\| \mathbf{\Sigma}_{\mathbf{\Pi}} \mathbf{V}_{\mathbf{\Pi}}^H \tilde{\mathbf{H}}^{(k)} \tilde{\mathbf{X}} - \mathbf{U}_{\mathbf{\Pi}}^H \mathbf{A} \right\|^2 - \left\| \mathbf{U}_{\mathbf{\Pi}}^H \mathbf{A} \right\|^2. \tag{46}$$

It can be observed that the first term of (46) falls into the well-known low rank matrix approximation problem, where we can find a matrix $\mathbf{\Sigma}_{\mathbf{\Pi}} \mathbf{V}_{\mathbf{\Pi}}^H \tilde{\mathbf{H}}^{(k)} \tilde{\mathbf{X}}$ whose rank is at most L_h and could approximate the matrix $\mathbf{U}_{\mathbf{\Pi}}^H \mathbf{A}$ in the best way. Denote the M left and right singular vectors and the corresponding singular values of $\mathbf{U}_{\mathbf{\Pi}}^H \mathbf{A}$ by $\boldsymbol{u}_l$, $\boldsymbol{v}_l$ and σ_l, respectively, i.e.,

$$\mathbf{U}_{\mathbf{\Pi}}^H \mathbf{A} = \sum_{l=1}^{M} \sigma_l \boldsymbol{u}_l \boldsymbol{v}_l^H, \tag{47}$$

where the singular values are listed in descending order.

From the SVD property [27], we know the first term of (46) achieves its minimum

$$\sum_{l=L_h+1}^{M} \sigma_l^2 \tag{48}$$

when

$$\mathbf{\Sigma}_{\mathbf{\Pi}} \mathbf{V}_{\mathbf{\Pi}}^H \tilde{\mathbf{H}}^{(k)} \tilde{\mathbf{X}} = \sum_{l=1}^{L_h} \sigma_l \boldsymbol{u}_l \boldsymbol{v}_l^H, \tag{49}$$

or equivalently,

$$\tilde{\mathbf{H}}^{(k)} \tilde{\mathbf{X}} = \mathbf{V}_{\mathbf{\Pi}} \mathbf{\Sigma}_{\mathbf{\Pi}}^{-1} \sum_{l=1}^{L_h} \sigma_l \boldsymbol{u}_l \boldsymbol{v}_l^H. \tag{50}$$

With the aid of (48)-(50), the multi-dimensional problem (46) is simplified to a 1-D problem

$$\hat{\xi}_{i+1}^{(k)} = \arg \min_{\tilde{\xi}^{(k)}} \left(\sum_{l=L_h+1}^{M} \sigma_l^2 - \sum_{l=1}^{M} \sigma_l^2 \right) = \arg \max_{\tilde{\xi}^{(k)}} \sum_{l=1}^{L_h} \sigma_l^2, \tag{51}$$

which yields the CFO estimation for the kth user at the $(i+1)$th cycle.

Based on (50), the corresponding channel estimation at this step is then performed as follows. Suppose the SVD of $\mathbf{\Pi}(\hat{\boldsymbol{\omega}}_{|k,i}, \hat{\xi}_{i+1}^{(k)})$ is $\mathbf{\Pi}(\hat{\boldsymbol{\omega}}_{|k,i}, \hat{\xi}_{i+1}^{(k)}) = \hat{\mathbf{U}}_{\mathbf{\Pi}} \hat{\mathbf{\Sigma}}_{\mathbf{\Pi}} \hat{\mathbf{V}}_{\mathbf{\Pi}}^H$. Denote $\hat{\boldsymbol{u}}_l$ $(l = 1, 2, \cdots, L_h)$ as the left singular vectors that correspond to the largest L_h singular values of $\hat{\mathbf{U}}_{\mathbf{\Pi}}^H \mathbf{A}$. Note that in (50), we can only estimate the column space of channel matrix with some ambiguity. Specifically, according to (50) the kth user's channel estimation at this step can be obtained from

$$\hat{\mathbf{H}}_{i+1}^{(k)} = \hat{\mathbf{V}}_{\mathbf{\Pi}} \hat{\mathbf{\Sigma}}_{\mathbf{\Pi}}^{-1} \left[\hat{\boldsymbol{u}}_1, \hat{\boldsymbol{u}}_2, \cdots, \hat{\boldsymbol{u}}_{L_h} \right]. \tag{52}$$

The 1-D search at the kth step in the $(i+1)$th cycle is summarized as follows: For each trial $\tilde{\xi}^{(k)}$ from (-0.5,0.5),
1) calculate $\mathbf{\Pi}(\hat{\boldsymbol{\omega}}_{|k,i}, \tilde{\xi}^{(k)})$ according to (43);
2) calculate $\mathbf{U}_{\mathbf{\Pi}}$ according to (44);
3) perform SVD operation on $\mathbf{U}_{\mathbf{\Pi}}^H \mathbf{A}$, and obtain its largest L_h singular values σ_l;
4) calculate the utility for current trial as $\sum_{l=1}^{L_h} \sigma_l^2$.

The trial $\tilde{\xi}^{(k)}$ that yields the maximum utility is the estimated CFO at this step, and the channel estimation is then obtained from (52).

Remark: It is well known that the initialization procedure is important for the iterative ML solutions. Thanks to the proposed subspace estimator, we can take its output as the initial estimates. It is also observed in the later simulations that the ML approach with initial values obtained from subspace estimator work well under different scenarios.

V. SIMULATIONS

In this section, we assess the proposed CFO estimation algorithms from computer simulations. The total number of subcarriers is taken as $N = 64$ and are divided into $Q = 4$ subchannels. The quadrature phase-shift keying (QPSK) constellation is adopted. The normalized CFO of each user is randomly generated from -0.4 to 0.4, and the delays of the multipath from each user are randomly generated from 0 to

8. In simulations, we use 120 CFO trial values to derive the solutions in (27) and (51).[1] The root mean square error (RMSE) of the normalized CFO estimation is adopted as the figure of merit. Fully loaded transmissions are assumed in our estimators. In addition, we also include the following two CFO estimation schemes for comparison:

1) 'ESPRIT-I': the scheme proposed in [17]. For fairness, the multiple receive antenna diversity is exploited as described in [15, eq. (28)], in which the multiple observations from multiple antennas are collected together to improve the estimation of the covariance matrix $\mathbf{R}_{zz}$ in [17]. Note that in this scheme, one subchannel has to be reserved as null subcarriers; thus, the number of users is $Q - 1 = 3$ in ESPRIT-I.

2) 'ESPRIT-II': an extended scheme based on [17] which supports fully loaded transmissions, i.e., $K = Q$. We derive this scheme by exploiting the rotational invariance property from (18). Specifically, we define $\mathbf{E}_x \in \mathbb{C}^{(K-1)M \times KL_h}$ as the first $(K-1)M$ rows of $\mathbf{U_s}$ and $\mathbf{E}_y \in \mathbb{C}^{(K-1)M \times KL_h}$ as the last $(K-1)M$ rows of $\mathbf{U_s}$. Based on the observations from (18) to (21), we have $\mathbf{E}_x = \mathbf{V}_x \mathbf{T}$ and $\mathbf{E}_y = \mathbf{V}_x \mathbf{\Phi T}$, where $\mathbf{V}_x \in \mathbb{C}^{(K-1)M \times KL_h}$ denotes the first $(K-1)M$ rows of $\mathbf{V}$ and $\mathbf{T} \in \mathbb{C}^{KL_h \times KL_h}$ is a full-rank ambiguity matrix. Then, according to the ESPRIT algorithm [18], the KL_h eigenvalues of $\mathbf{E}_x^\dagger \mathbf{E}_y$ correspond to the KL_h diagonal elements of $\mathbf{\Phi}$, by which the CFOs of the multiple users can be further derived. Let us take the kth user for example. After eigenvalue decomposition of $\mathbf{E}_x^\dagger \mathbf{E}_y$, we obtain L_h eigenvalues corresponding to the kth user's effective CFO, i.e., $e^{\mathbf{j}2\pi\theta^{(k)}}$, each of which provides one CFO estimation value like [17]. Then, the final CFO estimation for the kth user is obtained by averaging these L_h values. Since in ESPRIT-II, $\mathbf{V}_x$ should have full column rank, $M \geq \frac{K}{K-1}L_h$ is required.

We start by evaluating the performance of our subspace estimator, referred to as 'SSE'. We consider the frequency selective fading scenarios with $L_h = 2$. Fig. 2 shows the CFO estimation RMSEs versus signal-to-noise ratio (SNR) under different number of antenna elements and different number of sample blocks. As expected, the estimation performance of SSE is improved with the increase of SNR, the antenna number, as well as the sample block number. We see that with $M = 4$ and only one sample block, the performance of SSE is worse than that of ESPRIT-I under low and moderate SNR condition. Nevertheless, SSE behaves better with a few more OFDMA blocks or more antennas. This may be explained as follows. The equivalent numbers of snapshots to estimate the covariance matrix in ESPRIT-I and SSE are $L_s PM$ and $L_s P$, respectively. Thus, the estimated covariance matrix of SSE exhibits larger perturbations than that of ESPRIT-I, making SSE perform worse due to occurrence of outliers when few sample blocks are available. On the other side, we see that the

[1]We first obtain a coarse CFO estimation value by searching from -0.5 to 0.5 with an interval of 10^{-2}. Then, the fine estimation is achieved by searching within the neighborhood with size of 0.02 around the coarse value using an interval of 10^{-3}. Thus, the number of total trail CFO values is $\alpha = 120$.
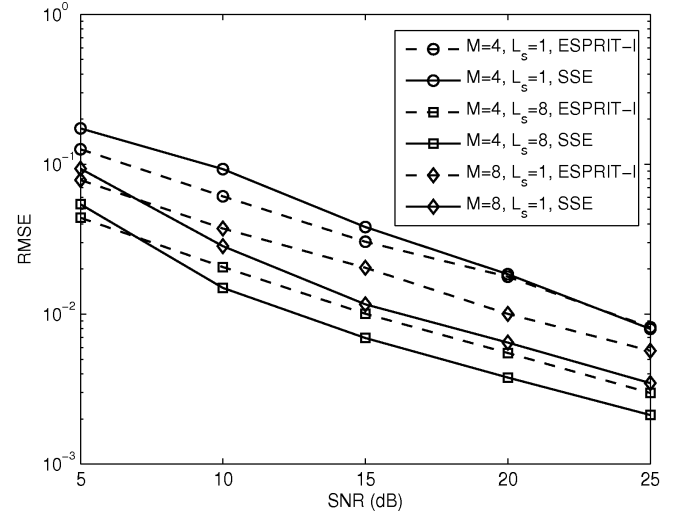

Fig. 2. The CFO estimation performance versus SNR between SSE and ESPRIT-I under frequency selective fading channels ($L_h = 2$). The numbers of users in SSE and ESPRIT-I are $K = 4$ and $K = 3$, respectively.
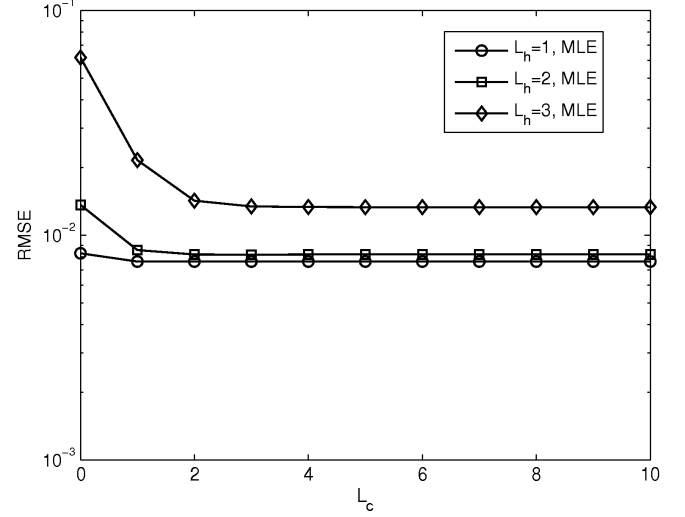

Fig. 3. The CFO estimation RMSE convergence process of MLE with different number of multipaths (SNR=20 dB, $M = 4$, $L_s = 1$). The results of $L_c = 0$ correspond to the performance of SSE.

equivalent array apertures of ESPRIT-I and SSE are Q and MQ, respectively, where the former is not related with the number of antennas; thus, SSE has a larger aperture, which provides the potential of higher resolution.

Next, we investigate the convergence rate of the proposed ML estimator, referred to as 'MLE'. Fig. 3 depicts the CFO estimation performance versus the number of cycles L_c under $M = 4$ and SNR=20 dB, where we vary the number of multipaths from $L_h = 1$ to $L_h = 3$. Only one block is utilized in this example, i.e., $L_s = 1$. Bearing in mind that, when $L_c = 0$, no iteration is conducted, and thus, it is equivalent to our proposed SSE. The results explicitly demonstrate the performance improvement introduced by the proposed MLE, especially under the scenarios with more multipaths. We also observe that for the three different scenarios, only two cycles
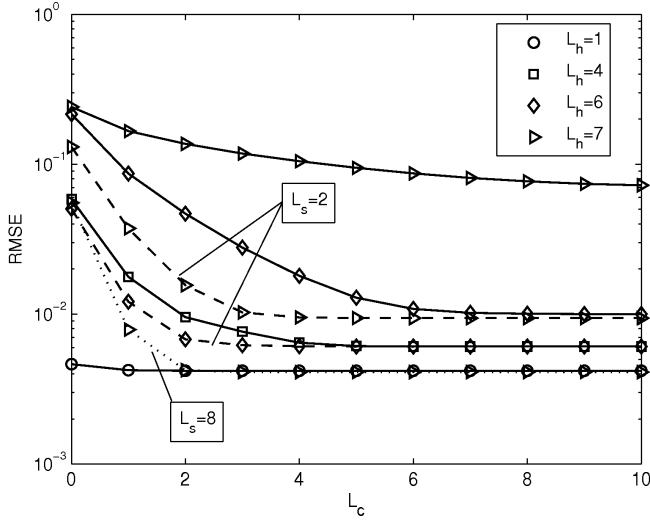
Fig. 4. The CFO estimation RMSE convergence process of MLE with different number of multipaths (SNR=20 dB, $M = 8$). The solid, dashed and dotted curves correspond to the performance with $L_s = 1$, $L_s = 2$, $L_s = 8$, respectively.
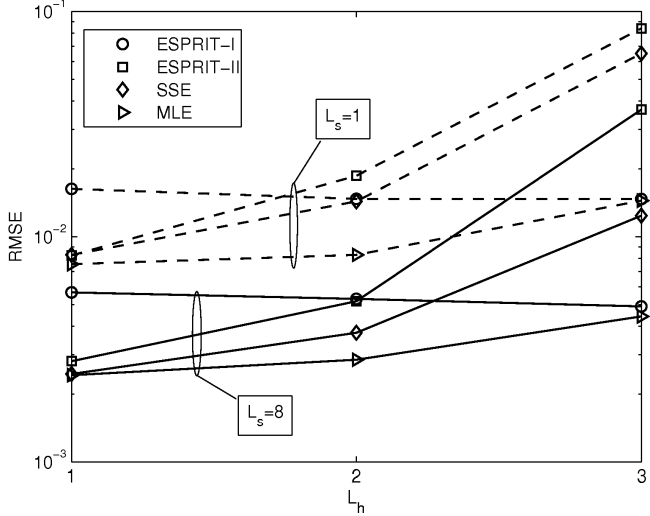

Fig. 6. The CFO estimation RMSE performance of different approaches versus L_h (SNR=20 dB, $M = 8$). The dashed and solid curves correspond to the performance with $L_s = 1$ and $L_s = 8$, respectively.

Fig. 5. The CFO estimation RMSE performance of different approaches versus L_h (SNR=20 dB, $M = 4$). The dashed and solid curves correspond to the performance with $L_s = 1$ and $L_s = 8$, respectively.

are needed for convergence and no improvement can be obtained with additional cycles, which demonstrates the high convergence rate of the proposed MLE. Moreover, it is clear that the increase of the number of multipaths will decrease the estimation convergence rate. From the results, only one cycle is enough for convergence when $L_h \leq 2$, whereas two cycles are required when $L_h = 3$.

Furthermore, we show the performance convergence process in Fig. 4 when the receiver is equipped with $M = 8$ antennas. In this example, the solid, dashed and dotted curves represent the performance with one block, two blocks and eight blocks, respectively. The number of multipaths is varied from 1 to 7. It is shown that, when only one block is utilized, MLE needs more cycles when the number of multipaths is large, e.g., $L_h \geq 6$, and it also suffers from great performance
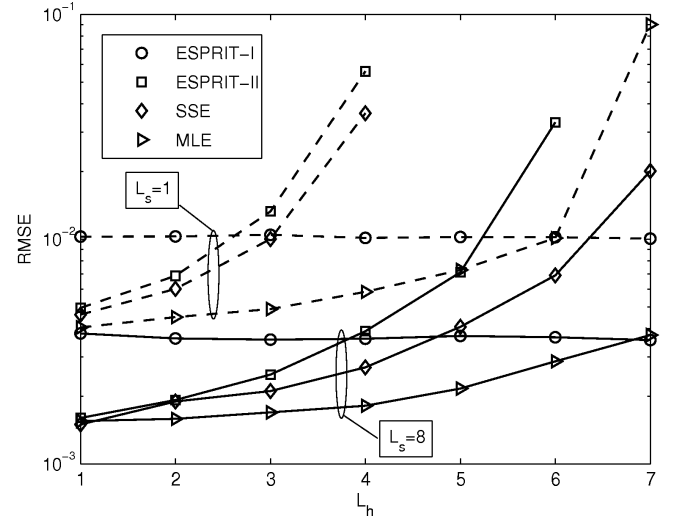
deterioration when $L_h = 7$. However, it can be observed that, when additional blocks are utilized, i.e., $L_s = 2$ or $L_s = 8$, both RMSE and the convergence rate can be greatly improved. We see that with $L_s = 8$ blocks, two cycles are enough for convergence of MLE under $L_h = 7$ multipaths.

The CFO estimation performance versus different values of L_h are shown in Fig. 5 and Fig. 6 under $M = 4$ and $M = 8$, respectively, where SNR=20 dB is taken. The performance from the four estimators, i.e., ESPRIT-I, ESPRIT-II, SSE and MLE, are all included for comparison. Clearly, we see that the ESPRIT-I has the advantage of being insensitive to the channel frequency selectivity, whereas both ESPRIT-II and our estimators suffer from the performance degradation with a larger number of multipaths. From these two figures, it is seen that, SSE could achieve better performance than ESPRIT-I when L_h is under some threshold, otherwise, SSE behaves worse. On the other side, we see in Fig. 5 that MLE behaves better than ESPRIT-I under all the scenarios when $M = 4$. With $M = 8$ antennas at the receiver in Fig. 6, MLE is superior to ESPRIT-I when $L_h \leq 6$ with only one block, while with eight blocks, MLE almost behaves better than ESPRIT-I under all values of L_h from 1 to 7. The reason why MLE yields excellent performance is twofold. First, MLE is derived based on ML criterion. Second, MLE has exploited the inherent relationship among the observations from multiple receiving antennas. Moreover, we need to emphasize that, our estimators could estimate not only the CFOs for all users but also the channels, which is a property that ESPRIT-I does not possess. Meanwhile, the proposed SSE and MLE can handle more users than ESPRIT-I. On the other side, although ESPRIT-II could also support fully loaded transmissions, it presents higher RMSEs than our SSE and MLE under all scenarios, especially with a larger number of multipaths.

Next, we display the CFO estimation performance versus SNR in Fig. 7, where the results of ESPRIT-I, ESPRIT-II and MLE are included. We assume $L_s = 8$ and $M = 8$ in this
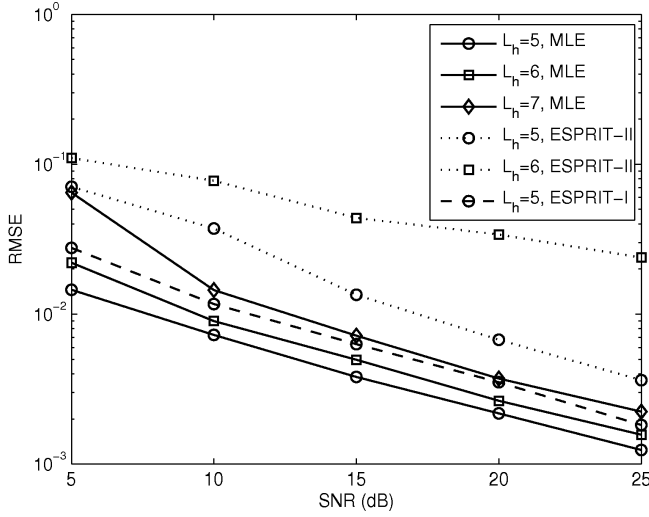
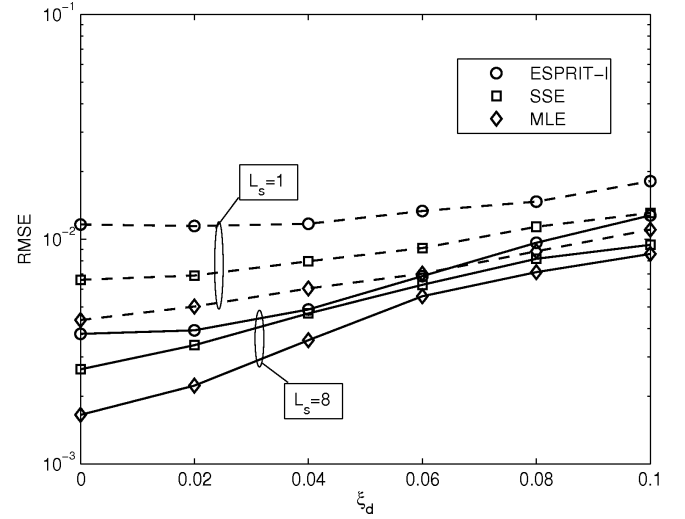Fig. 7. The CFO estimation RMSE performance versus SNR of different approaches ($L_s = 8$, $M = 8$).

Fig. 8. The CFO estimation RMSE performance of different approaches versus the maximum normalized Doppler frequency (SNR=20 dB, $M = 8$, $L_h = 2$). The dashed and solid curves correspond to the performance with $L_s = 1$ and $L_s = 8$, respectively.

TABLE I
COMPUTATIONAL COMPLEXITY ANALYSIS.

Estimator	Complexity
ESPRIT-I ($K=Q-1$)	$\mathcal{O}\left((K+1)^2 MPL_s + (K+1)^3 + 3K^3\right)$
ESPRIT-II	$\mathcal{O}\left(K^2 M^2 PL_s + K^3 M^3 + 2(K-1)MK^2 L_h^2 + K^3 L_h^3\right)$
SSE	$\mathcal{O}\left(K^2 M^2 PL_s + K^3 M^3 + \alpha K(M^2 K^2/2 + M^3)\right)$
One cycle of MLE	$\mathcal{O}\left(\alpha K(M^2 K^2 + M^3 K^2 + 2M^3 K)\right)$

example. Note that we only show the performance of ESPRIT-I with $L_h = 5$, since the value of L_h is verified having little effect on ESPRIT-I. It is demonstrated that, our MLE behaves better than ESPRIT-I when $L_h \leq 6$. When $L_h = 7$, it still achieves a comparable performance with ESPRIT-I under moderate and high SNR region. We should note that the system is fully loaded in our estimators, whereas one subchannel should be reserved in ESPRIT-I. Moreover, although ESPRIT-II also supports fully loaded transmissions, we see that its performance is much worse than our MLE.

We then evaluate the performance of our estimators in fast time-varying channels. Denote ξ_d as the maximum Doppler frequency normalized by subcarrier spacing. We increase ξ_d from 0 to 0.1 and display the RMSE results with $L_s = 1$ and $L_s = 8$ in Fig. 8. The performance of ESPRIT-I is also included for comparison. Meanwhile, we assume SNR=20 dB, $M = 8$ and $L_h = 2$. The sum-of-sinusoids statistical simulation model proposed in [28] is adopted. Since in fast time-varying channels, the channels hardly stay constant over multiple consecutive block durations, we let SSE and MLE first estimate the CFOs for each block duration individually, and then average the results as the final estimation. As expected, the increase of maximum Doppler frequency degrades the performance of all estimators. However, our estimators still perform better than ESPRIT-I. Another observation is that, as compared to the case of $L_s = 1$, the increase of maximum Doppler frequency also weakens the benefit from multiple block durations. This may arise from the effect of estimation bias introduced by the Doppler frequency.

Let us now evaluate the computational complexities of the proposed estimators in terms of complex multiplications. We denote α as the number of trial CFO values to derive the solutions in (27) and (51). In SSE, calculation of $\hat{\mathbf{R}}_a$ and its SVD requires $\mathcal{O}(K^2 M^2 PL_s + K^3 M^3)$; For each trial CFO, calculation of $\mathbf{\Pi}^{(k)}(\xi)$ and its SVD requires $\mathcal{O}(M^2 K^2/2 + M^3)$; The number of total trial CFOs is αK; Then the total complexity of SSE is in the order of

$\mathcal{O}(K^2 M^2 PL_s + K^3 M^3 + \alpha K(M^2 K^2/2 + M^3))$. On the side of MLE, for each trial CFO in one cycle, calculation of $\mathbf{\Pi}(\hat{\omega}_{|k,i}, \tilde{\xi}^{(k)})$ and its SVD requires $\mathcal{O}(M^2 K^2 + M^3 K)$; Calculating the singular values of $\mathbf{U}_{\mathbf{\Pi}}^H \mathbf{A}$ requires $\mathcal{O}(M^3 K^2 + M^3 K)^2$; Then the total complexity of MLE in one cycle is in the order of $\mathcal{O}(\alpha K(M^2 K^2 + M^3 K^2 + 2M^3 K))$. In summary, we list the required complexities for different estimators in Table I, including the results of ESPRIT-I and ESPRIT-II for comparison. For example, assume $M = 8$, $L_s = 8$, $L_h = 6$ and $\alpha = 120$. The computational complexities of ESPRIT-I, ESPRIT-II, SSE, one cycle of MLE are in the order of $\mathcal{O}(1.6 \times 10^4)$, $\mathcal{O}(2 \times 10^5)$, $\mathcal{O}(6.5 \times 10^5)$ and $\mathcal{O}(6.4 \times 10^6)$, respectively. It shows that our estimators suffer from a higher computational burden. Note that as compared to our estimators, ESPRIT-II supports fully loaded transmissions but requires lower computational complexity; however, simulation results have demonstrated that our estimators could provide much better performance. It also needs to be mentioned that the computational complexities of our estimators are still significantly smaller than that of [23] which is in the order of $\mathcal{O}(3(M-1)^3 N^3) = \mathcal{O}(2.7 \times 10^8)$ [23]. Moreover, bearing in mind that the OFDMA receiver usually plays a role as the base

---

[2]We can perform the decomposition $\mathbf{A}\mathbf{A}^H = \mathbf{Q}\mathbf{Q}^H$ in advance where $\mathbf{Q}$ is a $MK \times MK$ matrix. Then, $\mathbf{U}_{\mathbf{\Pi}}^H \mathbf{Q}$ has the same singular values with $\mathbf{U}_{\mathbf{\Pi}}^H \mathbf{A}$. Calculating $\mathbf{U}_{\mathbf{\Pi}}^H \mathbf{Q}$ and its SVD require $\mathcal{O}(M^3 K^2)$ and $\mathcal{O}(M^3 K)$, respectively.

station (BS) in cellular networks, the required complexities should be affordable.

VI. Conclusions

We have developed subspace based and ML based blind CFO estimators for interleaved OFDMA uplink transmission, where multi-antenna is employed at the receiver. We applied the rank reduction approach for the subspace based estimator while adopted the alternating projection for the ML estimator. The proposed schemes support full-load data transmission that allows all subcarriers being allocated to users, which provides better bandwidth efficiency than the existing schemes. Simulation results have shown that the proposed two estimators achieve comparable performance with several existing competitors, especially when there are a few multipaths from each user. Although these advantages are achieved at the price of an increased computational burden, the required complexity may not be the bottleneck, because the receiver usually plays a role as the BS in cellular networks.

Appendix A
Property of the Matrix $\mathbf{V}$

To guarantee $\mathbf{V}$ being a tall matrix, the number of antennas should be larger than the number of multipaths of each user, i.e., $M > L_h$. In the following, we prove that when $M > L_h$, $\mathbf{V}$ will have the full column rank. Rewrite $\mathbf{V}$ into the following row permutation version

$$\bar{\mathbf{V}} = \left[\bar{\mathbf{V}}^{(1)}, \bar{\mathbf{V}}^{(2)}, \cdots, \bar{\mathbf{V}}^{(K)} \right] \tag{53}$$

with $\bar{\mathbf{V}}^{(k)} = \mathbf{H}^{(k)} \otimes \mathbf{v}^{(k)}$. Due to the fact that $\left[\mathbf{v}^{(1)}, \mathbf{v}^{(2)}, \cdots, \mathbf{v}^{(K)} \right]$ is a full rank matrix, the submatrices $\bar{\mathbf{V}}^{(k)}$'s are mutually disjoint, which implies that

$$\text{rank}(\mathbf{V}) = \text{rank}(\bar{\mathbf{V}}) = \sum_{k=1}^{K} \text{rank}(\bar{\mathbf{V}}^{(k)}) = \sum_{k=1}^{K} \text{rank}(\mathbf{H}^{(k)}).$$

Since its entries are i.i.d. Gaussian random variables, when $M > L_h$, $\mathbf{H}^{(k)}$ has full column rank with probability one [29]. Hence, we conclude that $\text{rank}(\mathbf{V}) = KL_h$, i.e., $\mathbf{V}$ has full column rank.

Appendix B
Covariance Matrix $\mathbf{R}_{ZZ}$

Since the transmitted symbols of different users are uncorrelated, there holds

$$E\left[\mathbf{z}_{p,g}^{(k)} (\mathbf{z}_{p,g}^{(k')})^H \right] = \mathbf{0} \tag{54}$$

for $k \neq k'$. On the other hand, we can rewrite $\mathbf{z}_{p,g}^{(k)}$ as

$$\mathbf{z}_{p,g}^{(k)} = \frac{1}{\sqrt{N}} e^{\mathbf{j}\frac{2\pi(g-1)(N+G)\xi^{(k)}}{N}} e^{\mathbf{j}\frac{2\pi(q^{(k)}+\xi^{(k)})p}{N}} \boldsymbol{\varsigma}^{(k)} \boldsymbol{\zeta}^{(k)} \boldsymbol{\chi}_p \mathbf{s}_g^{(k)},$$

where

$$\boldsymbol{\varsigma}^{(k)} = \text{diag}\left(e^{-\mathbf{j}\frac{2\pi q^{(k)}\tau_1^{(k)}}{N}}, e^{-\mathbf{j}\frac{2\pi q^{(k)}\tau_2^{(k)}}{N}}, \cdots, e^{-\mathbf{j}\frac{2\pi q^{(k)}\tau_{L_h}^{(k)}}{N}} \right),$$

$$\boldsymbol{\zeta}^{(k)} = \left[\boldsymbol{\zeta}_1^{(k)}, \boldsymbol{\zeta}_2^{(k)}, \cdots, \boldsymbol{\zeta}_{L_h}^{(k)} \right]^T,$$

$$\boldsymbol{\zeta}_l^{(k)} = \left[1, e^{-\mathbf{j}\frac{2\pi\tau_l^{(k)}}{P}}, \cdots, e^{-\mathbf{j}\frac{2\pi(P-1)\tau_l^{(k)}}{P}} \right]^T,$$

$$\boldsymbol{\chi}_p = \text{diag}\left(1, e^{\mathbf{j}\frac{2\pi p}{P}}, \cdots, e^{\mathbf{j}\frac{2\pi(P-1)p}{P}} \right).$$

Let us define

$$\mathbf{R}_{ZZ}^{(k)} = E\left[\mathbf{z}_{p,g}^{(k)} (\mathbf{z}_{p,g}^{(k)})^H \right] = \frac{\rho_s^{(k)}}{N} \boldsymbol{\varsigma}^{(k)} \boldsymbol{\zeta}^{(k)} \left(\boldsymbol{\varsigma}^{(k)} \boldsymbol{\zeta}^{(k)} \right)^H, \tag{55}$$

where $\rho_s^{(k)}$ denotes the average transmit power of the kth user. We consider $L_h \leq P$ and $\tau_{L_h}^{(k)} < P$. Then the $L_h \times P$ Vandermonde matrix $\boldsymbol{\zeta}^{(k)}$ has rank L_h. As a result, $\mathbf{R}_{ZZ}^{(k)}$ has full rank. Therefore, combining (54) and (55), we obtain that

$$\mathbf{R}_{ZZ} = E[\mathbf{z}_{p,g} \mathbf{z}_{p,g}^H] = \text{blkdiag}\left(\mathbf{R}_{ZZ}^{(1)}, \mathbf{R}_{ZZ}^{(2)}, \cdots, \mathbf{R}_{ZZ}^{(K)} \right), \tag{56}$$

which also possesses the full rank.

Appendix C
Proof of *Lemma 1*

When $\boldsymbol{x} \in \text{Span}(\mathbf{H}^{(k)})$, there exist ϵ_i, $i = 1, 2, \cdots, L_h$, not all zero, such that $\boldsymbol{x} = \sum_{i=1}^{L_h} \epsilon_i \mathbf{h}_i^{(k)}$. Since $\mathbb{B}^{(k)}(\xi^{(k)}) = \mathbf{v}^{(k)}$, we have

$$(\mathbb{B}^{(k)}(\xi^{(k)}) \otimes \boldsymbol{x})^H \mathbf{U_n} = \sum_{i=1}^{L_h} \epsilon_i^* (\mathbf{v}^{(k)} \otimes \mathbf{h}_i^{(k)})^H \mathbf{U_n} = \mathbf{0}. \tag{57}$$

The equality to zero is due to the fact that each $\mathbf{v}^{(k)} \otimes \mathbf{h}_i^{(k)}$, $i = 1, 2, \cdots, L_h$, is one column vector of $\mathbf{V}$, and thus it is orthogonal to the noise space.

When $\boldsymbol{x} \notin \text{Span}(\mathbf{H}^{(k)})$, we rewrite the matrix $[\mathbf{V}, \mathbf{v}^{(k)} \otimes \boldsymbol{x}]$ to its column permutation version as follow

$$\breve{\mathbf{V}} = \left[\mathbf{v}^{(1)} \otimes \mathbf{H}^{(1)}, \cdots, \mathbf{v}^{(k)} \otimes \breve{\mathbf{H}}^{(k)}, \cdots, \mathbf{v}^{(K)} \otimes \mathbf{H}^{(K)} \right], \tag{58}$$

where $\breve{\mathbf{H}}^{(k)} = [\mathbf{H}^{(k)}, \boldsymbol{x}]$. Since $M > L_h$ and $\boldsymbol{x} \notin \text{Span}(\mathbf{H}^{(k)})$, we know $\breve{\mathbf{H}}^{(k)}$ has full column rank. Following the discussion in Appendix A, we know $\breve{\mathbf{V}}$ has full column rank. The implication behind is that $\mathbf{v}^{(k)} \otimes \boldsymbol{x}$ does not belong to the column space of $\mathbf{V}$. Thus, $\mathbf{v}^{(k)} \otimes \boldsymbol{x}$ cannot be orthogonal to $\mathbf{U_n}$, i.e., $(\mathbf{v}^{(k)} \otimes \boldsymbol{x})^H \mathbf{U_n} \neq \mathbf{0}$.

Appendix D
Proof of *Lemma 2*

The Lemma can be proved from the method of contradiction. Denote $\boldsymbol{x}_l$ as the lth column of $\boldsymbol{X}$. First assume that $(\mathbb{B}^{(k)}(\xi) \otimes \boldsymbol{x}_l)^H \mathbf{U_n} = \mathbf{0}$ holds, which implies that $\mathbb{B}^{(k)}(\xi) \otimes \boldsymbol{x}_l$ belongs to the column space of $\mathbf{V}$. Hence, there are $\epsilon_{l'}^{(k')}$, $l' = 1, 2, \cdots, L_h$, $k' = 1, 2, \cdots, K$, not all zero, such that

$$\mathbb{B}^{(k)}(\xi) \otimes \boldsymbol{x}_l = \sum_{k'=1}^{K} \sum_{l'=1}^{L_h} \epsilon_{l'}^{(k')} (\mathbf{v}^{(k')} \otimes \mathbf{h}_{l'}^{(k')}). \tag{59}$$

By extracting the rows corresponding to the mth entry of $\boldsymbol{x}_l$ and $\mathbf{h}_{l'}^{(k')}$, we obtain

$$\boldsymbol{x}_l(m) \mathbb{B}^{(k)}(\xi) = \sum_{k'=1}^{K} \sum_{l'=1}^{L_h} \epsilon_{l'}^{(k')} \mathbf{v}^{(k')} \mathbf{h}_{l'}^{(k')}(m)$$

$$= \sum_{k'=1}^{K} \mathbf{v}^{(k')} \sum_{l'=1}^{L_h} \epsilon_{l'}^{(k')} \mathbf{h}_{l'}^{(k')}(m), \tag{60}$$

$m = 1, 2, \cdots, M$. Note that $\mathbf{h}_{l'}^{(k')}(m) = \mathbf{H}^{(k')}(m, l')$; thus, we can rewrite the right hand side of (60) as $\sum_{k'=1}^{K} \mathbf{v}^{(k')}\big(\mathbf{H}^{(k')}(m,:)\boldsymbol{\epsilon}^{(k')}\big)$ where $\boldsymbol{\epsilon}^{(k')} = [\epsilon_1^{(k')}, \epsilon_2^{(k')}, \cdots, \epsilon_{L_h}^{(k')}]^T$. Then, (60) becomes

$$\boldsymbol{x}_l(m)\mathbb{B}^{(k)}(\xi) \tag{61}$$
$$= \bar{\mathbf{v}} \cdot \text{blkdiag}\Big(\mathbf{H}^{(1)}(m,:), \mathbf{H}^{(2)}(m,:), \cdots, \mathbf{H}^{(K)}(m,:)\Big)\boldsymbol{\epsilon},$$

$m = 1, 2, \cdots, M$, where

$$\bar{\mathbf{v}} = [\mathbf{v}^{(1)}, \mathbf{v}^{(2)}, \cdots, \mathbf{v}^{(K)}],$$
$$\boldsymbol{\epsilon} = [(\boldsymbol{\epsilon}^{(1)})^T, (\boldsymbol{\epsilon}^{(2)})^T, \cdots, (\boldsymbol{\epsilon}^{(K)})^T]^T.$$

Denote $\boldsymbol{\kappa} = \bar{\mathbf{v}}^{-1}\mathbb{B}^{(k)}(\xi)$. It can be verified that $\boldsymbol{\kappa}$ has no zero entries:

Proof: From $\bar{\mathbf{v}}\boldsymbol{\kappa} = \mathbb{B}^{(k)}(\xi)$, we know that if the ith entry of $\boldsymbol{\kappa}$ is zero, $\mathbb{B}^{(k)}(\xi)$ should be represented by the remaining $K-1$ columns of $\bar{\mathbf{v}}$ after removing its ith column. However, $\mathbb{B}^{(k)}(\xi)$ is not linearly related to any $K-1$ columns of $\bar{\mathbf{v}}$ when $\xi \neq \xi^{(k)}$, which indicates no zero entry in $\boldsymbol{\kappa}$. ∎

From (61), we further obtain

$$\boldsymbol{x}_l(m)\boldsymbol{\kappa} \tag{62}$$
$$= \Big[\mathbf{H}^{(1)}(m,:)\boldsymbol{\epsilon}^{(1)}, \mathbf{H}^{(2)}(m,:)\boldsymbol{\epsilon}^{(2)}, \cdots, \mathbf{H}^{(K)}(m,:)\boldsymbol{\epsilon}^{(K)}\Big]^T,$$

$m = 1, 2, \cdots, M$. Then, the k'th entry on both sides of (62) is related by

$$\boldsymbol{x}_l(m)\boldsymbol{\kappa}(k') = \mathbf{H}^{(k')}(m,:)\boldsymbol{\epsilon}^{(k')}, \tag{63}$$

$k' = 1, 2, \cdots, K$. Stacking all $\boldsymbol{x}_l(m)\boldsymbol{\kappa}(k')$, $m = 1, \ldots, M$ into one vector, we obtain the following K equations:

$$\boldsymbol{\kappa}(k') \cdot \boldsymbol{x}_l = \mathbf{H}^{(k')}\boldsymbol{\epsilon}^{(k')}, \tag{64}$$

$k' = 1, 2, \cdots, K$, which implies that $\boldsymbol{x}_l$ could be linearly represented by $\mathbf{H}^{(k')}$, $k' = 1, 2, \cdots, K$. In other words, $\boldsymbol{x}_l$ should be a vector that belongs to the column space of each $\mathbf{H}^{(k')}$.

Afterwards, we assume the equation $(\mathbb{B}^{(k)}(\xi) \otimes \boldsymbol{X})^H \mathbf{U_n} = \mathbf{0}$ holds. Following the above discussion, we know that the column space of $\boldsymbol{X}$ should be equal to the column space of each $\mathbf{H}^{(k')}$, $k' = 1, 2, \cdots, K$. This implies that all K matrices $\mathbf{H}^{(k')}$ should have the same column space, which has zero possibility due to the random property of the wireless channels. From the contradiction, we arrive at $(\mathbb{B}^{(k)}(\xi) \otimes \boldsymbol{X})^H \mathbf{U_n} \neq \mathbf{0}$.

APPENDIX E
PROOF OF THE FULL COLUMN RANK OF $\boldsymbol{\Pi}(\hat{\boldsymbol{\omega}}_{|k,i}, \tilde{\xi}^{(k)})$

We equivalently prove that $\boldsymbol{\Pi}^H(\hat{\boldsymbol{\omega}}_{|k,i}, \tilde{\xi}^{(k)})\boldsymbol{\Pi}(\hat{\boldsymbol{\omega}}_{|k,i}, \tilde{\xi}^{(k)})$ is a positive definite matrix. For any length-M non-zero vector $\boldsymbol{x}$, there holds

$$\boldsymbol{x}^H \boldsymbol{\Pi}^H(\hat{\boldsymbol{\omega}}_{|k,i}, \tilde{\xi}^{(k)})\boldsymbol{\Pi}(\hat{\boldsymbol{\omega}}_{|k,i}, \tilde{\xi}^{(k)})\boldsymbol{x} \tag{65}$$
$$= \big(\mathbf{v}^{(k)}(\tilde{\xi}^{(k)}) \otimes \boldsymbol{x}\big)^H \mathbf{P}_{\mathbf{B}}^{\perp}(\hat{\boldsymbol{\omega}}_{|k,i})\big(\mathbf{v}^{(k)}(\tilde{\xi}^{(k)}) \otimes \boldsymbol{x}\big).$$

Note that the channel matrices in the parameter sets $\hat{\boldsymbol{\omega}}_{|k,i}$ have full column rank, which always holds in the iterative process. Similar to the discussion in Appendix A, we observe that for any nonzero $\boldsymbol{x}$, the matrix

$$\Big[\mathbf{B}(\hat{\boldsymbol{\omega}}_{|k,i}), \mathbf{v}^{(k)}(\tilde{\xi}^{(k)}) \otimes \boldsymbol{x}\Big] \tag{66}$$

has full column rank, which means that $\mathbf{v}^{(k)}(\tilde{\xi}^{(k)}) \otimes \boldsymbol{x}$ is linearly independent with the column vectors of $\mathbf{B}(\hat{\boldsymbol{\omega}}_{|k,i})$. As a result, (65) cannot be equal to zero with any nonzero vector $\boldsymbol{x}$. This implies that $\boldsymbol{\Pi}^H(\hat{\boldsymbol{\omega}}_{|k,i}, \tilde{\xi}^{(k)})\boldsymbol{\Pi}(\hat{\boldsymbol{\omega}}_{|k,i}, \tilde{\xi}^{(k)})$ is a positive definite matrix.

REFERENCES

[1] G.L. Stuber, J Barry, S.W. McLaughlin, Y. Li, M.A. Ingram, and T.G. Pratt, "Broadband MIMO-OFDM wireless communications," in *Proc. IEEE,* vol. 92, no. 2, pp. 271–294, Feb. 2004.

[2] S. Sergi, F. Pancaldi and G.M. Vitetta, "Cooperative communication techniques for wireless OFDMA-based Ad-hoc Networks," in *Proc of IEEE Int. Conf. Commun. (ICC)*, Dresden, Germany, June 2009, pp. 1–6.

[3] U. Tureli, H. Liu, and M. D. Zoltowski, "OFDM blind carrier offset estimation: ESPRIT," *IEEE Trans. Commun.*, vol. 48, no. 9, Sep. 2001.

[4] L. Wu, X.-D. Zhang, and P.-S. Li, "A low-complexity blind carrier frequency offset estimator for MIMO-OFDM systems," *IEEE Signal Process. Letters*, vol. 15, pp. 769–772, 2008.

[5] F. Gao and A. Nallanathan, "Blind maximum likelihood CFO estimation for OFDM systems via polynomial rooting," *IEEE Signal Processing Lett.*, vol. 13, no. 2, pp. 73-76, Feb. 2006.

[6] Z. Zhang, C. Tellambura, "The effect of imperfect carrier frequency offset estimation on an OFDMA uplink," *IEEE Trans. Commun.*, vol. 57, no. 4, pp. 1025-1030, Apr. 2009

[7] P. Sun and L. Zhang, "Low complexity pilot aided frequency synchronization for OFDMA uplink transmission," *IEEE Trans. Wireless Commun.,* vol. 8, no. 7, pp. 3758–3769, July 2009.

[8] M. R. Raghavendra, E. Lior, S. Bhashyam, and K. Giridhar, "Parametric channel estimation for pseudo-random tile-allocation in uplink OFDMA," *IEEE Trans. Signal Processing,* vol. 55, no. 11, pp. 5307–5381, Nov. 2007.

[9] Z. Wang, Y. Xin, and G. Mathew, "Carrier-frequency offset estimation for OFDMA uplink with generalized subcarrier-assignment," in *Proc. IEEE Int. Conf. Commun. (ICC),* Beijing, China, May 2008, pp. 3490–3494.

[10] Z. Wang, Y. Xin, and G. Mathew, "Iterative carrier-frequency offset estimation for generalized OFDMA uplink transmission," *IEEE Trans. Wireless Commun.*, vol. 8, no. 3, pp. 1373–1383, Mar. 2009.

[11] S. Kay and S. Saha, "Mean likelihood frequency estimation," *IEEE Trans. Signal Process,* vol. 48, no. 7, pp. 1937-1946, Jul. 2000.

[12] J. Chen, Y.-C. Wu, S.C. Chan, and T.-S. Ng, "Joint maximum-likelihood CFO and channel estimation for OFDMA uplink using importance sampling," *IEEE Trans. Vehicular Technology,* vol. 57, no. 6, pp. 3462–3470, Nov. 2008.

[13] S. Barbarossa, M. Pompili, and G. B. Giannakis, "Channel-independent synchronization of orthogonal frequency division multiple access systems," *IEEE J. Select. Areas Commun.*, vol. 20, no. 2, pp. 474–486, Feb. 2002.

[14] Y. Yao and G. B. Giannakis, "Blind carrier frequency offset estimation in SISO, MIMO, and multiuser OFDM systems," *IEEE Trans. Commun.*, vol. 53, no. 1, pp. 173–183, Jan. 2005.

[15] Z. Cao, U. Tureli, and Y.-D. Yao, "Deterministic multiuser carrier-frequency offset estimation for interleaved OFDMA uplink," *IEEE Trans. Commun.*, vol. 52, no. 9, pp. 1585–1594, Sept. 2004.

[16] R. O. Schmidt, "Multiple emitter location and signal parameter estimation," in *Proc. RADC Spectral Estimation Workshop,* Griffiss AFB, NY, 1979, pp. 243–258.

[17] J. Lee, S. Lee, K.-J. Bang, S. Cha, and D. Hong, "Carrier frequency offset estimation using ESPRIT for interleaved OFDMA uplink systems," *IEEE Trans. Vehicular Technology,* vol. 56, no. 9, pp. 3227–3231, Sept. 2007.

[18] R. Roy, A. Paulraj, and T. Kailath, "ESPRIT-a subspace rotation approach to estimation of parameters of cisoids in noise," *IEEE Trans. on Acoust., Speech, Signal Processing,* vol. ASSP-34, pp. 1340–1342, Feb. 1986.

[19] D. Fan and Z. Cao, "Subspace-based uplink carrier-frequency offset estimation for interleaved-OFDMA," *Acta Electronica Sinica,* vol. 35, no. 4, pp. 629–633, Apr. 2007.

[20] K.-H. Wu, W.-H. Fang, and Y.-T. Chen, "Joint carrier frequency offset and direction of arrival estimation via hierarchical ESPRIT for interleaved OFDMA/SDMA uplink systems," in *Proc. IEEE Vehicular Technology Conf. (VTC Spring),* Taipei, Taiwan, May 2010, pp. 1–5.

[21] Y. Zhu and K.B. Letaief, "CFO estimation and compensation in single carrier interleaved FDMA systems," in *Proc. IEEE Global Telecommun. Conf. (GLOBECOM),* Honolulu, USA, Nov. 2009, pp. 1–5.

[22] H.T. Hsieh and W.R. Wu, "Blind maximum-likelihood carrier-frequency-offset estimation for interleaved OFDMA uplink systems," *IEEE Trans. Vehicular Tech.*, vol. 60, no. 1, pp. 160–173, Jan. 2011.

[23] H. Wang and Q. Yin, "Multiuser carrier frequency offsets estimation for OFDMA uplink with Generalized Carrier Assignment Scheme," *IEEE Trans. Wireless Commun.*, vol. 8, no. 7, pp. 3347–3353, July 2009.

[24] W. Zhang, Q. Yin, and W. Wang, "Carrier frequency offset estimation for OFDMA uplink using multi-antenna," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Cape Town, South Africa, May 2010, pp. 1–5.

[25] I. Ziskind and M. Wax, "Maximum likelihood localization of multiple sources by alternating projection," *IEEE Trans. Acoust., Speech, and Signal Proccessing*, vol. 36, no. 10, pp. 1553–1560, Oct. 1988.

[26] X. Liao, L. Fan, and F. Gao, "Blind channel estimation for OFDM modulated two-way relay network," in *IEEE Wireless Commun. Networking Conf. (WCNC),* Syndey, Austria, Apr. 2010, pp. 1–5.

[27] G.W. Stewart, "On the early history of the singular value decomposition," *SIAM Review,* vol. 35, no. 4, pp. 551–566, 1993.

[28] Y. R. Zheng and C. Xiao, "Simulation models with correct statistical properties for Rayleigh fading channels," *IEEE Trans. Commun.*, vol. 51, no. 6, pp. 920–928, June 2003.

[29] E. Telatar, "Capacity of multi-antenna Gaussian Channels," *European Trans. Telecommun.,* vol. 10, no. 6, pp. 585–595, Dec. 1999.

Qinye Yin received the B.S., M.S., and Ph.D. degrees in Communication and Electronic Systems from Xi'an Jiaotong University, Xi'an, China, in 1982, 1985, and 1989, respectively. Since 1989, he has been on the faculty at Xi'an Jiaotong University, where he is currently a Professor of Information and Communications Engineering Department and Chair of Academy Committee of School of Electronic and Information Engineering. From 1987 to 1989, he was a visiting scholar at the University of Maryland, MD, USA. From June to December 1996, he was a visiting scholar at the University of Taxes, Austin, TX, USA. His research interests focus on the joint time-frequency analysis and synthesis, the theory and applications of wireless sensor networks, multiple antenna MIMO broadband communication systems (including smart antenna systems), parameter estimation, and array signal processing.

Weile Zhang received the B.S. degree in Information Engineering from Xi'an Jiaotong University, Xi'an, China, in 2006. He is currently pursuing his Ph.D. degree in Xi'an Jiaotong University. From October 2010 to October 2011, he was a visiting scholar at the Department of Computer Science, University of California, Santa Barbara, CA, USA. His research interests include broadband wireless communications, array signal processing and localization in wireless networks.

Arumugam Nallanathan (S'97-M'00-SM'05) is the Head of Graduate Studies in the School of Natural and Mathematical Sciences and a Reader in Communications at King's College London, London, U.K. From August 2000 to December 2007, he was an Assistant Professor in the Department of Electrical and Computer Engineering at the National University of Singapore, Singapore. His research interests include smart grid, cognitive radio, and relay networks. He has authored nearly 200 journal and conference papers. He is an Editor for IEEE Transactions on Communications, IEEE Transactions on Vehicular Technology, IEEE Wireless Communications Letters and IEEE Signal Processing Letters. He served as an Editor for IEEE Transactions on Wireless Communications (2006-2011) and as a Guest Editor for the *EURASIP Journal of Wireless Communications and Networking: Special Issue on UWB Communication Systems-Technology and Applications*. He served as the General Track Chair for the Spring 2008 IEEE Vehicular Technology Conference (VTC'2008) and Co-Chair for the 2008 IEEE Global Communications Conference (GLOBECOM'2008) Signal Processing for Communications Symposium, 2009 IEEE International Conference on Communications (ICC'2009) Wireless Communications Symposium, 2011 IEEE Global Communications Conference (GLOBECOM'2011) Signal Processing for Communications Symposium and the 2012 IEEE International Conference on Communications (ICC'2012) Signal Processing for Communications Symposium. He also served as the Technical Program Co-Chair for the 2011 IEEE International Conference on Ultra-Wideband (ICUWB'2011). He is the Vice-Chair for the Signal Processing for Communication Electronics Technical Committee of the IEEE Communications Society. He was a co-recipient of the Best Paper Award presented at the 2007 IEEE International Conference on Ultra-Wideband.

Feifei Gao (S'05-M'09) received the B.Eng. degree from Xi'an Jiaotong University, Xi'an, China in 2002, the M.Sc. degree from McMaster University, Hamilton, ON, Canada, in 2004, and the Ph.D degree from National University of Singapore, Singapore, in 2007.

He was a Research Fellow with the Institute for Infocomm Research, A*STAR, Singapore, in 2008 and an Assistant Professor with the School of Engineering and Science, Jacobs University, Bremen, Germany, from 2009 to 2010. In 2011, he joined the Department of Automation, Tsinghua University, Beijing, China, where he is currently an Associate Professor. He also serves as an Adjunct Professor with the School of Engineering and Science, Jacobs University. He has coauthored more than 90 refereed IEEE journal and conference proceeding papers. His research interests are communication theory, broadband wireless communications, signal processing, multiple-input-multiple-output systems, and array signal processing.

Prof. Gao has served on the Editorial Board of IEEE Wireless Communications Letters, and as a Technical Program Committee member for the IEEE International Conference on Communications, the IEEE Global Communications Conference, the IEEE Vehicular Technology Conference, and the IEEE Personal, Indoor, and Mobile Radio Communications Conference.