Joint CFO and Channel Estimation for Multiuser MIMO-OFDM Systems With Optimal Training Sequences

Jianwu Chen, Yik-Chung Wu, Shaodan Ma, and Tung-Sang Ng, *Fellow, IEEE*

***Abstract*—This paper addresses the problem of joint carrier frequency offset (CFO) and channel estimation in multiuser multiple-input–multiple-output (MIMO) orthogonal frequency-division multiplexing (OFDM) systems. To choose the optimal training sequences with the goal of providing the smallest estimation mean square error (MSE), the asymptotic Cramér–Rao bounds (asCRBs) are derived. The optimal training sequences are designed that minimize the asCRBs for both CFO and channel estimation under the constraint that the asCRBs being channel independent. A joint CFO and channel estimator is derived based on the maximum likelihood (ML) criterion. A computationally efficient method using importance sampling technique is proposed to solve the highly demanding multidimensional exhaustive search required by the ML multi-CFO estimation. Simulation results illustrate the merits of the proposed training sequences and also verify the effectiveness of the proposed estimation scheme.**

***Index Terms*—Carrier frequency offset, channel estimation, Cramér–Rao bound (CRB), importance sampling, maximum likelihood, multiple-input–multiple-output (MIMO), orthogonal frequency-division multiplexing (OFDM).**

I. Introduction

MULTIPLE-INPUT–MULTIPLE-OUTPUT (MIMO) orthogonal frequency-division multiplexing (OFDM) system has attracted much attention recently and it is widely regarded as a potential candidate for fourth-generation (4G) broadband wireless communication networks due to its ability to mitigate intersymbol interference (ISI) and to enhance system capacity [1]. With spatial multiplexing (e.g., Bell Labs layered space–time (BLAST) architecture), the system throughput can be improved by transmitting independent data streams on different antennas simultaneously [2]. With OFDM, the frequency selective fading channel is transformed into a set of frequency flat fading channels, making equalization much simpler than single carrier systems. Compared with other OFDM-based techniques, such as MIMO-OFDMA, MIMO-OFDM systems promise a much higher spectral efficiency due to the use of spatial multiplexing. While researches on detection and decoding of MIMO-OFDM systems [3]–[5] have been recently reported, efficient synchronization and channel estimation of such systems are relatively unexplored. The purpose of this paper is to fill this gap.

In general, to fully exploit the potentials of MIMO-OFDM systems, three issues should be considered: time synchronization, frequency synchronization and channel estimation. Similar to other multicarrier-based techniques, MIMO-OFDM system is highly sensitive to carrier frequency offsets (CFOs) caused by oscillator mismatches and/or Doppler shifts. For coherent data detection, the knowledge of channel impulse response (CIR) is indispensable at receiver. Compared to frequency synchronization and channel estimation, time synchronization is relatively less critical since OFDM systems have some tolerances to timing errors due to the cyclic prefix (CP) insertion [6]. In this paper, the focus is therefore on frequency synchronization and channel estimation.

In many OFDM systems, CFO and channel estimation are tackled separately with the aid of different training sequences [7]–[10]. In these well established schemes, channel estimation is performed after frequency synchronization procedure and assumes frequency synchronization is perfectly achieved [9], [10]. Unfortunately, in practice, such an assumption is rarely valid in the presence of noise [7], [8]. The residual CFO due to frequency synchronization error will degrade the performance of channel estimation significantly [11]. A reasonable solution to this problem is to estimate CFO and channel simultaneously based on common training sequences, which has been highlighted in [12] and [13]. However, in these schemes [12], [13], training sequences are proposed without proof of optimality.

In the literature, most of the existing works addressing the problem of optimal training sequence design for joint CFO and channel estimation focus on single carrier systems (e.g., [14], [15] and [16]). In [14], the design of optimal training sequence is first investigated for SISO systems using worst case asymptotic CRBs (asCRB). It was found that a white sequence is optimal in the sense of minimizing the worst case asCRBs. This approach was extended to single user MIMO case in [16]. In the considered system of [16], only one CFO between the transmitter and receiver is assumed. To remove the channel effects on CFO estimation, the training sequences are chosen such that the performance of CFO estimation is independent to the channel realizations. In [15], a more general MIMO system with different CFOs for distinct transmit antennas is considered for the flat-fading

Manuscript received April 17, 2007; revised April 9, 2008. This work was supported by a Grant CERG HKU 7160/06E. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Anthanasios P. Liavas.

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

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TSP.2008.925896

channel only. More recently, in [17], the optimal training sequences for MIMO-OFDMA have been derived in the sense of minimizing the asCRBs for CFO and channel separately.

In this paper, the problem of joint CFO and channel estimation in multiuser MIMO-OFDM systems is investigated, where all users can utilize all available subcarriers simultaneously and each user is assumed to have distinct CFO with respect to the receiver. The channels in this paper are frequency selective fading channels, which are different from the case in [15]. The contributions of this paper are as follows. First, the conditions for the optimal training sequences are derived based on asymptotic analyses. The proposed training sequences are optimal in the sense of simultaneously minimizing both the asCRBs for CFO and channel estimation under the constraint that the asCRBs being channel independent. Compared with [17], the training sequences in this paper are designed to jointly minimize both asCRBs for frequency offset and channel, while the training sequences in [17] are designed by minimizing the asCRBs for CFO and channel separately. Furthermore, based on the asymptotic optimal condition, we present explicitly how to construct the optimal training sequences with finite length. Second, a joint maximum likelihood (ML) estimator is derived. It is found that the ML multi-CFO estimation is very challenging due to the need of multidimensional exhaustive search. To overcome this problem, the importance sampling technique is exploited. Based on the derived optimal training sequences, we designed an efficient and simple importance function considering both the estimation performance and ease of sample generation. Third, with the proposed importance function, we designed a novel modified ratio-of-uniform method in sample generation, which remains computationally efficient for any CFO estimation range.

The rest of this paper is organized as follows. In Section II, the considered MIMO-OFDM system model is presented. The optimal training sequence design for joint CFO and channel estimation is investigated in Section III through CRB analyses. In Section IV, the ML estimator for the joint CFO and channel estimation is derived and an efficient algorithm based on importance sampling is proposed to reduce the complexity. Section V presents simulation results to validate the proposed training design and estimation scheme. Concluding remarks are given in Section VI.

Notations: $(\cdot)^{-1}$, $(\cdot)^T$ and $(\cdot)^H$ denote the inverse, the transpose, and the conjugate transpose operations, respectively, and $\otimes$ denotes the Kronecker product. The $k \times k$ identity and zero matrix are denoted by $\mathbf{I}_{k\times k}$ and $\mathbf{0}_{k\times k}$, respectively, and $\|\mathbf{x}\|$ represents the norm operation for a vector $\mathbf{x}$. Throughout the paper, MATLAB notations for matrix and vector are used.

II. Signal Model

In the considered MIMO-OFDM system, K users transmit different data streams simultaneously using the same set of subcarriers to the base station (BS), which is equipped with M_r antennas and is responsible for decoding the symbols for each user. For each user, only one transmit antenna is assumed. It is reasonable to assume that the receive antennas at BS share the same oscillator while all users are driven by different oscillators. Thus, the data streams from different users will experience different CFOs. Before initiating the transmission, the timing for each user is acquired by using the downlink synchronization channel from BS. Consequently, the transmissions from all users can be regarded to be quasi-synchronous [6].

The data stream from user i is first segmented into blocks of length N (denoted as $\mathbf{d}_i = [d_i(0), \ldots, d_i(N-1)]^T$) and then modulated onto different subcarriers by left multiplying an N-point inverse FFT matrix $\mathbf{F}^H$, where $\mathbf{F}$ is the FFT matrix with $\mathbf{F}(k,l) = (1/\sqrt{N})\exp(-(j2\pi kl/N))$. After inserting a CP of length L_{cp} into each block of the time domain signal (denoted as $\mathbf{F}^H\mathbf{d}_i$), the augmented block is serially transmitted through the multipath channel. Let the channel impulse response (including all transmit/receive filtering effects) between the ith user and the jth receive antenna be denoted as $\boldsymbol{\xi}_{i,j} = [\xi_{i,j}(0), \ldots, \xi_{i,j}(L_{i,j}-1)]^T$ where $L_{i,j}$ is the channel length. Denoting the timing offset on this ray caused by propagation delay as $\theta_{i,j}$, the compound channel response can be written as $\mathbf{h}_{i,j} \triangleq [\mathbf{0}^T_{\theta_{i,j}\times 1}\ \boldsymbol{\xi}^T_{i,j}\ \mathbf{0}^T_{(L-\theta_{i,j}-L_{i,j})\times 1}]^T$, where L is the upper bound on the compound channel length. Since for coherent data detection, only the estimates of the compound channels $\mathbf{h}_{i,j}$ are required, the estimation of timing offsets becomes dispensable. Through combining the timing offsets $\theta_{i,j}$ and the exact channel $\boldsymbol{\xi}_{i,j}$, the system can be regarded as a perfectly synchronous system with the compound channels $\mathbf{h}_{i,j}$. Thus, without loss of generality, in the following we assume $\theta_{i,j} = 0, \forall i,j$ for simplicity. For user i, the normalized CFO (between the oscillator at user i and that of the base station) is denoted as ε_i. At the BS, after removal of CP, the signal from user i to the jth receive antenna is given by

$$\mathbf{x}_{ij} \triangleq [x_{ij}(0), \ldots, x_{ij}(N-1)]^T = \boldsymbol{\Gamma}(\omega_i)\mathbf{A}_i\mathbf{h}_{i,j} \tag{1}$$

where

$$\boldsymbol{\Gamma}(\omega_i) \triangleq \text{diag}\left(1, \ldots, \exp\left(j(N-1)\omega_i\right)\right) \tag{2}$$
$$\mathbf{A}_i = \mathbf{F}^H\mathbf{D}_i\mathbf{F}_L \tag{3}$$
$$\mathbf{D}_i \triangleq \text{diag}(\mathbf{d}_i) \tag{4}$$
$$\mathbf{F}_L = \mathbf{F}(:, 1:L) \tag{5}$$
$$\omega_i \triangleq 2\pi\varepsilon_i/N. \tag{6}$$

Since the received signal at the jth receive antenna of the base station is the sum of signals from all users and noise, the received signal at this antenna is given by

$$\mathbf{x}_j = \sum_{i=1}^{K}\mathbf{x}_{ij} + \mathbf{n}_j = \sum_{i=1}^{K}\boldsymbol{\Gamma}(\omega_i)\mathbf{A}_i\mathbf{h}_{i,j} + \mathbf{n}_j. \tag{7}$$

In above, the vector $\mathbf{n}_j$ is complex white Gaussian noise with zero mean and covariance matrix $\sigma^2\mathbf{I}_{N\times N}$.

Denoting $\mathbf{x} = [\mathbf{x}_1^T\ \mathbf{x}_2^T\ \ldots\ \mathbf{x}_{M_r}^T]^T$, $\boldsymbol{\omega} = [\omega_1\ \omega_2\ \ldots\ \omega_K]$, $\mathbf{n} = [\mathbf{n}_1^T\ \mathbf{n}_2^T\ \ldots\ \mathbf{n}_{M_r}^T]^T$, $\mathbf{h} = [\mathbf{h}_1^T\ \mathbf{h}_2^T\ \ldots\ \mathbf{h}_{M_r}^T]^T$ with $\mathbf{h}_j = [\mathbf{h}_{1,j}^T\ \mathbf{h}_{2,j}^T\ \ldots\ \mathbf{h}_{K,j}^T]^T$, the signal model from (7) can be rewritten as

$$\mathbf{x} = \mathbf{Q}(\boldsymbol{\omega})\mathbf{h} + \mathbf{n} \tag{8}$$

where

$$\mathbf{Q}(\boldsymbol{\omega}) = \mathbf{I}_{M_r\times M_r} \otimes \left[\boldsymbol{\Gamma}(\omega_1)\mathbf{A}_1\ \ldots\ \boldsymbol{\Gamma}(\omega_K)\mathbf{A}_K\right]. \tag{9}$$

Remark 1: Although in this paper, only the case where each user equips with one antenna is considered, the above signal model can be easily generalized to the case where each user equips with more than one antenna. The only extra constraint required is that the unknown CFOs for different antennas of the same user should be equal.

Remark 2: Although in the above signal model, there is an implicit assumption that the upper bound of the sum of the channel delay spread and propagation delay for each user (L) is less than the length of CP (L_{cp}), the considered system model is practical. The reasons are that the timing offsets due to different propagation delays are limited to several samples only, and in practical OFDM systems the CP is always longer than the exact channel order. It should also be noted that with fixed L_{cp}, this assumption limits the distance between the base station and mobile users (i.e., the cell radius in a cellular system). However, this limitation can be eliminated by using a longer CP, which is a system design parameter.

III. Training Sequence Design

The optimal training sequence design in this paper is based on CRB analyses due to the fact that the ML estimator can asymptotically approach this bound ([14]–[16]). Since the additive noise $\mathbf{n}$ in the signal model (8) is white and circular, the received signal $\mathbf{x}$ is complex circular with mean $\boldsymbol{\mu} \triangleq \mathbf{Q}(\boldsymbol{\omega})\mathbf{h}$ and covariance matrix $\sigma^2\mathbf{I}_{M_rN\times M_rN}$. The Fisher information matrix for $\boldsymbol{\eta} \triangleq [\boldsymbol{\omega}, \Re\{\mathbf{h}\}, \Im\{\mathbf{h}\}]$ is given by

$$\boldsymbol{\Phi} = \frac{2}{\sigma^2}\Re\left[\frac{\partial\boldsymbol{\mu}^H}{\partial\boldsymbol{\eta}}\frac{\partial\boldsymbol{\mu}}{\partial\boldsymbol{\eta}^T}\right] = \frac{2}{\sigma^2}\begin{bmatrix} \Re\{\mathbf{Z}^H\mathbf{Z}\} & \Im\{\mathbf{Z}^H\mathbf{Q}\} & \Re\{\mathbf{Z}^H\mathbf{Q}\} \\ -\Im\{\mathbf{Q}^H\mathbf{Z}\} & \Re\{\mathbf{Q}^H\mathbf{Q}\} & -\Im\{\mathbf{Q}^H\mathbf{Q}\} \\ \Re\{\mathbf{Q}^H\mathbf{Z}\} & \Im\{\mathbf{Q}^H\mathbf{Q}\} & \Re\{\mathbf{Q}^H\mathbf{Q}\} \end{bmatrix} \tag{10}$$

where $\mathbf{Q}$ denotes $\mathbf{Q}(\boldsymbol{\omega})$ for expression simplicity, and

$$\mathbf{Z} = \begin{bmatrix} \mathbf{Z}_{11} & \mathbf{Z}_{21} & \dots & \mathbf{Z}_{K1} \\ \mathbf{Z}_{12} & \mathbf{Z}_{22} & \dots & \mathbf{Z}_{K2} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{Z}_{1M_R} & \mathbf{Z}_{2M_R} & \dots & \mathbf{Z}_{KM_R} \end{bmatrix} \tag{11}$$

$$\mathbf{Z}_{ij} = \mathbf{M}\boldsymbol{\Gamma}(\omega_i)\mathbf{A}_i\mathbf{h}_{i,j} \tag{12}$$

$$\mathbf{M} = \mathrm{diag}(0, 1, \dots, N-1). \tag{13}$$

The CRB matrix can be obtained by inverting the above Fisher information matrix $\boldsymbol{\Phi}$. Through some manipulations, the result is given by (14), shown at the bottom of the page, where

$$\boldsymbol{\beta} = (\mathbf{Q}^H\mathbf{Q})^{-1}\mathbf{Q}^H\mathbf{Z} \tag{15}$$

$$\boldsymbol{\gamma}^{-1} = \left(\Re\left\{\mathbf{Z}^H\boldsymbol{\Pi}_Q^{\perp}\mathbf{Z}\right\}\right)^{-1} \tag{16}$$

$$\boldsymbol{\Pi}_Q^{\perp} = \mathbf{I}_{M_rN\times M_rN} - \mathbf{Q}(\boldsymbol{\omega})\left(\mathbf{Q}^H(\boldsymbol{\omega})\mathbf{Q}(\boldsymbol{\omega})\right)^{-1}\mathbf{Q}^H(\boldsymbol{\omega}). \tag{17}$$

Notice that when there is only one user, $\boldsymbol{\gamma}^{-1}$ will be a scalar and (14) reduces to (9) of [14]. Using similar mathematical manipulations as those in [14, Appendix I], the CRB for $\boldsymbol{\omega}$ and $\mathbf{h}$ are obtained as

$$\mathrm{CRB}(\boldsymbol{\omega}) = \frac{\sigma^2}{2}\left(\Re\left\{\mathbf{Z}^H\boldsymbol{\Pi}_Q^{\perp}\mathbf{Z}\right\}\right)^{-1} \tag{18}$$

$$\begin{aligned}\mathrm{CRB}(\mathbf{h}) = \frac{\sigma^2}{2}\cdot\Big(2\left(\mathbf{Q}^H(\boldsymbol{\omega})\mathbf{Q}(\boldsymbol{\omega})\right)^{-1} + \left(\mathbf{Q}^H(\boldsymbol{\omega})\mathbf{Q}(\boldsymbol{\omega})\right)^{-1} \\ \cdot\mathbf{Q}^H(\boldsymbol{\omega})\mathbf{Z}\left(\Re\left\{\mathbf{Z}^H\boldsymbol{\Pi}_Q^{\perp}\mathbf{Z}\right\}\right)^{-1} \\ \times\mathbf{Z}^H\mathbf{Q}(\boldsymbol{\omega})\left(\mathbf{Q}^H(\boldsymbol{\omega})\mathbf{Q}(\boldsymbol{\omega})\right)^{-H}\Big).\end{aligned} \tag{19}$$

In the CFO-free case, the joint CFO and channel estimation reduces to a pure channel estimation problem which has been extensively studied in [16], [18] and [19]. For this case, the CRB for the channel estimation in (19) is simplified to

$$\mathrm{CRB}(\mathbf{h})_{CFO-free} = \sigma^2\mathbf{I}_{M_r\times M_r}\otimes(\mathbf{A}^H\mathbf{A})^{-1} \tag{20}$$

where

$$\mathbf{A} \triangleq [\mathbf{A}_1 \quad \mathbf{A}_2 \quad \dots \quad \mathbf{A}_K] \tag{21}$$

is assumed to be of full column rank to guarantee $\mathbf{A}^H\mathbf{A}$ is invertible. It is known that the condition for optimal training in the sense of minimizing $\mathrm{Tr}\{\mathrm{CRB}(\mathbf{h})_{CFO-free}\}$ is [18]

$$\mathbf{A}^H\mathbf{A} \propto \mathbf{I}_{KL\times KL}. \tag{22}$$

With this condition, two optimal training sequences over one OFDM symbol are proposed in [18] using the basic properties of the discrete Fourier transform: the first type has code-division multiplexing (CDM) pilot allocation in frequency domain, called CDM(F) pilot; the other type has equally spaced disjoint pilot tones in the frequency domain, called FDM pilot.

$$\mathrm{CRB} = \frac{\sigma^2}{2}\cdot\begin{bmatrix} \boldsymbol{\gamma}^{-1} & \boldsymbol{\gamma}^{-1}\Im\{\boldsymbol{\beta}^T\} & -\boldsymbol{\gamma}^{-1}\Re\{\boldsymbol{\beta}^T\} \\ \Im\{\boldsymbol{\beta}\}\boldsymbol{\gamma}^{-1} & \Re\left\{(\mathbf{Q}^H\mathbf{Q})^{-1}\right\} + \Im\{\boldsymbol{\beta}\}\boldsymbol{\gamma}^{-1}\Im\{\boldsymbol{\beta}^T\} & -\Im\left\{(\mathbf{Q}^H\mathbf{Q})^{-1}\right\} - \Im\{\boldsymbol{\beta}\}\boldsymbol{\gamma}^{-1}\Re\{\boldsymbol{\beta}^T\} \\ -\Re\{\boldsymbol{\beta}\}\boldsymbol{\gamma}^{-1} & \Im\left\{(\mathbf{Q}^H\mathbf{Q})^{-1}\right\} - \Re\{\boldsymbol{\beta}\}\boldsymbol{\gamma}^{-1}\Im\{\boldsymbol{\beta}^T\} & \Re\left\{(\mathbf{Q}^H\mathbf{Q})^{-1}\right\} + \Re\{\boldsymbol{\beta}\}\boldsymbol{\gamma}^{-1}\Re\{\boldsymbol{\beta}^T\} \end{bmatrix} \tag{14}$$

In the presence of CFO, two cases are possible: 1) if all the users share the same CFO (only one CFO needs to be estimated), a special kind of FDM sequence is proposed in [16] such that the CRB for the CFO estimation is channel-independent; 2) if multiple CFOs are present, the conditions for the optimal training sequence become very complicated because the exact CRBs have a complex dependence on the channel gains and the frequency offsets. To proceed, we turn to the asymptotic CRBs (when the number of subcarriers is infinite) instead. In the asymptotic analysis, it is reasonable to assume that the time domain training signals are zero-mean random sequences. Then, exploiting the results of [15, eq. (51)], we have the following equations:

$$\lim_{N\to\infty}\frac{\mathbf{Q}^H\mathbf{Q}}{N}=\mathbf{R} \tag{23a}$$

$$\lim_{N\to\infty}\frac{\mathbf{Q}^H\left(\mathbf{I}_{M_r\times M_r}\otimes\mathbf{M}\right)\mathbf{Q}}{N^2}=\frac{1}{2}\mathbf{R} \tag{23b}$$

$$\lim_{N\to\infty}\frac{\mathbf{Q}^H\left(\mathbf{I}_{M_r\times M_r}\otimes\mathbf{M}\right)^2\mathbf{Q}}{N^3}=\frac{1}{3}\mathbf{R} \tag{23c}$$

where

$$\mathbf{R}=\mathbf{I}_{M_r\times M_r}\otimes\begin{bmatrix}\mathbf{R}_{1,1} & \mathbf{R}_{1,2} & \dots & \mathbf{R}_{1,K}\\ \mathbf{R}_{2,1} & \mathbf{R}_{2,2} & \dots & \mathbf{R}_{2,K}\\ \vdots & \vdots & \ddots & \vdots\\ \mathbf{R}_{K,1} & \mathbf{R}_{K,2} & \dots & \mathbf{R}_{K,K}\end{bmatrix} \tag{24}$$

$$\mathbf{R}_{i,j}=\frac{\mathbf{A}_i^H\mathbf{A}_j}{N}\delta(\omega_j-\omega_i). \tag{25}$$

It is noted that the matrix $\mathbf{R}$ is a positive definite matrix since $\mathbf{A}$ has full column rank. Similar to [14], the normalized CRB ($\overline{\mathrm{CRB}}$) is denoted as $\mathbf{K}_N\mathrm{CRB}\mathbf{K}_N^T$ with the block-diagonal matrix $\mathbf{K}_N = \mathrm{diag}(N^{3/2}\mathbf{I}_{K\times K}, N^{1/2}\mathbf{I}_{M_rKL\times M_rKL}, N^{1/2}\mathbf{I}_{M_rKL\times M_rKL})$ and the corresponding FIM is denoted as $\overline{\mathbf{\Phi}}$. Using the results of (23), it is readily to obtain $\lim_{N\to\infty}\overline{\mathbf{\Phi}}$. Since $\lim_{N\to\infty}\overline{\mathrm{CRB}}=[\lim_{N\to\infty}\overline{\mathbf{\Phi}}]^{-1}$, using a similar approach in above CRB derivation, the asymptotic CRB for $\boldsymbol{\omega}$ and $\mathbf{h}$ are obtained as [20]

$$\mathrm{asCRB}(\boldsymbol{\omega})=\frac{6\sigma^2}{N^3}\left(\Re\{\mathbf{H}^H\mathbf{R}\mathbf{H}\}\right)^{-1} \tag{26}$$

$$\mathrm{asCRB}(\mathbf{h})=\frac{\sigma^2}{N}\left(\mathbf{R}^{-1}+\frac{3}{2}\mathbf{H}\left(\Re\{\mathbf{H}^H\mathbf{R}\mathbf{H}\}\right)^{-1}\mathbf{H}^H\right) \tag{27}$$

where

$$\mathbf{H}=\left[\mathbf{H}_1^T\ \mathbf{H}_2^T\ \dots\ \mathbf{H}_{M_r}^T\right]^T \tag{28}$$

$$\mathbf{H}_j=\mathrm{diag}\left([\mathbf{h}_{1j}\ \mathbf{h}_{2j}\ \dots\ \mathbf{h}_{Kj}]\right). \tag{29}$$

Based on the fact that $\Re\{\mathbf{H}^H\mathbf{R}\mathbf{H}\}$ is a real positive-definite matrix, it is readily to prove that [20]

$$\left(\left(\Re\{\mathbf{H}^H\mathbf{R}\mathbf{H}\}\right)^{-1}\right)_{k,k}\geq\frac{1}{\left(\Re\{\mathbf{H}^H\mathbf{R}\mathbf{H}\}\right)_{k,k}}=\frac{1}{\sum_{i=1}^{M_r}\mathbf{h}_{ki}^H\mathbf{R}_{k,k}\mathbf{h}_{ki}}. \tag{30}$$

Furthermore, since $\mathbf{R}$ is a positive-definite matrix, based on [21, Theorem 7.7.8], we have

$$\mathrm{Tr}(\mathbf{R}^{-1})\geq M_r\sum_{k=1}^{K}\mathrm{Tr}\left(\mathbf{R}_{kk}^{-1}\right). \tag{31}$$

In (30) and (31), the equalities hold if and only if $\mathbf{R}_{i,j}=\mathbf{0}_{L\times L}$, $i,j=1,\dots,K; i\neq j$. From (26), (27), (30), and (31), we can obtain the following two inequalities:

$$\mathrm{Tr}\left(\mathrm{asCRB}(\boldsymbol{\omega})\right)\geq\frac{6\sigma^2}{N^3}\sum_{k=1}^{K}\frac{1}{\sum_{i=1}^{M_r}\mathbf{h}_{ki}^H\mathbf{R}_{k,k}\mathbf{h}_{ki}} \tag{32}$$

$$\mathrm{Tr}\left(\mathrm{asCRB}(\mathbf{h})\right)\geq\frac{\sigma^2}{N}\left(M_r\sum_{k=1}^{K}\mathrm{Tr}\left(\mathbf{R}_{kk}^{-1}\right)+\frac{3}{2}\sum_{k=1}^{K}\frac{\sum_{i=1}^{M_r}\mathbf{h}_{ki}^H\mathbf{h}_{ki}}{\sum_{i=1}^{M_r}\mathbf{h}_{ki}^H\mathbf{R}_{k,k}\mathbf{h}_{ki}}\right) \tag{33}$$

with the equalities hold if and only if

$$\mathbf{R}_{i,j}=\mathbf{0}_{L\times L},\quad i,j=1,\dots,K;\quad i\neq j. \tag{34}$$

The above two inequalities (32) and (33) mean that for any non-block diagonal matrix $\mathbf{R}$, its block diagonal version defined as $\bar{\mathbf{R}}=\mathbf{I}_{M_r\times M_r}\otimes\mathrm{diag}\{\mathbf{R}_{11},\dots,\mathbf{R}_{KK}\}$ always has smaller $\mathrm{Tr}(\mathrm{asCRB}(\boldsymbol{\omega}))$ and $\mathrm{Tr}(\mathrm{asCRB}(\mathbf{h}))$. That implies the matrix $\mathbf{R}$ should be designed as a block diagonal matrix, which means the training sequences from different users should be uncorrelated. Based on this condition, the direct approach to proceed is to choose the sequences (contained in $\mathbf{R}_{k,k}$) that minimize the right-hand side (RHS) terms in (32) and (33). However, the solutions to this problem are found to be channel dependent. This renders the training design impractical. Here, inspired by the idea in [16], we choose training sequences that make the RHS terms in (32) and (33) channel independent. In order to fulfill this objective, the training sequences have to satisfy the following condition:

$$\mathbf{R}_{k,k}=P_k\mathbf{I}_{L\times L},\quad k=1,\dots,K \tag{35}$$

where P_k is the transmission power of the user k. Under this condition (together with $\mathbf{R}$ is block diagonal), we have

$$\mathrm{Tr}\left(\mathrm{asCRB}(\boldsymbol{\omega})\right)=\frac{6\sigma^2}{N^3}\sum_{k=1}^{K}\frac{1}{P_k\sum_{i=1}^{M_r}\|\mathbf{h}_{ki}\|^2} \tag{36}$$

$$\mathrm{Tr}(\mathrm{asCRB}(\mathbf{h}))=\frac{\sigma^2}{N}\left(\sum_{k=1}^{K}\frac{M_rL}{P_k}+\frac{3}{2}\sum_{k=1}^{K}\frac{1}{P_k}\right). \tag{37}$$

From above, it is clear that given any channel realization, all users should transmit the training sequences using the maximum power permitted in order to obtain the best performance [i.e., minimize (36) and (37)]. Without loss of generality, all training sequences are assumed to be transmitted at the same power, so the condition for the optimal training in the sense of minimizing

$\mathrm{Tr}\{\mathrm{asCRB}(\boldsymbol{\omega})\}$ and $\mathrm{Tr}\{\mathrm{asCRB}(\mathbf{h})\}$ under the constraint that the asCRBs are channel independent is given by

$$\mathbf{R} \propto \mathbf{I}_{M_r KL \times M_r KL}. \tag{38}$$

It is observed that the conditions for (38) are equivalent to $\mathbf{A}_i^H \mathbf{A}_i \propto \mathbf{I}_{L\times L}$ plus one of the following conditions: 1) $\mathbf{A}_i^H \mathbf{A}_j = \mathbf{0}_{L\times L}, \forall i \neq j$; or 2) $\omega_i \neq \omega_j \ \forall i \neq j$. Since the CFOs of user i and j may be equal (i.e., $\omega_i = \omega_j$), we need $\mathbf{A}_i^H \mathbf{A}_j = \mathbf{0}_{L\times L}, \forall i \neq j$ in order to make the design valid in all possible situations. Together with $\mathbf{A}_i^H \mathbf{A}_i \propto \mathbf{I}_{L\times L}$, we have

$$\mathbf{A}^H \mathbf{A} \propto \mathbf{I}_{KL\times KL}. \tag{39}$$

From (22) and (39), it is clear that for the considered MIMO-OFDM system, the optimal training sequences proposed in [18] (i.e., CDM(F) and FDM sequences) for channel estimation in the absence of CFO are also asymptotically optimal for joint CFO and channel estimation. However, for a practical system with finite subcarriers, it is still unclear how the CDM(F) and FDM training sequences behave.

Since the sequences in [18] satisfy (39) and, thus, share the common asCRBs, the optimal sequence can be regarded as the one whose finite sample CRBs (18) and (19) are closest to the asCRBs (26) and (27). It is pointed out in [14] that colored training sequences with high correlation coefficients could have their finite sample CRBs depart from the asCRBs, so it is sensible to compare the sequences with respect to their correlation properties. For the FDM sequence, the corresponding time domain signal is repetitive due to the equally paced pilot tones in the frequency domain. Thus if the number of nonzero tones in the FDM training is small, the time domain training signal will be highly correlated because of the small repetition length. Compared with the FDM sequence, the CDM(F) sequence has a long correlation length, so it is expected to have a better performance than the FDM like sequence. In the next section, this point will be verified by simulations. Thus, we propose the CDM(F) pilots [18] for the joint CFO and channel estimation

$$d_k(n) = \exp(j\zeta_n)\exp(j\phi_k)\exp\left(-j2\pi(k-1)n/K\right), \quad k = 1,\ldots,K;\ n = 0,\ldots,N-1 \tag{40}$$

where ζ_n and ϕ_k are random variables in $[0, 2\pi]$ with respect to n and k, respectively.

Remark 3: Actually, the training design (35) can also be shown to be minmax optimal using the similar approach in [14]. In other words, the training sequences (35) minimize the RHS of (32) and (33) corresponding to the worst-case channel.

Remark 4: From (39), the optimal training design for joint CFO and channel estimation is the one that can asymptotically decouple the parameter estimations for different users. That means for systems with infinite subcarriers, the best performance of estimation is achieved when the CFO and channels for each user can be estimated separately. Another scheme for separating the estimations for different users is to use time division multiplexing (TDM) sequences similar to the work in [22]. Though exploiting the TDM sequences leads to a scheme with low complexity, it will be shown in the simulation section that with the same training overhead, this TDM training scheme suffers a significant performance loss compared to the proposed training scheme.

Remark 5: It might seem that both [17] and this work design the optimal training sequences based on time domain properties, therefore, from this perspective, there is no difference for MIMO-OFDMA or MIMO-OFDM/SDMA in the pilot design. However, the assumptions and scopes of the results in [17] and this paper are different. The differences of the two works are summarized as follows.

- In this paper, the optimal training sequences are derived by simultaneously minimizing both asCRBs for CFO $(\mathrm{Tr}\{\mathrm{asCRB}(\boldsymbol{\omega})\})$ and channel $(\mathrm{Tr}\{\mathrm{asCRB}(\mathbf{h})\})$ under the constraint that the asCRBs being channel independent. On the other hand, in [17], the optimal sequences are designed in the sense of minimizing the asCRB for frequency and channel estimation separately. Besides, the results in [17] require both the subcarrier number N and the channel length L to be large enough. On the contrary, our proof is valid for any L.
- In practical construction of the training sequence with finite length, we further present the explicit design.

IV. Joint CFO and Channel Estimation

A. ML Estimation

Based on the signal model in (8), the ML estimates of parameters $\{\mathbf{h}, \boldsymbol{\omega}\}$ is given by maximizing

$$\psi(\mathbf{x};\tilde{\mathbf{h}},\tilde{\boldsymbol{\omega}}) = \frac{1}{(\pi\sigma^2)^N}\cdot\exp\left\{-\frac{1}{\sigma^2}\left[\mathbf{x}-\mathbf{Q}(\tilde{\boldsymbol{\omega}})\tilde{\mathbf{h}}\right]^H\left[\mathbf{x}-\mathbf{Q}(\tilde{\boldsymbol{\omega}})\tilde{\mathbf{h}}\right]\right\} \tag{41}$$

or equivalently minimizing

$$\Lambda(\mathbf{x};\tilde{\mathbf{h}},\tilde{\boldsymbol{\omega}}) = \left[\mathbf{x}-\mathbf{Q}(\tilde{\boldsymbol{\omega}})\tilde{\mathbf{h}}\right]^H\left[\mathbf{x}-\mathbf{Q}(\tilde{\boldsymbol{\omega}})\tilde{\mathbf{h}}\right] \tag{42}$$

where $\tilde{\mathbf{h}}$ and $\tilde{\boldsymbol{\omega}}$ are trial values of $\mathbf{h}$ and $\boldsymbol{\omega}$, respectively. Due to the linear dependence of parameter $\mathbf{h}$ in (8), the ML estimate for the channel vector $\mathbf{h}$ (when $\tilde{\boldsymbol{\omega}}$ are fixed) is given by

$$\hat{\mathbf{h}} = \left(\mathbf{Q}^H(\tilde{\boldsymbol{\omega}})\mathbf{Q}(\tilde{\boldsymbol{\omega}})\right)^{-1}\mathbf{Q}^H(\tilde{\boldsymbol{\omega}})\mathbf{x}. \tag{43}$$

Putting $\hat{\mathbf{h}}$ into (42), the estimate of $\boldsymbol{\omega}$ can be obtained as

$$\hat{\boldsymbol{\omega}} = \arg\max_{\tilde{\boldsymbol{\omega}}}\left\{J'(\tilde{\boldsymbol{\omega}}) \triangleq \|\mathbf{P}_{\mathbf{Q}}(\tilde{\boldsymbol{\omega}})\mathbf{x}\|^2\right\} \tag{44}$$

where $\mathbf{P}_{\mathbf{Q}}(\tilde{\boldsymbol{\omega}}) = \mathbf{Q}(\tilde{\boldsymbol{\omega}})(\mathbf{Q}^H(\tilde{\boldsymbol{\omega}})\mathbf{Q}(\tilde{\boldsymbol{\omega}}))^{-1}\mathbf{Q}^H(\tilde{\boldsymbol{\omega}})$. The above CFO estimation in (44) requires an exhaustive search over the multidimensional space spanned by $\tilde{\boldsymbol{\omega}}$, which may be too computationally expensive in implementation.

Remark 6: If the training sequences are designed such that (39) is satisfied and when the number of subcarriers is large enough, we have the following:

$$\begin{aligned}\left(\mathbf{Q}^H(\boldsymbol{\omega})\mathbf{Q}(\boldsymbol{\omega})\right)^{-1} =& \mathbf{I}_{M_r\times M_r} \\ &\otimes \begin{bmatrix} \left(\mathbf{A}_1^H\mathbf{A}_1\right)^{-1} & \dots & \mathbf{0} \\ \vdots & \ddots & \vdots \\ \mathbf{0} & \vdots & \left(\mathbf{A}_k^H\mathbf{A}_k\right)^{-1}\end{bmatrix} \\ \triangleq& \mathbf{B}. \end{aligned} \quad (45)$$

Then the CFO estimation for all users in (44) will be decoupled as

$$\begin{aligned}\hat{\omega}_k = \arg\max_{\tilde{\omega}_k} &\left\{\sum_{i=1}^{Mr}\mathbf{x}_i^H\left\{\boldsymbol{\Gamma}(\tilde{\omega}_k)\mathbf{A}_k\left(\mathbf{A}_k^H\mathbf{A}_k\right)^{-1}\right.\right. \\ &\left.\left.\times\mathbf{A}_k^H\boldsymbol{\Gamma}^H(\tilde{\omega}_k)\right\}\mathbf{x}_i\right\} \quad (k=1,\dots,K). \end{aligned}\quad (46)$$

Unfortunately, (45) and (46) hold exactly only when the number of subcarriers is infinite. For a practical system with finite subcarriers, the results from (46) are only approximated solutions to the original estimation problem. In the simulation section, we will show that the decoupled estimator in (46) suffers great performance loss when the number of subcarriers is not large enough. Thus, an efficient algorithm which can find the exact solution of (44) with affordable complexity is needed.

Remark 7: For the pilot-aided estimation problem to be valid, a necessary condition is the existence of $(\mathbf{Q}^H(\boldsymbol{\omega})\mathbf{Q}(\boldsymbol{\omega}))^{-1}$. Due to the fact that $\text{rank}\{\mathbf{Q}(\boldsymbol{\omega})\} = M_r \cdot \text{rank}\{\mathbf{A}\} \leq M_r \cdot \min\{N, KL\}$ and $\text{rank}\{\mathbf{A}^H\mathbf{A}\} = KL$, the following requirement is needed: $KL \leq N$. Thus, we conclude that at most N/L users can be involved in the system during the joint CFO and channel estimation period. This constraint is not restrictive in practical wireless network applications due to the large number of subcarriers and the relatively small L. For example, for WiMax setup ($N = 1024$, $L = 16$), a large number of users $(N/L = 64)$ can be supported.

B. Estimation via Importance Sampling

For the problem in (44), though iterative approaches, such as alternating projection and EM algorithms ([6], [23]) can be used, a good initial guess of the CFOs is required. Besides, there is no guarantee that an estimate obtained iteratively will be the global maximum. To avoid this, Pincus [24] showed that it is possible to obtain a closed form solution for the parameter $\boldsymbol{\omega}$ that guarantees to be the global solution. Based on the theorem given by Pincus, the $\boldsymbol{\omega}$ that yields the global maximum of $J'(\boldsymbol{\omega})$, is given by

$$\hat{\omega}_k = \lim_{\rho\to\infty}\frac{\int\cdots\int\omega_k\exp\left(\rho J'(\boldsymbol{\omega})\right)d\boldsymbol{\omega}}{\int\cdots\int\exp\left(\rho J'(\boldsymbol{\omega})\right)d\boldsymbol{\omega}} \quad k=1,\dots,K \quad (47)$$

where ρ is a design parameter. If we denote $J(\boldsymbol{\omega}) = \exp(\rho J'(\boldsymbol{\omega}))$, the normalized version of $J(\boldsymbol{\omega})$ can be obtained by

$$\bar{J}(\boldsymbol{\omega}) = \frac{J(\boldsymbol{\omega})}{\int\cdots\int J(\boldsymbol{\omega})d\boldsymbol{\omega}}. \quad (48)$$

Then, the function $\bar{J}(\boldsymbol{\omega})$ has all the properties of a probability density function (pdf), so it is termed as the pseudo-pdf in $\boldsymbol{\omega}$. With this definition and (47), the optimal solution of $\boldsymbol{\omega}$ in (44) is

$$\hat{\omega}_k = \int\cdots\int\omega_k\bar{J}(\boldsymbol{\omega})d\boldsymbol{\omega}, \quad k=1,\dots,K \quad (49)$$

for some large value of ρ. Due to the fact that the frequency offsets $(\omega_k, k=1,\dots,K)$ has the properties of a circular random variable, the estimate for ω_k in (49) can be rewritten as

$$\hat{\omega}_k = \frac{1}{2\pi}\angle\int\cdots\int\exp(j2\pi\omega_k)\bar{J}(\boldsymbol{\omega})d\boldsymbol{\omega}, \quad k=1,\dots,K \quad (50)$$

where $\angle$ denotes the operation of finding the angle of the complex number. The advantage of using (50) instead of (49) is that (50) eliminates a potential bias in $\hat{\omega}_k$.

In [25], importance sampling is used to compute the multidimensional integral in (50). This approach is based on the observation that integrals of the type $\int\cdots\int h(\boldsymbol{\omega})\bar{J}(\boldsymbol{\omega})d\boldsymbol{\omega}$ can be expressed as

$$\int\cdots\int h(\boldsymbol{\omega})\bar{J}(\boldsymbol{\omega})d\boldsymbol{\omega} = \int\cdots\int h(\boldsymbol{\omega})\frac{\bar{J}(\boldsymbol{\omega})}{\bar{g}(\boldsymbol{\omega})}\bar{g}(\boldsymbol{\omega})d\boldsymbol{\omega} \quad (51)$$

where

$$\bar{g}(\boldsymbol{\omega}) = \frac{g(\boldsymbol{\omega})}{\int\cdots\int g(\boldsymbol{\omega})d\boldsymbol{\omega}} \quad (52)$$

with $g(\boldsymbol{\omega}) > 0$. In above, the function $g(\boldsymbol{\omega})$ is called the importance function and its normalized version $\bar{g}(\boldsymbol{\omega})$ has all the properties of a pdf. Then, the RHS of (51) can be expressed as the expected value of $h(\boldsymbol{\omega})(\bar{J}(\boldsymbol{\omega})/\bar{g}(\boldsymbol{\omega}))$ with respect to the pseudo-pdf $\bar{g}(\boldsymbol{\omega})$. If we can generate realizations of $\boldsymbol{\omega}$ according to $\bar{g}(\boldsymbol{\omega})$, the value of the integral in (51) can be found by the Monte Carlo approximation as

$$\int\cdots\int h(\boldsymbol{\omega})\bar{J}(\boldsymbol{\omega})d\boldsymbol{\omega} \approx \frac{1}{T}\sum_{i=1}^{T}h(\boldsymbol{\omega}^i)\frac{\bar{J}(\boldsymbol{\omega}^i)}{\bar{g}(\boldsymbol{\omega}^i)} \quad (53)$$

where T is the number of realizations and $\boldsymbol{\omega}^i$ is the ith realization of the vector $\boldsymbol{\omega}$ generated according to the pseudo-pdf $\bar{g}(\boldsymbol{\omega})$. With $h(\boldsymbol{\omega}) = \exp(j2\pi\omega_k)$, we can obtain the estimate of CFO using importance sampling as

$$\hat{\omega}_k = \frac{1}{2\pi}\angle\frac{1}{T}\sum_{i=1}^{T}\frac{\bar{J}(\boldsymbol{\omega}^i)}{\bar{g}(\boldsymbol{\omega}^i)}\exp\left(j2\pi\omega_k^i\right) \quad k=1,\dots,K. \quad (54)$$

Note that it is only necessary to calculate the angle of the complex value in (54). The equivalent but simplified estimator is given by

$$\hat{\omega}_k = \frac{1}{2\pi}\angle\frac{1}{T}\sum_{i=1}^{T}\frac{J(\boldsymbol{\omega}^i)}{g(\boldsymbol{\omega}^i)}\exp\left(j2\pi\omega_k^i\right) \quad k=1,\dots,K. \quad (55)$$

In general, the estimator (55) converges to (47) by the Strong Law of Large Number regardless the choice of the function $g(\boldsymbol{\omega})$. However, there are obviously some choices of $g(\boldsymbol{\omega})$ that are better than others in terms of computational complexity. From [26], the optimal importance function $g_o(\boldsymbol{\omega})$ which minimizes the variance of the estimator for a fixed number of realizations T is given by

$$\bar{g}_o(\boldsymbol{\omega}) = \frac{|h(\boldsymbol{\omega})|\,\bar{J}(\boldsymbol{\omega})}{\int\cdots\int|h(\boldsymbol{\omega})|\,\bar{J}(\boldsymbol{\omega})d\boldsymbol{\omega}}. \quad (56)$$

Since $h(\boldsymbol{\omega}) = \exp(j2\pi\omega_k)$, $|\exp(j2\pi\omega_k)| = 1$ holds. Thus we have

$$g_o(\boldsymbol{\omega}) = \frac{\int \cdots \int g_o(\boldsymbol{\omega}) d\boldsymbol{\omega}}{\int \cdots \int J(\boldsymbol{\omega}) d\boldsymbol{\omega}} J(\boldsymbol{\omega}) \propto J(\boldsymbol{\omega}). \tag{57}$$

It is obvious that the optimal importance function $g_o(\boldsymbol{\omega})$ for the estimator at hand should be a scaled version of $J(\boldsymbol{\omega})$. However, when $g(\boldsymbol{\omega})$ is chosen like this, the implementation of the estimator in (55) will be challenging due to the difficulty in sample generation from a multidimensional pdf. Considering the performance and ease of sample generation, the exploited importance function $g(\boldsymbol{\omega})$ is relaxed to be a close approximation of $J(\boldsymbol{\omega})$, and at the same time it should be as simple as possible to facilitate sample generation [27].

From (46), it is noted that if the training sequences are designed such that (39) is satisfied, $J'(\boldsymbol{\omega})$ in (44) become separable in ω_i. Based on this observation and the considerations aforementioned, the importance function $g(\boldsymbol{\omega})$ is designed as

$$\begin{aligned} g(\boldsymbol{\omega}) &= \exp\left(\rho \mathbf{x}^H \mathbf{Q}(\boldsymbol{\omega}) \mathbf{B} \mathbf{Q}^H(\boldsymbol{\omega}) \mathbf{x}\right) \\ &= \exp\left(\sum_{i=1}^{Mr} \sum_{k=1}^{K} \rho \mathbf{x}_i^H \left\{ \boldsymbol{\Gamma}(\omega_k) \mathbf{A}_k \left(\mathbf{A}_k^H \mathbf{A}_k\right)^{-1} \right.\right. \\ &\qquad \left.\left. \times \mathbf{A}_k^H \boldsymbol{\Gamma}^H(\omega_k) \right\} \mathbf{x}_i \right) \\ &= \prod_{k=1}^{K} \prod_{i=1}^{Mr} \exp\left(\rho \mathbf{x}_i^H \boldsymbol{\Gamma}(\omega_k) \mathbf{A}_k \left(\mathbf{A}_k^H \mathbf{A}_k\right)^{-1} \right. \\ &\qquad \left. \times \mathbf{A}_k^H \boldsymbol{\Gamma}^H(\omega_k) \mathbf{x}_i \right) \\ &\triangleq \prod_{k=1}^{K} g_k(\omega_k). \end{aligned} \tag{58}$$

Then the generation of realizations of $\boldsymbol{\omega}$ reduces to the generation of K independent realizations of ω_k following the pseudo-pdfs

$$\bar{g}_k(\omega_k) = \frac{g_k(\omega_k)}{\int_{-\alpha}^{\alpha} g_k(\omega_k) d\omega_k}, \quad k = 1, \ldots, K \tag{59}$$

where α is the estimation range of CFOs.

Remark 8: In general, with any importance function, the proposed estimator (55) can guarantee to obtain the global optimal solution as long as T is large enough [27]. All the above efforts on importance function selection are based on the complexity consideration. Since the cost function (44) is derived based on the ML criterion, the performance of the proposed estimator (55) would approach the derived CRB asymptotically.

Remark 9: In the literature, several sequential monte carlo based methods (e.g., particle filter) have been proposed for tracking time varying synchronization and channel parameters [28]–[30]. Notice that the present work is dealing with one-shot estimation based on ML framework. Due to the difference in problem natures, the methods in [28]–[30] are not applicable to the present problem.

C. Efficient Algorithms for Sample Generation

It is pointed out in [25] that there are two major sources of computations in the importance sampling approach: the first one is the generation of realizations using the pseudo-pdf (59); and the second one is the calculation of the coefficient $L(\boldsymbol{\omega}^i)/g(\boldsymbol{\omega}^i)$ in (55). Since the complexity of the second step is constant for a fixed T, to reduce the complexity of the whole algorithm, it is critical to focus on the first one.

To generate realizations of ω_k according to the pseudo-pdf $\bar{g}_k(\omega_k)$, several methods are available in the literature [31]–[33]. The conceptually simplest one is the inverse cumulative density function (CDF) approach, which is employed in [25]. For the inverse CDF approach, a random variable U uniformly distributed in (0,1) is generated first and then the realization of ω_k can be obtained by solving the equation $U = G_k(\omega_k)$, where $G_k(\omega_k) = \int_{-\alpha}^{\omega_k} \bar{g}_k(x) dx$ can be regarded as the CDF of the pseudo-pdf $\bar{g}_k(\omega_k)$. To solve the equation, 24 evaluations of $G_k(\omega_k)$ are needed even if the golden search is exploited [25]. Besides, for one evaluation of $G_k(\omega_k)$, the $\bar{g}_k(\omega_k)$ has to be evaluated for many times due to the inherent complexity in numerical integration. Thus, the complexity for the inverse CDF method is very high and not suitable for the problem at hand. It is desirable to exploit other techniques which can reduce the evaluation times of pdf $\bar{g}_k(\omega_k)$ for generating ω_k.

Here we propose to use the ratio-of-uniform method, since this technique is quite fast and has a moderate performance [33]. This method is based on the fact that, if (U,V) is uniformly distributed over the set $C = \{(u, v); 0 \le u \le \sqrt{\bar{g}_k(v/u)}\}$, then v/u has the probability density function as $\bar{g}_k(\omega_k)$. To implement this method, the following procedures are needed [33]:

1) Choose a rectangle E which encloses C. From [34], a simple form of the rectangle E that encloses C can be obtained by

$$E = \left\{ (u, v); 0 \le u \le \max_{\omega_k} \sqrt{\bar{g}_k(\omega_k)}, \right. \\ \left. \min_{\omega_k} \omega_k \sqrt{\bar{g}_k(\omega_k)} \le v \le \max_{\omega_k} \omega_k \sqrt{\bar{g}_k(\omega_k)} \right\} \tag{60}$$

where $\omega_k \in [-\alpha, \alpha]$.

2) Generate two random variables (u, v) in the domain E under a uniform distribution.

3) If the generated random variables (u, v) satisfies $u^2 \le \bar{g}_k(v/u)$, one realization of ω_k is generated as $\omega_k = v/u$; otherwise reject the (u, v) and go back to the step 2.

For the system with small estimation range α, the above method can work efficiently. However, in practical systems, there definitely are some cases where a moderate or large CFO estimation range is preferred, such as in the presence of large frequency synchronization errors during initialization or with large Doppler shift caused by rapid increase of moving speed. When the estimation range α is moderate or large, the region for random variable v in (60) may be large. In this case, after generating (u, v), the chance for rejecting this pair is large such that the step 2 has to be repeated for many times to generate one realization of ω_k. In other words, the number of evaluation times of $\bar{g}_k(\omega_k)$ is significant for generating one ω_k successfully.

To reduce the complexity in this situation, we revised the first step of the above method as follows. Denoting $\bar{\omega}_k$ as a rough estimate of the maximal point of $\sqrt{\bar{g}_k(\omega_k)}$, we define

$$E' = \left\{ (u', v'); 0 \le u' \le \max_{\omega_k} \sqrt{\bar{g}_k(\omega_k)}, \right. \\ \left. \min_{\omega'_k} \omega'_k \sqrt{\bar{g}_k\left(\omega'_k + \bar{\omega}_k\right)} \le v' \le \max_{\omega'_k} \omega'_k \sqrt{\bar{g}_k\left(\omega'_k + \bar{\omega}_k\right)} \right\} \tag{61}$$

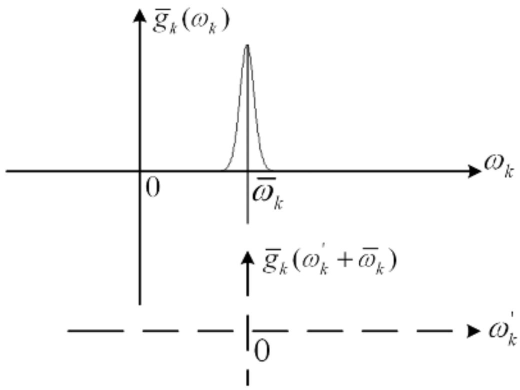


Fig. 1. Idea of the proposed sample generation method.

where $\omega_k' \in [-\alpha-\bar{\omega}_k, \alpha-\bar{\omega}_k]$. If the generated (u', v') satisfies $u'^2 \le \bar{g}_k(v'/u'+\bar{\omega}_k)$, the value of $(v'/u'+\bar{\omega}_k)$ will be accepted as a realization of ω_k. This idea can be illustrated by Fig. 1. Since $\bar{\omega}_k$ approximately corresponds to the maximal point of the function $\bar{g}_k(\omega_k)$, which is normally near the real CFO value, most realizations of ω_k will be around the $\bar{\omega}_k$. If we shift the zero point to $\bar{\omega}_k$ and construct a new axis framework, the bounds of v' in (61) can always be small values even if the real CFO is large.

Remark 10: Although a matrix inverse $(\mathbf{A}_k^H \mathbf{A}_k)^{-1}$ (see (58)) is included in $g_k(\omega_k)$, this term can be calculated offline. Furthermore, because the estimation range α is a design parameter which is known at receiver, the term $\int_{-\alpha}^{\alpha} g_k(\omega_k) d\omega_k$ in (59) can also be calculated once and stored.

Remark 11: It is noticed that searching the maximum of the function $\bar{g}_k(\omega_k)$ in step one is equivalent to finding the solution of (46). The complexity is affordable since only multiple one-dimensional search are needed. This step can be regarded as a coarse search for the parameters to be estimated.

Remark 12: Using the proposed method, the generation of ω_k is only loosely coupled with the estimation range through the bound of v'. It is shown in the simulation section that, with different estimation ranges, the complexity of the proposed sample generation algorithm is almost constant. This is a significant advantage over the conventional CFO estimation schemes, in which the required computation increases in proportion to the estimation range.

V. Simulation Results and Discussions

A. Simulation Setup

In this section, simulation results are presented to demonstrate the effectiveness of the proposed schemes. The considered MIMO-OFDM system has the following parameters: $N = 64$, $L_{\text{cp}} = 16$, $M_r = 2$. The channel response for each user, is generated according to the HIPERLAN/2 channel model with eight paths $(L = 8)$ [35]. In details, the channel taps are modeled as independent and complex Gaussian random variables with zero mean and an exponential power delay profile

$$E\left\{|\mathbf{h}_{i,j}(l)|^2\right\} = \lambda_{i,j} \cdot \exp\{-l\}, \quad l = 0, \ldots, L-1 \quad (62)$$

where parameter $\lambda_{i,j}$ denotes the power of user i observed at the receive antenna j. Since the receive antennas at the base station are co-located, we have $\lambda_{i,1} = \cdots = \lambda_{i,M_r} \triangleq \lambda_i$. In the simulation, it is assumed that the users' signals arrive at the base station with equal power, so $\lambda_1 = \cdots = \lambda_K$ holds. Without loss of generality, the normalized CFOs for all user $(\varepsilon_k, k = 1, \ldots, K)$ in each packet are generated as random variables uniformly distributed in $[-N/2, N/2]$. Notice that this corresponds to a large CFO range. The proper choice of the design parameter ρ can significantly reduce the number of realizations T needed for one estimation. Following the suggestions in [25], in the simulation, we choose $\rho = 1/K$. For each CFO estimate, 5000 realizations $(T = 5000)$ are generated for the importance sampling estimation. The signal-to-noise ratio (SNR) used in the simulation is defined as the $\text{SNR} = \eta/\sigma^2$, where η is the total power of the received signal at the receiver. Since there are multiple CFOs and channels to be estimated in the scheme, the MSE performance is defined as

$$\text{MSE}_{CFO} = \sum_{k=1}^{K} (\hat{\omega}_k - \omega_k)^2$$
$$\text{MSE}_{channel} = ||\hat{\mathbf{h}} - \mathbf{h}||^2.$$

All the simulation results of the proposed algorithm are averaged over 200 Monte Carlo runs.

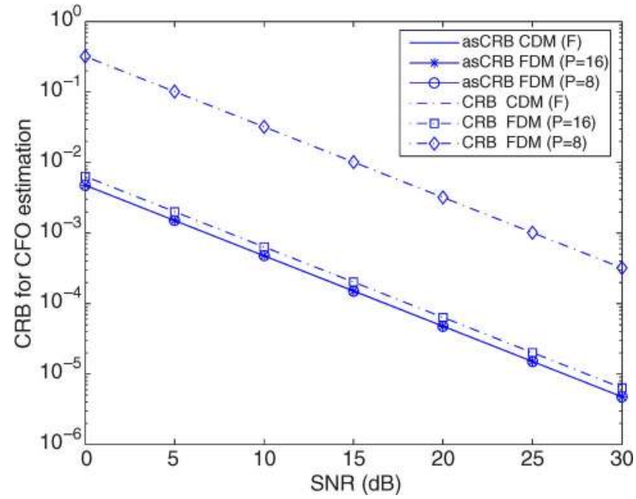


Fig. 2. CRB, asCRB of CFO estimation for different training sequences.

B. Validity of the Proposed Pilot

In Fig. 2 and Fig. 3, the CRBs and asCRBs for the CFO and channel estimation with different training sequences are plotted $(K = 2)$. Three kinds of training sequences are assessed in the simulation, one is the proposed CDM(F) sequence in (40) and the other two are FDM sequences in [18]. For the FDM sequences, the number of non-zero tones is denoted as P. In both figures, it is noticed that all the test sequences have the same asCRBs as expected. For the CDM(F) sequence, its CRB is the closest to the asCRB while the CRBs of the two FDM sequences depart from the asCRBs with the decrease of P. Recall that when P becomes smaller, the FDM sequences will be more correlated due to its shorter correlation length. Fig. 2 and Fig. 3 clearly validate the pilot design in Section III.

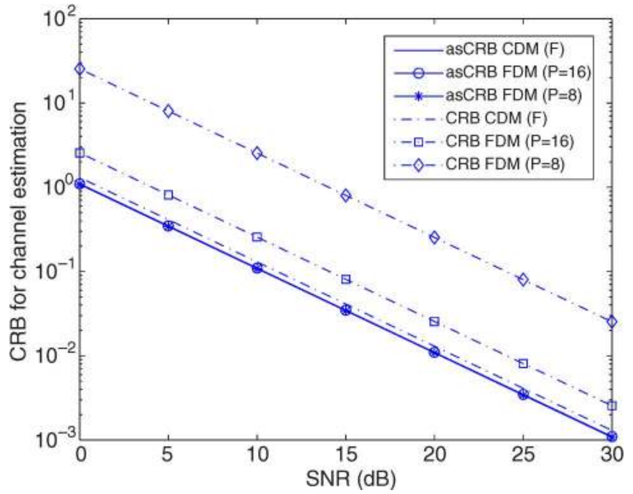


Fig. 3. CRB, asCRB of channel estimation for different training sequences.

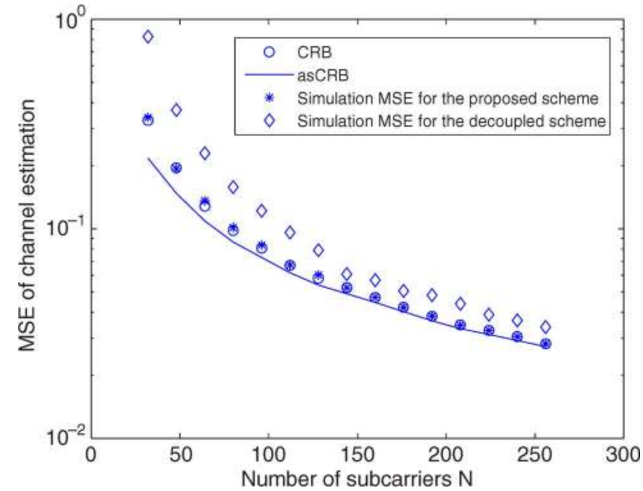


Fig. 5. CRB, asCRB, and MSE for the channel estimator versus different number of subcarriers.

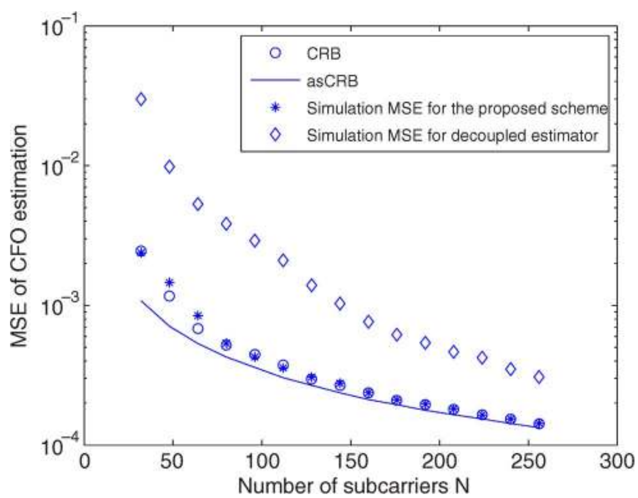


Fig. 4. CRB, asCRB, and MSE for the proposed CFO estimator versus different number of subcarriers.

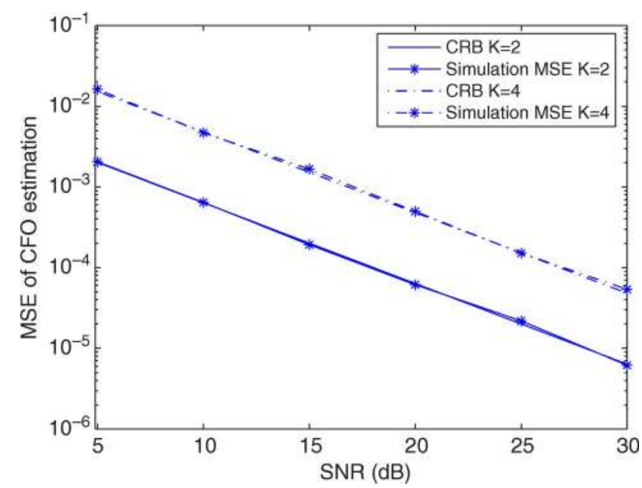


Fig. 6. The MSE performance for the proposed CFO estimator $K = 2, 4$.

C. Asymptotic Performance

In Figs. 4 and 5, the CRB, asCRB, and MSE for the CFO and channel estimation $(K = 2)$ using the proposed scheme are presented versus different number of subcarriers (N). The SNR in this simulation is set as 10 dB. It is noticed that even when N is small, there is only a small gap between the CRB and asCRB, which further validates the training design based on the asCRB in Section III. Besides, it should also be pointed out that the gap between CRB and asCRB becomes minimal with the increase of N. In both figures, the MSEs from simulations meet the respective CRB quite well, which show that the proposed scheme is efficient. The performances of the joint CFO and channel estimation scheme using the decoupled CFO estimator in (46) are also included in the two figures as references. It is noticed that for the decoupled estimator, a significant performance loss occurs when the number of subcarriers is small. On the other hand, the gap between the decoupled estimator and the proposed estimator decreases when the number of subcarriers increases This implies that when the number of subcarriers is large, we can use the decoupled estimator, which is computational efficient with marginal loss of performance.

D. Efficiency of the Proposed Estimation Scheme

To show the effectiveness of the proposed scheme in the whole SNR region of interested, the MSE performance of the CFO and channel estimators are plotted versus SNR (with 2 and 4 users) in Fig. 6 and Fig. 7. The CRBs are also shown in the two figures as references. For the proposed CFO and channel estimator, the MSEs always coincide with the respective CRBs which means that the proposed estimation scheme is efficient.

E. Proposed Pilot Versus TDM Pilot

If the TDM sequences are used in the training, the estimation of the CFO and channel for each user can be performed separately like a single user case. To obtain a fair comparison, the length of the whole training block is fixed and all users share the entire training block equally. For example, if there are 2 user in the system, the TDM training sequence for each user has 32

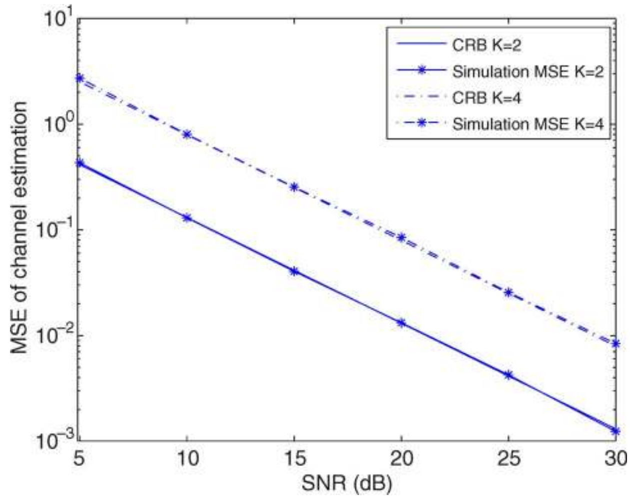


Fig. 7. The MSE performance for the proposed channel estimator $K = 2, 4$.

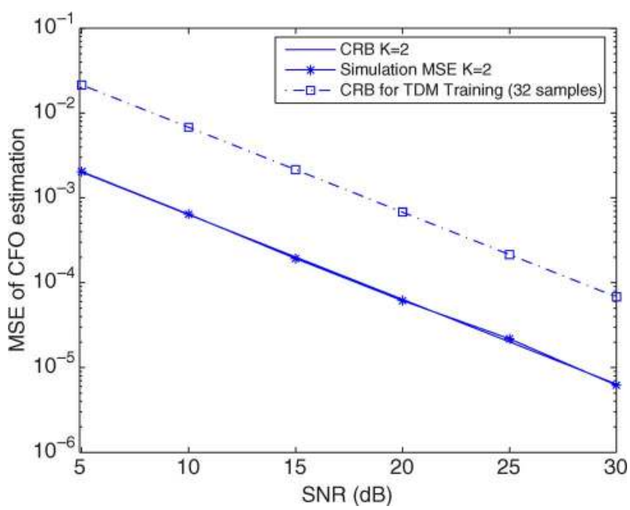


Fig. 8. The MSE comparison for CFO estimation using the proposed training sequences and TDM type training sequences.

samples in the time domain. All the training sequences are constructed by white sequences. Since CRB is the lower bound of estimation performance, it is sensible to compare the CRB of CFO and channel estimation in TDM training case (derived in [14]) with the performance of the proposed scheme. It is shown in Fig. 8 and Fig. 9 that for the TDM training, there will be a significant performance loss compared to the proposed scheme.

F. Complexity Comparison

At last, the complexity analysis is performed. For the problem in (44), the ML estimator using exhaustive search is impractical for implementation due to the high complexity. The decoupled CFO estimator in (46) has a reduced complexity, but this estimator cannot yield the optimal solution when the number of subcarriers is finite. For the proposed estimator using importance sampling, as mentioned previously, the complexity is determined by the samples generation. To show the computational advantage of the proposed approach, in Fig. 10, the averaged pdf evaluation times for generating one realization of ω_k is plotted versus $\alpha \times N/(2\pi)$. The same results for the inverse CDF method and the original ratio-of-uniform method are also included as comparisons. The SNR in the simulation is set as

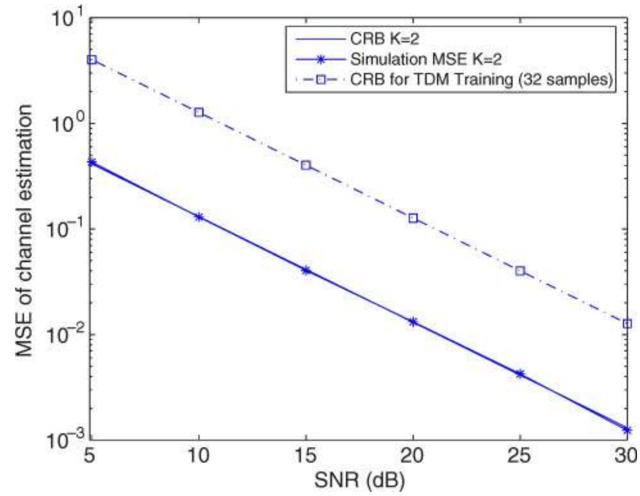


Fig. 9. The MSE comparison for channel estimation using the proposed training sequences and TDM type training sequences.

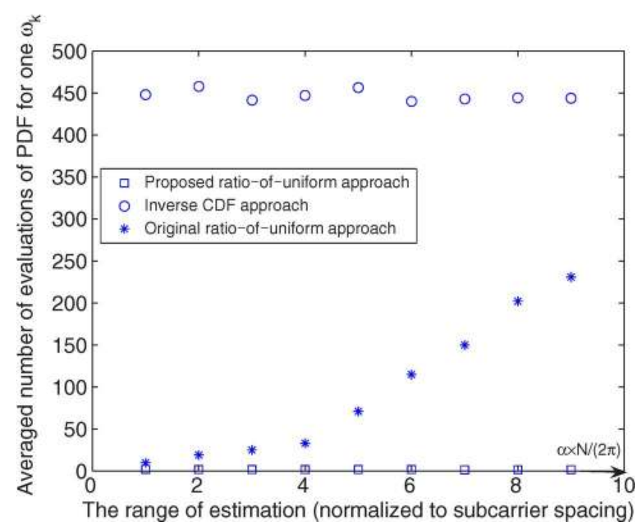


Fig. 10. The averaged pdf evaluation times for generating one realization of ω_k.

20 dB. It is noticed that for any estimation range, the pdf evaluation times needed for inverse CDF method is far larger than that of the other two approaches. For the original ratio-of-uniform method, the complexity is comparable with the proposed method when estimation range α is small. However, its complexity increases quickly with the increase of α. Compared with the above two methods, the proposed method has the lowest complexity, which remains constant for all estimation range. In fact, while cannot be seen in Fig. 10 because of the scale of the figure, the proposed method only needs 2–3 evaluations of $\bar{g}_k(\omega_k)$ for each successful generation of ω_k, regardless of the estimation range α.

VI. CONCLUSION

In this paper, the problem of joint CFO and channel estimation in multiuser MIMO-OFDM system has been considered. Based on the asymptotic CRB analyses and under the constraint that the performance limits of CFO and channel estimation are channel independent, the condition for optimal training sequences has been derived. Among the two well-known classes of sequences that satisfy the optimal condition, it was found that the CDM(F) sequences perform better than FDM sequences when the number of subcarriers is finite. The joint ML CFO and channel estimator was then derived. However, direct implementation is impractical due to the multidimensional search required in the CFO estimation. Unlike the conventional iterative algorithms usually used in multidimensional optimization, a monte carlo importance sampling based estimator has been proposed to solve this problem. To reduce the complexity in implementation, an efficient importance function and a novel modified sample generation algorithm have been proposed. The proposed scheme has two merits: 1) global optimality is naturally guaranteed; 2) the algorithm has almost constant complexity for different estimation ranges. Simulation results have clearly illustrated the merits of the proposed training sequence design and also verified the effectiveness of the proposed scheme.

REFERENCES

[1] H. Sampath, S. Talwar, J. Tellado, V. Erceg, and A. Paulraj, "A fourth-generation MIMO-OFDM broadband wireless system: Design, performance, and field trial results," *IEEE Commun. Mag.*, vol. 40, pp. 143–149, Sep. 2002.

[2] P. W. Wolniansky, G. J. Foschini, G. D. Golden, and R. A. Valenzuela, "V-BLAST: An architecture for realizing very high data rates over the rich-scattering wireless channel," in *Proc. ISSSE-98*, Pisa, Italy, Sept. 29, 1998, invited paper.

[3] M. Jiang, J. Akhtman, and L. Hanzo, "Soft-information assisted near-optimum nonlinear detection for BLAST-type space division multiplexing OFDM systems," *IEEE Trans. Wireless Commun.*, vol. 6, no. 4, pp. 1230–1234, Apr. 2007.

[4] H. Lee and I. Lee, "New approach for error compensation in coded V-BLAST OFDM systems," *IEEE Trans. Commun.*, vol. 55, no. 2, pp. 345–355, Feb. 2007.

[5] H. Lee, B. Lee, and I. Lee, "Iterative detection and decoding with an improved V-BLAST for MIMO-OFDM systems," *IEEE J. Sel. Areas Commun.*, vol. 24, no. 3, pp. 504–513, Mar. 2006.

[6] M.-O. Pun, M. Morelli, and C.-C. J. Kuo, "Iterative detection and frequency synchronization for OFDMA uplink transmissions," *IEEE Trans. Wireless Commun.*, vol. 6, no. 2, pp. 629–639, Feb. 2007.

[7] P. H. Moose, "A technique for orthogonal frequency division multiplexing frequency offset correction," *IEEE Trans. Commun.*, vol. 42, no. 10, pp. 2908–2914, 1994.

[8] A. J. Coulson, "Maximum likelihood synchronization for OFDM using a pilot symbol: Algorithm," *IEEE J. Sel. Areas Commun.*, vol. 19, no. 12, pp. 2486–2494, Dec. 2001.

[9] J. J. van de Beek, O. Edfors, and M. Sandell, "On channel estimation in OFDM systems," in *Proc. Veh. Technol. Conf.*, Chicago, IL, Jul. 1995, vol. 2, pp. 815–819.

[10] M. Morelli and U. Mengali, "A comparison of pilot-aided channel estimation methods for OFDM systems," *IEEE Trans. Signal Process.*, vol. 49, no. 12, pp. 3065–3073, Dec. 2001.

[11] L. Weng, R. D. Murch, and V. K. N. Lau *et al.*, "Effect of carrier frequency offset on channel estimation for SISO/MIMO-OFDM systems," *IEEE Trans. Wireless Commun.*, vol. 6, no. 5, pp. 1854–1863, May 2007.

[12] Y. Zeng, W. Leon, Y. C. Liang, and A. R. Leyman, "A new method for frequency offset and channel estimation in OFDM," in *Proc. IEEE ICC*, Jun. 2006, vol. 10, pp. 4606–4611.

[13] J. Chen, Y.-C. Wu, and T. S. Ng, "Timing robust joint carrier frequency offset and channel estimation for OFDM systems," in *Proc. IEEE WCNC*, Mar. 2007, pp. 1025–1029.

[14] 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.

[15] O. Besson and P. Stoica, "On parameter estimation of MIMO flat-fading channels with frequency offsets," *IEEE Trans. Signal Process.*, vol. 51, no. 3, pp. 602–613, Mar. 2003.

[16] M. Ghogho and A. Swami, "Training design for multipath channel and frequency-offset estimation in MIMO systems," *IEEE Trans. Signal Process.*, vol. 54, no. 10, pp. 3957–3965, Oct. 2006.

[17] S. Sezginer, P. Bianchi, and W. Hachem, "Asymptotic Cramér-Rao bounds and training design for uplink MIMO-OFDMA systems with frequency offsets," *IEEE Trans. Signal Process.*, vol. 55, no. 7, pp. 3606–3622, Jul. 2007.

[18] H. Minn and N. Al-Dhahir, "Optimal training signals for MIMO OFDM channel estimation," *IEEE Trans. Wireless Commun.*, vol. 5, no. 5, pp. 1158–1168, May 2006.

[19] H. Minn, N. Al-Dhahir, and Y. Li, "Optimal training signals for MIMO OFDM channel estimation in the presence of frequency offset and phase noise," *IEEE Trans. Commun.*, vol. 54, no. 10, pp. 1754–1759, Oct. 2006.

[20] J. Chen, Y.-C. Wu, S. Ma, and T. S. Ng, "Training design for joint CFO and channel estimation in multiuser MIMO OFDM system," in *Proc. IEEE Globecom*, Nov. 2007, pp. 3008–3012.

[21] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge, London, U.K.: Cambridge Univ. Press, 1985.

[22] X. Fu and H. Minn, "TDMA-type preamble for low complexity multiuser frequency synchronization in OFDMA uplink," in *IEEE VTC*, Sep. 25–28, 2005, vol. 2, pp. 1093–1097.

[23] M.-O. Pun, M. Morelli, and C.-C. Jay Kuo, "Maximum-likelihood synchronization and channel estimation for OFDMA uplink transmissions," *IEEE Trans. Commun.*, vol. 54, no. 4, pp. 726–736, Apr. 2006.

[24] M. Pincus, "A closed form solution for certain programming problems," *Oper. Res.*, pp. 690–694, 1962.

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

[26] C. P. Robert and G. Casella, *Monte Carlo Statistical Methods*. New York: Springer, 2004.

[27] A. F. M. Smith and A. Gelfand, *Bayesian Statistics Without Tears: A Sampling-Resampling Framework*. New York: Amer. Stat. Assoc., 1992.

[28] K. J. Kim and R. A. Iltis, "A sequential monte-carlo kalman filter based frequency offset and channel estimation method in the MIMO-OFDM system," in *Proc. VTC 2005-Fall*, 2005, pp. 277–281.

[29] K. J. Kim and R. A. Iltis, "Frequency offset synchronization and channel estimation for the MIMO-OFDM system using Rao-Blackwellized Gauss-Hermite filter," in *Proc. Asilomar Conf. Signals Syst. Comput.*, Pacific Grove, CA, Oct. 2005, pp. 829–833.

[30] T. Nyblom, T. Roman, M. Enescu, and V. Koivunen, "Time-varying carrier offset tracking in OFDM systems using particle filtering," in *Proc. 4th IEEE Int. Symp. Signal Process. Inf. Technol.*, Dec. 2004, pp. 217–220.

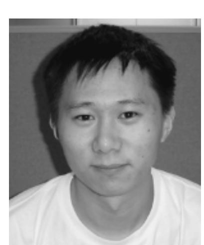
[31] A. Hirosha, "A new method of interpolation and smooth curve fitting based on local procedures," *J. ACM*, vol. 17, pp. 589–602.

[32] G. Marsaglia, "The exact-approximation method for generating random variables in a computer," *J. Amer. Statist. Assoc.*, vol. 79, pp. 218–221.

[33] A. J. Kinderman and J. F. Monahan, "Computer generation of random variables using the ratio of uniform deviates," *ACM Trans. Math. Software*, vol. 3, pp. 257–260, 1977.

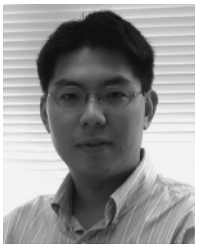
[34] J. E. Gentle, *Random Number Generation and Monte Carlo Methods*, 2nd ed. New York: Springer, 2003.

[35] ETSI, 2001, BRAN; HIPERLAN Type 2; Physical (PHY) Layer Specification (2nd ed.) Technical Specification 101 475 [Online]. Available: http://www.etsi.org



Jianwu Chen received the B.S. degree in 2002 and the M.S. degree in 2004 from Huazhong University of Science and Technology (HUST), China, both in electronic engineering.

He is currently pursuing the Ph.D. degree at the Department of Electrical and Electronic Engineering, the University of Hong Kong. His current research interests include synchronization for OFDM systems, channel equalization, OFDMA, MIMO, and Monte Carlo algorithms.



Yik-Chung Wu received the B.Eng. (EEE) degree in 1998 and the M.Phil. degree in 2001 from The University of Hong Kong (HKU). He received the Ph.D. degree in 2005 from Texas A&M University. During his study at Texas A&M University, he was fully supported by the prestigious Croucher Foundation scholarship.

In 2001, he was a Research Assistant with HKU. From August 2005 to August 2006, he was with Thomson Corporate Research, Princeton, NJ, as a Member of Technical Staff. Since September 2006, he has been with the University of Hong Kong as an Assistant Professor. His research interests are in general area of signal processing for communication systems, and in particular, receiver algorithm design, synchronization techniques, channel equalization, software radio, OFDM, MIMO, cooperative systems, Bayesian inference, Monte Carlo statistical methods, and low-complexity implementation of DSP algorithms.

Dr. Wu was a Technical Program Committee Member for IEEE VTC Fall 2005, Globecom 2006, 2008, ICC 2007, and 2008. He currently serves as an Associate Editor for the IEEE COMMUNICATIONS LETTERS.

Shaodan Ma received the B.Sc. (Eng.) and M.Eng.Sc. 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 Postdoctoral Fellow. Her research interests include wireless communication systems, spread spectrum techniques, MIMO systems, OFDM technique, and digital signal processing.

Tung-Sang Ng (S'74–M'78–SM'90–F'03) received the B.Sc. (Eng.) degree from The University of Hong Kong in 1972, and 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 he returned to The University of Hong Kong in 1991, where he took 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 more than 250 international journal and conference papers.

Dr. Ng was the General Chair of ISCAS'97 and the VP-Region 10 of IEEE CAS Society in 1999 and 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, in 1997, the Senior Croucher Foundation Fellowship in 1999, the IEEE Third Millenium medal in 2000, and the Outstanding Researcher Award by The University of Hong Kong in 2003. He is a Fellow of the IEE and HKIE.