Training Signal Design and Tradeoffs for Spectrally-Efficient Multi-User MIMO-OFDM Systems

Yuejie Chi, *Student Member, IEEE,* Ahmad Gomaa, *Student Member, IEEE,* Naofal Al-Dhahir, *Fellow, IEEE,* and A. Robert Calderbank, *Fellow, IEEE*

***Abstract*—In this paper, we design MMSE-optimal training sequences for multi-user MIMO-OFDM systems with an arbitrary number of transmit antennas and an arbitrary number of training symbols. It addresses spectrally-efficient uplink transmission scenarios where the users overlap in time and frequency and are separated using spatial processing at the base station. The robustness of the proposed training sequences to residual carrier frequency offset and phase noise is evaluated. This analysis reveals an interesting design tradeoff between the peak-to-average power ratio of a training sequence and the increase in channel estimation mean squared error over the ideal case when these two impairments are not present.**

***Index Terms*—Training sequences design, Pilot design, MIMO-OFDM, Multi-user systems, Carrier frequency offset, Phase noise, RF impairments**

I. INTRODUCTION

Information-theoretic analysis by Foschini [1] and by Telatar [2] has shown that multiple antennas at the transmitter and receiver enable high-rate wireless communication. Space-time codes, introduced by Tarokh et al. [3], improve the reliability of communication over fading channels by correlating signals across different transmit antennas. Orthogonal Frequency Division Multiplexing (OFDM) [4] is widely adopted in broadband communications standards for its efficient implementation, high spectral efficiency, and robustness to Inter-Symbol Interference (ISI). OFDM offers great flexibility in that multiple streams with diverse rates and Quality-of-Service (QoS) requirements can be transmitted over the parallel frequency subchannels. However, there are two main drawbacks in OFDM; the first is high Peak-to-Average Power Ratio (PAPR) which results in larger backoff with nonlinear amplifiers, and the second is high sensitivity to frequency errors and phase noise. We will address both issues in this paper. Our focus is on training sequence design for the combination of Multiple-Input-Multiple-Output (MIMO) systems and OFDM technology (see [5] and references therein), and we aim to make this combination more attractive by reducing the overhead that is necessary for channel estimation.

Current multi-user MIMO-OFDM systems [6] support multiple users by assigning each time/frequency slot to only one user. For example, in OFDMA systems (adopted in the WiMAX [7] and LTE standards [8]), different users are assigned different subcarriers within the same OFDMA symbol. A different method of separating users is through the random-access CSMA/CA medium access control (MAC) protocol used in WLAN standards, e.g. IEEE 802.11n [9]. Both methods require that users not overlap in either time or frequency and this restriction results in a significant loss in spectral efficiency. The introduction of multiple receive antennas at the base station means that it is possible to improve spectral efficiency by allowing users to overlap while maintaining decodability, as in the recently-proposed Coordinated MultiPoint transmission (CoMP) techniques in the LTE-Advanced standard [10].

Accurate Channel State Information (CSI) is required at the receiver for coherent detection and is typically acquired by sending known training sequences from the transmit antennas and inferring channel parameters from the received signals. Various OFDM channel estimation schemes [11]-[13] have been proposed for Single-Input Single-Output (SISO) systems. However channel estimation is more challenging in a multi-user MIMO-OFDM system because there are more link parameters to calculate, and their estimation is complicated by interference between different transmissions. The direct approach is to invert a large matrix that describes cross-antenna interference at each OFDM tone [14]. Complexity can be reduced by exploiting the correlation between adjacent subchannels [15]. It is also possible to develop solutions in the time domain [16] where the challenge is to estimate time of arrivals. Here it is possible to reduce complexity by exploiting the power-delay profiles of the typical urban and hilly terrain propagation models. MIMO Channel estimation schemes were investigated in [17] for single-carrier single-user systems in the context of GSM-EDGE.

Linear Least-Squares (LLS) channel estimation is of great practical importance since it does not require prior knowledge of the channel statistics and enjoys low implementation complexity. We consider frequency-selective block-fading channels where the Time Domain (TD) representation requires fewer parameters than the Frequency Domain (FD) representation. Our focus is on the design of (optimal) training sequences for

Y. Chi is with the Department of Electrical Engineering, Princeton University, Princeton, NJ 08544, USA (email: ychi@princeton.edu).

A. Gomaa and N. Al-Dhahir are with the Department of Electrical Engineering, University of Texas at Dallas, Richardson, TX 75083, USA (e-mail: {aag083000, aldhahir}@utdallas.edu).

A. R. Calderbank is with the Department of Computer Science, Duke University, Durham, NC 27708, USA (email: robert.calderbank@duke.edu).

Multi-User MIMO OFDM systems that minimize the mean squared error of time-domain LLS channel estimation. The design of optimal training sequences for single-user MIMO-OFDM systems is investigated in [18] and [19]. The Fourier methods used in [18] provide some control over PAPR and some resilience to frequency offsets. The construction of optimal training sequences for multi-user MIMO-OFDM systems has been investigated in both the time domain [20] and the frequency domain [21], but these designs do not easily extend to multiple OFDM training symbols. It is also possible to take advantage of the similarities between communications and radar signal processing, where the path gains and delays are the range / Doppler coordinates of a scattering source and the problem is to estimate them. The unitary filter bank developed for Instantaneous Radar Polarimetry [22] supports frequency domain LLS channel estimation in a 2x2 MIMO OFDM system [23] and is able to suppress interference over two OFDM symbols with linear complexity. This example is a special case of a more general construction of filter banks for the analysis of acoustic surface waves [24], [25]. A limitation of these methods is that the number of OFDM training symbols is at least the number of transmit antennas.

In contrast, our framework supports the design of optimal training sequences for an arbitrary number of transmit antennas and an arbitrary number of training symbols. It provides the first general solution to the channel estimation problem for Multi-User MIMO-OFDM systems where Spatial Division Multiple Access (SDMA) is employed to increase the spectral efficiency. The optimality of our designs holds irrespective of the number of transmit antennas per user, the number of OFDM sub-carriers, the channel delay spread, and the number of users *provided* that the number of tones dedicated to estimation exceeds the product of the number of transmit antennas and the worst case delay spread. Not only does our design algorithm generate training sequences that minimize mean squared channel estimation error, but the designs have additional properties that make them very attractive from several implementation perspectives: 1) Individual training sequences can be drawn from standard signal constellations, 2) Low PAPR, and 3) Low channel estimation complexity without sacrificing optimality.

We start by considering the optimal training sequence design for uplink Multi-User MIMO-OFDM systems where all users are assumed to be synchronized. Then, we analyze the average performance degradation when the users are asynchronous, i.e. with residual Carrier Frequency Offsets (CFO). Next, we investigate the impact of Phase Noise (PN) perturbing transmit and receive oscillators on the channel estimation accuracy. This analysis leads to an interesting design tradeoff between the PAPR of a training sequence and its robustness to CFO and PN. The main contributions of this paper are

- Optimal training sequences design for Multi-User MIMO-OFDM systems with an *arbitrary* number of transmit antennas per user and an *arbitrary* number of training OFDM symbols as long as the rank condition (20) holds.
- Allowing users to overlap in time and frequency to increase the spectral efficiency.
- Analytical study of CFO and PN effects on the channel estimation performance for any training sequence taking into account PN at both the transmit and receive oscillators.
- Investigating the trade-off between the PAPR of the training sequence and its immunity against CFO and PN.

The rest of this paper is organized as follows. The uplink Multi-User MIMO-OFDM communication system model is described in Section II. The design of optimal training sequences is given for one and multiple training symbol senarios separately in Section III. Practical issues such as CFO and PN are discussed in Section IV. Design trade-offs are discussed in Section V. Simulation results are presented in Section VI. Finally, conclusions are drawn in Section VII.

A note on notation: We use boldface to denote matrices and vectors. For a matrix $\mathbf{A}$, $\mathbf{A}^T$ denotes its transpose, $\mathbf{A}^H$ denotes its complex-conjugate transpose, $\mathbf{A}^\dagger$ denotes its Penrose-Moore pseudo-inverse, $\mathbf{A}^{-1}$ denotes its inverse if it exists, and $\text{Tr}(\mathbf{A})$ denotes its trace. $\mathbf{I}_n$ denotes an identity matrix of dimension n and $\mathbf{0}_{m\times n}$ denotes an all-zero matrix of size $m \times n$. The notation $\text{diag}\,(x_1, x_2, \ldots, x_N)$ denotes an $N \times N$ diagonal matrix whose diagonal elements are $\{x_1, x_2, \ldots, x_N\}$. The operator $\otimes$ denotes the Kronecker product, and the operator $\circ$ denotes the entry-wise Hadamard product. We also summarize the key variables used throughout the paper in Table I.

II. System Model

We consider the uplink of a Multi-User MIMO-OFDM system, as shown in Fig. 1. We denote the Discrete Fourier Transform (DFT) size by N and the number of users by L ($L \geq 1$) where the ith user is equipped with M_i transmit antennas, $0 \leq i \leq L-1$. Therefore, the total number of transmit antennas among all users is given by $M = \sum_{i=0}^{L-1} M_i$.

We assume that the channel is quasi-static and remains constant over K successive OFDM training symbols. The channel from the jth transmit antenna of the ith user to the Base Station (BST) can be represented either in TD or FD. Let the Channel Frequency Response (CFR) be $\mathbf{H}_{i,j} = [H_{i,j}(0), \cdots, H_{i,j}(N-1)]^T$ where $H_{i,j}(k)$, $0 \leq k \leq N-1$, is the frequency response at the kth subcarrier. However, the Channel Impulse Response (CIR) in TD is represented by a much smaller number of parameters. We assume that the maximal memory over all CIRs is $\nu_{\max}$, and write the CIR as $\mathbf{h}_{i,j} = [h_{i,j}(0), \cdots, h_{i,j}(\nu_{\max})]^T$. Estimating the CIR instead of the CFR leads to the reduction of the number of unknowns from MN to $M(\nu_{\max}+1)$. Hence, a more accurate channel estimate is attainable using the same amount of training. Furthermore, the CFR can be reconstructed from the CIR as follows

$$H_{i,j}(k) = \frac{1}{\sqrt{N}} \sum_{t=0}^{\nu_{\max}} h_{i,j}(t) e^{-j\frac{2\pi}{N}tk}. \tag{1}$$

At the jth ($0 \leq j \leq M_i - 1$) transmit antenna of the ith ($0 \leq i \leq L-1$) user, an OFDM symbol $\mathbf{X}_{i,j}$ of size N is given by $\mathbf{X}_{i,j} = [X_{i,j}(0), \cdots, X_{i,j}(N-1)]^T$.

TABLE I
KEY VARIABLES USED THROUGHOUT THE PAPER.

Variable	Meaning	Domain
L	number of users	$\mathbb{Z}^+$
M_i	number of transmit antennas for the ith user	$\mathbb{Z}^+$
M	total number of transmit antennas	$\mathbb{Z}^+$
N	DFT size of the OFDM system	$\mathbb{Z}^+$
L_p	cyclic prefix length	$\mathbb{Z}^+$
$\nu_{\max}$	the maximal memory of all CIRs	$\mathbb{Z}^+$
$\mathbf{X}_{i,j}$	OFDM symbol from the jth antenna of the ith user	$\mathbb{C}^N$
$\mathbf{h}_{i,j}$	CIR from the jth antenna of the ith user	$\mathbb{C}^{\nu_{\max}+1}$
$\mathbf{H}_{i,j}$	CFR from the jth antenna of the ith user	$\mathbb{C}^N$
$\mathbf{F}$	DFT matrix of size N	$\mathbb{C}^{N\times N}$
$\mathbf{F}_0$	the first $(\nu_{\max}+1)$ columns of $\mathbf{F}$	$\mathbb{C}^{N\times(\nu_{\max}+1)}$
$\mathbf{\Lambda}_m$	the transform operator between training sequences	$\mathbb{C}^{N\times N}$
$\mathbf{h}_m$	CIR from the mth antenna	$\mathbb{C}^{\nu_{\max}+1}$
$\mathbf{X}_{tm}$	training sequence from the mth antenna in the tth OFDM symbol	$\mathbb{C}^N$
$\mathbf{S}_{tm}$	circulant training matrix from the mth antenna in the tth OFDM symbol	$\mathbb{C}^{N\times(\nu_{\max}+1)}$
$\mathbf{S}_t$	training matrix in the tth OFDM symbol	$\mathbb{C}^{N\times M(\nu_{\max}+1)}$
$\mathbf{y}_t$	received signal in the tth OFDM symbol	$\mathbb{C}^N$
α	normalized carrier frequency offset	$\mathbb{R}$
f_{sub}	subcarrier frequency spacing (Hz)	$\mathbb{R}^+$
β	two-sided 3-dB linewidth of the oscillator power spectrum density (Hz)	$\mathbb{R}^+$
ϕ_n	nth phase noise sample	$\mathbb{R}$

Let $\mathbf{x}_{i,j} = [x_{i,j}(0), x_{i,j}(1), \cdots, x_{i,j}(N-1)]^T$ be the Inverse Discrete Fourier Transform (IDFT) of $\mathbf{X}_{i,j}$. We use a Cyclic-Prefix (CP) of length L_p for the guard interval in the OFDM system so that

$$\tilde{\mathbf{x}}_{i,j} = [x(N-L_p+1), \cdots, x(N-1), x(0), \cdots, x(N-1)]^T \tag{2}$$

where L_p is chosen to be greater than the channel memory, i.e. $L_p \geq (\nu_{\max}+1)$. Finally, $\tilde{\mathbf{x}}_{i,j}$ goes through Parallel-to-Serial (P/S) conversion and is modulated to the carrier frequency for transmission.

At the base station, all users are assumed to be in frequency synchronization with the BST. In Section IV, we examine the robustness of our proposed optimal training sequence design when this condition is not satisfied. In addition, we assume that all users are synchronized in time with the BST, where the received signal is down-converted to baseband and passed through a Serial-to-Parallel (S/P) converter. Then, the CP is removed and the Fast Fourier Transform (FFT) is applied. The received OFDM symbol $\mathbf{Y} = [Y(0), \cdots, Y(N-1)]^T$ in one symbol time can be written as

$$\mathbf{Y} = \sum_{i=0}^{L-1}\sum_{j=0}^{M_i-1} \text{diag}(\mathbf{H}_{i,j})\mathbf{X}_{i,j} + \mathbf{N}, \tag{3}$$

where $\mathbf{N} \sim \mathcal{N}(\mathbf{0}_{N\times 1}, \sigma^2\mathbf{I}_N)$ is Additive White Gaussian Noise (AWGN). We consider the mapping $(i,j) \mapsto m : m = \sum_{s=0}^{i} M_s + j - M_i$, $0 \leq m \leq M-1$, and re-label $\mathbf{H}_{i,j}$ and $\mathbf{X}_{i,j}$ as $\mathbf{H}_m$ and $\mathbf{X}_m$, respectively. The label can be inverted easily as

$$i = \underset{0\leq i^*\leq L-1}{\text{argmin}}\ i^* \text{ s.t. } m \leq \sum_{s=0}^{i^*} M_s,\ j = m + M_i - \sum_{s=0}^{i} M_s. \tag{4}$$

Then, equation (3) can be written as

$$\mathbf{Y} = \sum_{m=0}^{M-1} \text{diag}(\mathbf{H}_m)\mathbf{X}_m + \mathbf{N}. \tag{5}$$

Remark: The development of the algorithm requires labeling of transmit antennas among all users, and that both the BST and all the users are aware of that labeling.

In Section III, we first consider TD LLS channel estimation when only one training symbol is allowed by leveraging the channel representation in TD. A general approach when $K \geq 2$ training symbol is given by further incorporating space-time code structure into the design. Then, a special construction utilizing Quaternions is given when $K = 2$. Finally, an alternative scheme using equally-spaced pilots instead of the whole symbol for training is given under some mild conditions. In the following sections, we assume one receive antenna, since the same channel estimation scheme can be applied at all receive antennas without loss of generality.

III. MAIN RESULTS

A. One OFDM Training Symbol

Since there are fewer parameters to be estimated in the TD, we apply the IDFT of size N to Eq. (5), and get

$$\begin{aligned}\mathbf{y} &= \sum_{m=0}^{M-1} \mathbf{S}_m\mathbf{h}_m + \mathbf{n} \\ &= \begin{bmatrix}\mathbf{S}_0 & \mathbf{S}_1 & \cdots & \mathbf{S}_{M-1}\end{bmatrix}\begin{bmatrix}\mathbf{h}_0^H & \mathbf{h}_1^H & \ldots & \mathbf{h}_{M-1}^H\end{bmatrix}^H + \mathbf{n} \\ &\triangleq \mathbf{S}\mathbf{h} + \mathbf{n}\end{aligned} \tag{6}$$

where $\mathbf{y} \in \mathbb{C}^N$, $\mathbf{h}_m \in \mathbb{C}^{\nu_{\max}+1}$, $0 \leq m \leq M-1$, and $\mathbf{S}_m \in \mathbb{C}^{N\times(\nu_{\max}+1)}$ is the circulant training matrix constructed from the corresponding training sequence transmitted over the mth antenna.

Let $\mathbf{F} = [\mathbf{f}_0, \cdots, \mathbf{f}_{N-1}]$ be the DFT matrix of size N with $\mathbf{f}_i$ denoting its i^{th} column, and let $\mathbf{F}_0 = [\mathbf{f}_0, \cdots, \mathbf{f}_{\nu_{\max}}]$ be

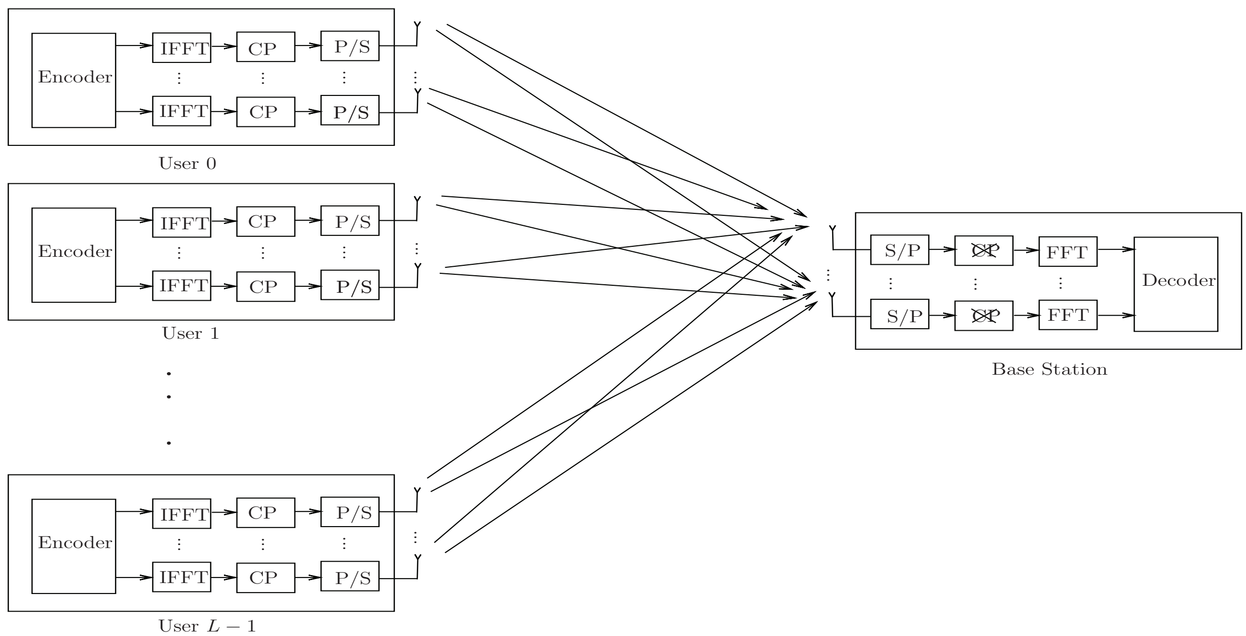


Fig. 1. The uplink of a Multi-User MIMO-OFDM communication system.

composed of the first $(\nu_{\max}+1)$ columns of $\mathbf{F}$. Then, $\mathbf{S}_m$ can be written as

$$\mathbf{S}_m = \mathbf{F}^H \mathbf{D}_m \mathbf{F}_0, \tag{7}$$

where $\mathbf{D}_m = \text{diag}\left(X_m(0), \cdots, X_m(N-1)\right)$. The matrix $\mathbf{S} \in \mathbb{C}^{N \times M(\nu_{\max}+1)}$ defined in Eq. (6) is formed by horizontally concatenating the matrices $\mathbf{S}_m$, $0 \leq m \leq M-1$. To enable LLS channel estimation, the following condition on dimensionality has to be satisfied [26]

$$N \geq M(\nu_{\max}+1) \quad \text{or,} \quad M \leq \frac{N}{(\nu_{\max}+1)}. \tag{8}$$

To minimize the variance of the channel estimation error, the matrix $\mathbf{S}$ is required to satisfy [26]

$$\mathbf{S}^H \mathbf{S} = c\mathbf{I}_{M(\nu_{\max}+1)} \tag{9}$$

and this requires that

$$\mathbf{S}_m^H \mathbf{S}_n = c\delta_{mn} \mathbf{I}_{(\nu_{\max}+1)}, \; 0 \leq m, n \leq M-1. \tag{10}$$

Given (7), the optimality condition becomes

$$\mathbf{F}_0^H \mathbf{D}_m^H \mathbf{D}_n \mathbf{F}_0 = c\delta_{mn} \mathbf{I}_{(\nu_{\max}+1)}, \; 0 \leq m, n \leq M-1. \tag{11}$$

Next, let $\mathbf{F}_m$ be composed of $(\nu_{max}+1)$ consecutive columns of $\mathbf{F}$ starting at index $m(\nu_{\max}+1)$, i.e.

$$\begin{aligned} \mathbf{F}_m &= \left[\mathbf{f}_{m(\nu_{\max}+1)}, \cdots, \mathbf{f}_{(m+1)(\nu_{\max}+1)-1}\right] \\ &= \mathbf{\Lambda}_m \mathbf{F}_0, \quad 0 \leq m \leq M-1, \end{aligned} \tag{12}$$

where

$$\mathbf{\Lambda}_m = \text{diag}\left(1, e^{j\frac{2\pi(\nu_{\max}+1)}{N}m}, \cdots, e^{j\frac{2\pi(\nu_{\max}+1)(N-1)}{N}m}\right). \tag{13}$$

It can be easily shown that $\mathbf{F}_m^H \mathbf{F}_n = \delta_{mn} \mathbf{I}_{(\nu_{\max}+1)}$. Now we present a general approach which gives a family of optimal training sequences. As a starting point, we choose the FD training sequence as an arbitrary constant-amplitude sequence $\mathbf{X}$. Let $\mathbf{D} = \text{diag}\left(X(0), \cdots, X(N-1)\right)$, then $\mathbf{D}^H \mathbf{D} = c\mathbf{I}_N$ where c is determined by the signal constellation and/or transmit power constraints. The FD training sequence at the mth transmit antenna is given by

$$\mathbf{X}_m = \mathbf{\Lambda}_m \mathbf{X}, \;\; 0 \leq m \leq M-1. \tag{14}$$

Equivalently, $\mathbf{D}_m = \mathbf{\Lambda}_m \mathbf{D} = \mathbf{D}\mathbf{\Lambda}_m$, $0 \leq m \leq M-1$. Furthermore, we have the following theorem.

Theorem 1: The choice of FD training sequences in Eq. (14) is optimal for a single training OFDM symbol.

Proof: It is enough to show that Eq. (10) holds. Since

$$\mathbf{S}_m = \mathbf{F}^H \mathbf{D}_m \mathbf{F}_0 = \mathbf{F}^H \mathbf{D} \mathbf{\Lambda}_m \mathbf{F}_0 = \mathbf{F}^H \mathbf{D} \mathbf{F}_m, \tag{15}$$

it follows that

$$\begin{aligned} \mathbf{S}_m^H \mathbf{S}_n &= (\mathbf{F}^H \mathbf{D} \mathbf{F}_m)^H \mathbf{F}^H \mathbf{D} \mathbf{F}_n \\ &= \mathbf{F}_m^H \mathbf{D}^H \mathbf{D} \mathbf{F}_n = c\delta_{mn} \mathbf{I}_{(\nu_{\max}+1)}. \end{aligned} \tag{16}$$

Therefore, the LLS estimate (LLSE) of $\mathbf{h}$ is given as $\hat{\mathbf{h}} = \frac{1}{c}\mathbf{S}^H \mathbf{y}$, where each CIR can be estimated as $\hat{\mathbf{h}}_m = \frac{1}{c}\mathbf{S}_m^H \mathbf{y}$. Then, the CFR is given by

$$\hat{\mathbf{H}}_m = \frac{1}{c}\mathbf{F}\mathbf{S}_m^H \mathbf{y} = \frac{1}{c}(\mathbf{F}\mathbf{F}_0^H)\mathbf{D}_m^H \mathbf{F}\mathbf{y}, \;\; 0 \leq m \leq M-1. \tag{17}$$

The resulting channel estimation error variance is given by

$$\sigma_e^2 = \sigma^2 \text{Tr}\left((\mathbf{S}^H \mathbf{S})^{-1}\right) = \frac{M(\nu_{\max}+1)}{c}\sigma^2. \tag{18}$$

B. K OFDM Training Symbols with $K \geq 2$

The major limitation of using only one training OFDM symbol is that the total number of transmit antennas is limited by $\frac{N}{(\nu_{\max}+1)}$. When the channel is quasi-static over $K \geq 2$ OFDM training symbols it is possible to increase the number of admissable transmit antennas and reduce MMSE by a factor of K.

Denoting the received TD OFDM symbol in the tth symbol time by $\mathbf{y}_t$, $0 \le t \le K-1$, we express the received symbol block $\mathbf{y}$ in Eq. 19 where $\mathbf{S}_{tm} = \mathbf{F}^H \mathbf{D}_{tm} \mathbf{F}_0$, and the matrices $\mathbf{D}_{tm}$'s are diagonal with the FD training sequences appearing on their main diagonals. Least-square estimation is possible when the following dimensionality condition for the matrix $\mathbf{S} \in \mathbb{C}^{KN \times M(\nu_{\max}+1)}$ holds

$$KN \ge M(\nu_{\max}+1), \text{ or, } \sum_{i=0}^{L-1} M_i = M \le \frac{KN}{(\nu_{\max}+1)}. \quad (20)$$

For $\mathbf{S}$ to be optimal, it has to satisfy $\mathbf{S}^H\mathbf{S} = \tilde{c}\mathbf{I}_{M(\nu_{\max}+1)}$ for some $\tilde{c}$. We extend our previous approach by constructing a unitary matrix of higher dimension with the space-time code structure. Let the matrix $\mathbf{\Sigma} \in \mathbb{C}^{KN \times KN}$ be constructed as a Kronecker product $\mathbf{\Sigma} = \mathbf{U} \otimes \mathbf{V}$ where $\mathbf{U} = [U_{tq}] \in \mathbb{C}^{K \times K}$ is a unitary matrix and $\mathbf{V} \in \mathbb{C}^{N \times N}$ is a diagonal matrix satisfying $\mathbf{V}^H\mathbf{V} = \tilde{c}\mathbf{I}_N$. Therefore the matrix $\mathbf{\Sigma}$ satisfies

$$\mathbf{\Sigma}^H\mathbf{\Sigma} = \mathbf{U}^H\mathbf{U} \otimes \mathbf{V}^H\mathbf{V} = \tilde{c}\mathbf{I}_{KN}. \quad (21)$$

We give the following general design of optimal training sequences. For $0 \le m \le M-1$, let $p = \left\lfloor \frac{m}{K} \right\rfloor$, $0 \le p \le \left\lfloor \frac{M-1}{K} \right\rfloor$ and $q = m - Kp \in \{0, \cdots, K-1\}$. For the mth transmit antenna, its FD training sequence matrix at the tth OFDM training symbol is given by

$$\mathbf{D}_{tm} = \mathbf{\Sigma}_{tq}\mathbf{\Lambda}_p, \quad \text{if} \quad m = Kp+q, \ 0 \le m \le M-1, \quad (22)$$

where $\mathbf{\Sigma}_{tq} = U_{tq}\mathbf{V}$ is the $N \times N$ diagonal matrix located at the (t,q) block of $\mathbf{\Sigma}$.

The bijection $\pi : m \mapsto \{p, q\}$ groups the antennas into K classes depending on the equivalence of the residue q. For two antennas *not* in the same class, their training sequences can be proved orthogonal over *any* OFDM training symbol. For two antennas in the same class, their training sequences can be proved orthogonal over *all* K OFDM training symbols. We give the detailed proof below.

Theorem 2: The training sequences in (22) are optimal for K training OFDM symbols.

Proof: It is enough to show that

$$\sum_{t=0}^{K-1} \mathbf{S}_{tm}^H\mathbf{S}_{tn} = \mathbf{F}_0^H \left(\sum_{t=0}^{K-1} \mathbf{D}_{tm}^H\mathbf{D}_{tn} \right) \mathbf{F}_0 = \tilde{c}\delta_{mn}\mathbf{I}_{(\nu_{\max}+1)} \quad (23)$$

It is obvious that when $m = n$, the above equation holds. When $m \ne n$, we write $m = Kp_1 + q_1$ and $n = Kp_2 + q_2$ and split the proof into two cases:

- $q_1 = q_2 = q \in \{0, \cdots, K-1\}$ but $p_1 \ne p_2$. Then,

$$\mathbf{D}_{tm} = \mathbf{\Sigma}_{tq}\mathbf{\Lambda}_{p_1} \quad \text{and} \quad \mathbf{D}_{tn} = \mathbf{\Sigma}_{tq}\mathbf{\Lambda}_{p_2}, \quad (24)$$

for $0 \le t \le K-1$, therefore,

$$\sum_{t=0}^{K-1} \mathbf{S}_{tm}^H\mathbf{S}_{tn} = \mathbf{F}_{p_1}^H \left(\sum_{t=0}^{K-1} \mathbf{\Sigma}_{tq}^H\mathbf{\Sigma}_{tq} \right) \mathbf{F}_{p_2} = \tilde{c}\mathbf{F}_{p_1}^H\mathbf{F}_{p_2} = \mathbf{0}_{(\nu_{\max}+1)}.$$

TABLE II
FD TRAINING SEQUENCES AT THE mTH TRANSMIT ANTENNA WHEN TWO TRAINING OFDM SYMBOLS ARE AVAILABLE.

$m = 2p+q$	$q = 0$	$q = 1$
0th symbol	$\mathbf{X}_p$	$-\mathbf{Z}_p^H$
1st symbol	$\mathbf{Z}_p$	$\mathbf{X}_p^H$

- $q_1 \ne q_2$. Then from Eq. (24), we have

$$\sum_{t=0}^{K-1} \mathbf{D}_{tm}^H\mathbf{D}_{tn} = \mathbf{\Lambda}_{p_1}^H \left(\sum_{t=0}^{K-1} \mathbf{\Sigma}_{t,q_1}^H\mathbf{\Sigma}_{t,q_2} \right) \mathbf{\Lambda}_{p_2} = \mathbf{0}_N. \quad (25)$$

Now, Eq. (23) follows trivially. ■

Finally, the LLSE of $\mathbf{h}$ is given by $\hat{\mathbf{h}} = \frac{1}{\tilde{c}}\sum_{t=0}^{K-1}\mathbf{S}_t^H\mathbf{y}_t$ where each CIR can be estimated as $\hat{\mathbf{h}}_m = \frac{1}{\tilde{c}}\sum_{t=0}^{K-1}\mathbf{S}_{tm}^H\mathbf{y}_t$. Then, the CFR at the mth transmit antenna is given by

$$\hat{\mathbf{H}}_m = \frac{1}{\tilde{c}}\mathbf{F}\hat{\mathbf{h}}_m = \frac{1}{\tilde{c}}(\mathbf{F}\mathbf{F}_0^H)\sum_{t=0}^{K-1}\mathbf{D}_{tm}^H\mathbf{F}\mathbf{y}_t. \quad (26)$$

Let $\tilde{c} = Kc$, the resulting channel estimation error variance is given by

$$\sigma_e^2 = \sigma^2 \mathrm{Tr}\left(\sum_{t=0}^{K-1} \mathbf{S}_t^H\mathbf{S}_t)^{-1} \right) = \frac{M(\nu_{\max}+1)}{Kc}\sigma^2.$$

C. Special case when $K = 2$

When $K = 2$, like the Alamouti Space-Time Block Code (STBC), our construction of training sequences makes use of Hamilton's Biquaternions. We will choose two FD training sequences $\mathbf{X}$ and $\mathbf{Z}$ where the sum of their squared amplitudes is constant, i.e.

$$\mathbf{D}_X^H\mathbf{D}_X + \mathbf{D}_Z^H\mathbf{D}_Z = \tilde{c}\mathbf{I}_N. \quad (27)$$

where $\mathbf{D}_X = \mathrm{diag}\,(X(0), \cdots, X(N-1))$, and $\mathbf{D}_Z = \mathrm{diag}\,(Z(0), \cdots, Z(N-1))$.

For $0 \le m \le M-1$, let $p = \left\lfloor \frac{m}{2} \right\rfloor$, $0 \le p \le \left\lfloor \frac{M-1}{2} \right\rfloor$ and $q = m - 2p \in \{0, 1\}$. Let $\mathbf{X}_p = \mathbf{\Lambda}_p\mathbf{X}$ and $\mathbf{Z}_p = \mathbf{\Lambda}_p\mathbf{Z}$, $0 \le p \le \left\lfloor \frac{M-1}{2} \right\rfloor$, where $\mathbf{\Lambda}_p$ is defined in Eq. (13).

The diagonal FD training matrices of the mth antenna in the 0th and 1st training symbols are given by $\mathbf{D}_{0m}$ and $\mathbf{D}_{1m}$ respectively:

$$\mathbf{D}_{0m} = \begin{cases} \mathbf{\Lambda}_p\mathbf{D}_X, & \text{if } q = 0, m = 2p \\ \mathbf{\Lambda}_p^H\mathbf{D}_Z^H, & \text{if } q = 1, m = 2p+1 \end{cases}, \quad (28)$$

$$\text{and } \mathbf{D}_{1m} = \begin{cases} \mathbf{\Lambda}_p\mathbf{D}_Z, & \text{if } q = 0, m = 2p \\ -\mathbf{\Lambda}_p^H\mathbf{D}_X^H, & \text{if } q = 1, m = 2p+1 \end{cases}. \quad (29)$$

We summarize the FD training sequences design in Table II.

Theorem 3: The FD training sequences in Table II are optimal for two training OFDM symbols.

Proof: It is enough to show that

$$\mathbf{S}_{0m}^H\mathbf{S}_{0n} + \mathbf{S}_{1m}^H\mathbf{S}_{1n} = \mathbf{F}_0^H(\mathbf{D}_{0m}^H\mathbf{D}_{0n} + \mathbf{D}_{1m}^H\mathbf{D}_{1n})\mathbf{F}_0 = \tilde{c}\delta_{mn}\mathbf{I}_{(\nu_{\max}+1)}. \quad (30)$$

It is obvious that the above equation holds when $m = n$. When $m \ne n$, we write $m = 2p_1 + q_1$ and $n = 2p_2 + q_2$ and consider two cases

$$\mathbf{y}=\begin{bmatrix}\mathbf{y}_0\\ \mathbf{y}_1\\ \vdots\\ \mathbf{y}_{K-1}\end{bmatrix}=\begin{bmatrix}\mathbf{S}_{00} & \mathbf{S}_{01} & \cdots & \mathbf{S}_{0,M-1}\\ \mathbf{S}_{10} & \mathbf{S}_{11} & \cdots & \mathbf{S}_{1,M-1}\\ \vdots & \vdots & \ddots & \vdots\\ \mathbf{S}_{K-1,0} & \mathbf{S}_{K-1,1} & \cdots & \mathbf{S}_{K-1,M-1}\end{bmatrix}\begin{bmatrix}\mathbf{h}_0\\ \mathbf{h}_1\\ \vdots\\ \mathbf{h}_{M-1}\end{bmatrix}+\mathbf{n}\triangleq\begin{bmatrix}\mathbf{S}_0\\ \mathbf{S}_1\\ \vdots\\ \mathbf{S}_{K-1}\end{bmatrix}\mathbf{h}+\mathbf{n}=\mathbf{S}\mathbf{h}+\mathbf{n} \tag{19}$$

- $q_1 = q_2$ but $p_1 \neq p_2$. Without loss of generality, we assume $q_1 = q_2 = 0$ and get

$$\mathbf{S}_{0m}^H\mathbf{S}_{0n}=\mathbf{F}_0^H\mathbf{D}_{0m}^H\mathbf{D}_{0n}\mathbf{F}_0=\mathbf{F}_{p_1}^H\mathbf{D}_X^H\mathbf{D}_X\mathbf{F}_{p_2},$$
$$\text{and}\quad \mathbf{S}_{1m}^H\mathbf{S}_{1n}=\mathbf{F}_0^H\mathbf{D}_{1m}^H\mathbf{D}_{1n}\mathbf{F}_0=\mathbf{F}_{p_1}^H\mathbf{D}_Z^H\mathbf{D}_Z\mathbf{F}_{p_2}$$

Hence,

$$\mathbf{S}_{0m}^H\mathbf{S}_{0n}+\mathbf{S}_{1m}^H\mathbf{S}_{1n}=\mathbf{F}_{p_1}^H\left(\mathbf{D}_X^H\mathbf{D}_X+\mathbf{D}_Z^H\mathbf{D}_Z\right)\mathbf{F}_{p_2}$$
$$=\tilde{c}\mathbf{F}_{p_1}^H\mathbf{F}_{p_2}=\mathbf{0}_{(\nu_{\max}+1)}.$$

- $q_1 \neq q_2$. Without loss of generality, we assume $q_1 = 0$ and $q_2 = 1$ and write

$$\mathbf{D}_{0m}^H\mathbf{D}_{0n}+\mathbf{D}_{1m}^H\mathbf{D}_{1n}$$
$$=(\mathbf{\Lambda}_{p_1}\mathbf{D}_X)^H(\mathbf{\Lambda}_{p_2}^H\mathbf{D}_Z^H)+(\mathbf{\Lambda}_{p_1}\mathbf{D}_Z)^H(-\mathbf{\Lambda}_{p_2}^H\mathbf{D}_X^H)$$
$$=\mathbf{\Lambda}_{p_1}^H(\mathbf{D}_X^H\mathbf{D}_Z^H-\mathbf{D}_Z^H\mathbf{D}_X^H)\mathbf{\Lambda}_{p_2}^H=\mathbf{0}_{(\nu_{\max}+1)}.$$

Eq. (30) follows directly. ■

If all the users employ two transmit antennas and Alamouti code, their training sequences in two symbol intervals are assigned according to Eq. (28) and (29) , and can be generated simply using the same Alamouti code generator which greatly reduce the training assignment complexity.

D. Peak-to-Average Power Ratio (PAPR) Property

The PAPR of the training sequence $S(n),\ 0 \le n \le N-1$, is given by

$$\text{PAPR}=\frac{\max_n |S(n)|^2}{\frac{1}{N}\sum_{n=0}^{N-1}|S(n)|^2}. \tag{31}$$

The transform operator $\mathbf{\Lambda}_m$ between different FD training sequences can be viewed as a frequency modulation, which is equivalent to circulant shift of the training sequence in the TD. Hence, we have the following proposition.

Proposition 1: All TD training sequences of $\mathbf{X}_m$'s in (14) have the same PAPR.

This property is important when designing the training sequences. As long as the PAPR of $\mathbf{X}$ is low, all training symbols will have low PAPR. Another merit of our design is that if $\frac{N}{(\nu_{\max}+1)} = 2^k$, for some integer k, and if $\mathbf{X}$ is chosen from a 2^k-phase shift keying (PSK) constellation, then the transform $\mathbf{\Lambda}_m$ guarantees that all FD training sequences $\{\mathbf{X}_m,\ 0 \le m \le M-1\}$ will belong to the same 2^k-PSK constellation, which is very easy to generate.

One possible choice for $\mathbf{X}$ is a Constant-Amplitude-Zero-Auto-Correlation (CAZAC) sequence [27], which is a complex-valued sequence with constant amplitude and zero autocorrelation at nonzero lags. One example of a CAZAC sequence of length N is the chirp sequence given by

$$X(k)=\begin{cases}\sqrt{c}\exp(j\frac{\pi u k^2}{N}), & \text{if } N \text{ is even}\\ \sqrt{c}\exp(j\frac{\pi u k(k+1)}{N}), & \text{if } N \text{ is odd}\end{cases},0\le k\le N-1. \tag{32}$$

where u is any integer relatively prime[1] to N. A disadvantage of this and other CAZAC sequences is that the entries are not restricted to a standard signal constellation. An alternative is provided by Golay complementary sequences [28] which only assume values from $\{-\sqrt{c}, \sqrt{c}\}$. A third possibility is the flat sequence (impulsive in TD) $\{\mathbf{X}:\ X(k)=\sqrt{c},\ \text{for all } k\}$. These three choices have different PAPRs, as summarized in Table III and perform differently under practical system impairments as will be discussed in Sections IV and VI. Given the above discussion, it is possible to generate a family of optimal training sequences with low PAPR from a standard signal constellation.

TABLE III
PAPR COMPARISON OF THREE TRAINING SEQUENCE CANDIDATES

	Chirp-based	Golay-based	TD Impulsive
PAPR	0 dB	$\le$ 3 dB	18 dB

E. Reducing the number of pilots per training symbol

Assume the number of subcarriers N can be decomposed as $N = N_pT$ where $N_p \ge (\nu_{\max}+1)$, then it is possible to use N_p equally-spaced pilots in each training symbol instead of the whole symbol. At the mth antenna, the training sequence is given by diagonal matrix $\hat{\mathbf{D}}_m \in \mathbb{C}^{N_p\times N_p}$ and the pilot locations are $\{sT\}_{s=0}^{N_p-1}$. Consider the one training symbol scenario without loss of generality. Instead of taking IDFT of Eq. (5) of length N, we now take IDFT of Eq. (5) *only at pilot tones* of length N_p, and get

$$\hat{\mathbf{y}}=\sum_{m=0}^{M-1}\hat{\mathbf{S}}_m\mathbf{h}_m+\mathbf{n}, \tag{33}$$

where $\hat{\mathbf{y}} \in \mathbb{C}^{N_p}$, $\hat{\mathbf{S}}_m = \hat{\mathbf{F}}^H\hat{\mathbf{D}}_m\hat{\mathbf{F}}_0 \in \mathbb{C}^{N_p\times(\nu_{\max}+1)}$, $\hat{\mathbf{F}}$ is the DFT matrix of size N_p, $\hat{\mathbf{F}}_0 = [\hat{f}_{st} = e^{j\frac{2\pi st}{N_p}}] \in \mathbb{C}^{N_p\times(\nu_{\max}+1)}$ is the submatrix of $\mathbf{F}_0$ at rows corresponding to pilot frequencies. It is obvious that $\hat{\mathbf{F}}_0$ is also the first $(\nu_{\max}+1)$ columns of the DFT matrix $\hat{\mathbf{F}}$, and $\hat{\mathbf{F}}_0^H\hat{\mathbf{F}}_0 = \frac{1}{T}\mathbf{I}_{\nu_{\max}+1}$. Therefore, it is clear that we can follow the same framework in both single and multiple training symbol scenarios, by replacing N by N_p in both the dimensionality conditions and design parameters at the cost of increasing the MMSE by a factor of T.

[1]Two integers are said to be relatively prime if their greatest common divisor is 1.

IV. Practical Issues

In this section, we study the performance of channel estimation using our proposed optimal training sequences under two practical impairments, namely, Carrier Frequency Offset (CFO) and oscillator Phase Noise (PN).

A. Residual CFO

In practical systems, CFO is first estimated and compensated for prior to channel estimation; however, a residual CFO remains uncompensated for due to the inaccuracy of the CFO estimate. In the sequel, we derive the channel estimation MSE in the presence of a residual CFO for the Multi-User MIMO-OFDM system. Taking the residual CFOs into consideration, the received signal over K training OFDM symbols is

$$\begin{aligned}\mathbf{y} &= \sum_{i=0}^{L-1} \mathbf{Q}_i \mathcal{S}_i \mathcal{h}_i + \mathbf{n} \\ &= \begin{bmatrix}\mathbf{Q}_0\mathcal{S}_0 & \mathbf{Q}_1\mathcal{S}_1 & \dots & \mathbf{Q}_{L-1}\mathcal{S}_{L-1}\end{bmatrix}\mathbf{h} + \mathbf{n} \\ &\triangleq \widetilde{\mathbf{S}}\mathbf{h} + \mathbf{n} \end{aligned} \tag{34}$$

where $\mathcal{S}_i$ and $\mathcal{h}_i$ concatenate the training matrices and the CIRs, respectively, of all the antennas used by the ith user and

$$\mathbf{Q}_i = \text{diag}\left\{\mathbf{d}, e^{\frac{j2\pi(N+\nu+1)\alpha_i}{N}}\mathbf{d}, \dots, e^{\frac{j2\pi(K-1)(N+\nu+1)\alpha_i}{N}}\mathbf{d}\right\}$$

where $\mathbf{d} = \left[1, e^{\frac{j2\pi\alpha_i}{N}}, \dots, e^{\frac{j2\pi(N-1)\alpha_i}{N}}\right]^T$. Furthermore, α_i is a random variable representing the normalized frequency offset between the ith user carrier frequency, f_{T}^i, and the receiver carrier frequency, f_{R}, defined as

$$\alpha_i \triangleq \frac{f_{\text{T}}^i - f_{\text{R}}}{f_{\text{sub}}}, \tag{35}$$

where f_{sub} denotes the subcarrier frequency spacing. Assuming the dimensionality condition in (20) is satisfied, the LLSE of $\mathbf{h}$ is

$$\hat{\mathbf{h}} = \frac{1}{\tilde{c}}\underbrace{\begin{bmatrix}\mathcal{S}_0 & \mathcal{S}_1 & \dots & \mathcal{S}_{L-1}\end{bmatrix}^H}_{=\mathbf{S}^H}\mathbf{y} = \frac{1}{\tilde{c}}\mathbf{S}^H\widetilde{\mathbf{S}}\mathbf{h} + \underbrace{\frac{1}{\tilde{c}}\mathbf{S}^H\mathbf{n}}_{\triangleq\tilde{\mathbf{n}}}. \tag{36}$$

Writing $\mathbf{Q}_i = \mathbf{I}_N + (\mathbf{Q}_i - \mathbf{I}_N) \triangleq \mathbf{I}_N + \widetilde{\mathbf{Q}}_i$, we express $\hat{\mathbf{h}}$ as

$$\hat{\mathbf{h}} = \mathbf{h} + \mathbf{S}_\Delta\mathbf{h} + \tilde{\mathbf{n}} \tag{37}$$

where

$$\mathbf{S}_\Delta = \frac{1}{\tilde{c}}\mathbf{S}^H\begin{bmatrix}\widetilde{\mathbf{Q}}_0\mathcal{S}_0 & \widetilde{\mathbf{Q}}_1\mathcal{S}_1 & \dots & \widetilde{\mathbf{Q}}_{L-1}\mathcal{S}_{L-1}\end{bmatrix}. \tag{38}$$

The trace of the error auto-correlation matrix is given by

$$\begin{aligned} t_e &= \text{Tr}\left(\mathbb{E}\left[\left(\mathbf{h}-\hat{\mathbf{h}}\right)\left(\mathbf{h}-\hat{\mathbf{h}}\right)^H\right]\right) \\ &= \text{Tr}\left(\mathbb{E}\left[\mathbf{S}_\Delta\mathbf{h}\mathbf{h}^H\mathbf{S}_\Delta^H\right]\right) + 2\sigma_e^2 \triangleq t_{\text{fo}} + 2\sigma_e^2, \end{aligned} \tag{39}$$

where $\mathbb{E}[\cdot]$ denotes the statistical expectation. For any two matrices $\mathbf{A}$ and $\mathbf{B}$, we know that

$$\text{Tr}(\mathbb{E}[\mathbf{A}]) = \mathbb{E}[\text{Tr}(\mathbf{A})] \quad \text{and} \quad \text{Tr}(\mathbf{AB}) = \text{Tr}(\mathbf{BA}). \tag{40}$$

Using these properties and assuming that the CIR coefficients and the normalized frequency offsets are statistically independent, we write

$$\begin{aligned} t_{\text{fo}} &= \mathbb{E}\left[\text{Tr}\left(\mathbf{h}\mathbf{h}^H\mathbf{S}_\Delta^H\mathbf{S}_\Delta\right)\right] \\ &= \text{Tr}\left(\mathbb{E}\left[\mathbf{h}\mathbf{h}^H\right]\mathbb{E}\left[\mathbf{S}_\Delta^H\mathbf{S}_\Delta\right]\right) = \text{Tr}\left(\mathbf{C}_\mathbf{h}\mathbf{C}_{\mathbf{S}_\Delta}\right) \end{aligned} \tag{41}$$

where $\mathbf{C}_\mathbf{h} \triangleq \mathbb{E}\left[\mathbf{h}\mathbf{h}^H\right]$ and $\mathbf{C}_{\mathbf{S}_\Delta} \triangleq \mathbb{E}\left[\mathbf{S}_\Delta^H\mathbf{S}_\Delta\right]$. The (i,j)th block matrix of $\mathbf{C}_{\mathbf{S}_\Delta}$ is given by

$$\begin{aligned} \mathbf{C}_{\mathbf{S}_\Delta}(i,j) &= \frac{1}{\tilde{c}^2}\mathbb{E}\left[\mathcal{S}_i^H\widetilde{\mathbf{Q}}_i^H\left(\sum_{k=0}^{L-1}\mathcal{S}_k\mathcal{S}_k^H\right)\widetilde{\mathbf{Q}}_j\mathcal{S}_j\right] \\ &= \frac{1}{\tilde{c}^2}\mathbb{E}\left[\mathcal{S}_i^H\widetilde{\mathbf{Q}}_i^H\mathbf{S}\mathbf{S}^H\widetilde{\mathbf{Q}}_j\mathcal{S}_j\right], \end{aligned} \tag{42}$$

where $0 \le i,j \le L-1$. Since $\widetilde{\mathbf{Q}}_i$ and $\widetilde{\mathbf{Q}}_j$ are diagonal, we express $\mathbf{C}_{\mathbf{S}_\Delta}(i,j)$ as follows

$$\mathbf{C}_{\mathbf{S}_\Delta}(i,j) = \frac{1}{\tilde{c}^2}\mathcal{S}_i^H\left(\left[\mathbf{S}\mathbf{S}^H\right]\circ\underbrace{\mathbb{E}\left[\mathbf{v}_i^*\mathbf{v}_j^T\right]}_{\triangleq\mathbf{C}_\mathbf{v}^{i,j}}\right)\mathcal{S}_j \tag{43}$$

where $\mathbf{v}_i$ and $\mathbf{v}_j$ are columns vectors containing the diagonal elements of $\widetilde{\mathbf{Q}}_i$ and $\widetilde{\mathbf{Q}}_j$, respectively.

Given the second-order statistics of the residual CFOs and the CIR coefficients, we can easily compute $\mathbf{C}_{\mathbf{S}_\Delta}$ and $\mathbf{C}_\mathbf{h}$, and hence t_e for any training sequence. Given a training sequence, we can then evaluate the impact of the residual CFO on the corresponding channel estimate. We shall assume throughout that the residual offsets α_i's are independent and identically distributed. One commonly-used distribution is the uniform distribution over the interval $[-\alpha_{\max}, \alpha_{\max}]$ where $0 \le \alpha_{\max} \le 0.5$. Using these assumptions and the fact that $\mathbb{E}[g(x)] = \int_x g(x)f(x)dx$, for any random variable x, where $f(x)$ is the probability density function of x, we find that, for all $i,j = 0,1,...,L-1$, the (m,n) element of the matrix $\mathbf{C}_\mathbf{v}^{i,j}$, where $m,n = 0,1,...,KN-1$, is given in Eq. (44) where $\text{sinc}(x) \triangleq \frac{\sin(\pi x)}{\pi x}$ for any real number x, and $f(m) = m + l(N+\nu+1)$ where $l = 0,1,...,K-1$ and $lN \le m \le (l+1)N-1$.

B. Phase Noise

In the presence of PN affecting the free-running voltage-controlled oscillators (VCOs) of the transmitters and the receiver, the received signal over $K = 1$ training OFDM symbol is given by

$$\mathbf{y} = \mathbf{P}^{\text{rx}}\sum_{i=0}^{M-1}\mathbf{H}_i^C\,\mathbf{P}_i^{\text{tx}}\,\mathbf{s}_i + \mathbf{n} \tag{45}$$

where $\mathbf{s}_i$ denotes the training sequence transmitted by the ith transmit antenna and $\mathbf{H}_i^C$ denotes the matrix of the channel experienced by the ith transmit antenna. Although $\mathbf{H}_i^C$ is not exactly circulant due to the edge effect introduced by PN at the transmitters, it can be considered circulant with this effect lumped into the noise vector $\mathbf{n}$. For large FFT sizes, the edge effect can be ignored [29]. The PN perturbing the VCO of the ith transmit antenna is modeled by the diagonal matrix $\mathbf{P}_i^{\text{tx}} \triangleq \text{diag}(\{e^{j\phi_n^i}\}_{n=0}^{N-1})$ with ϕ_n^i representing the PN

$$\mathbf{C}_{\mathbf{v}}^{i,j}(m,n) = \begin{cases} 1 - \text{sinc}\left(\frac{2f(m)\,\alpha_{\max}}{N}\right) - \text{sinc}\left(\frac{2f(n)\,\alpha_{\max}}{N}\right) + \text{sinc}\left(\frac{2f(n-m)\,\alpha_{\max}}{N}\right), & i = j \\ \left(\text{sinc}\left(\frac{2f(m)\,\alpha_{\max}}{N}\right) - 1\right)\left(\text{sinc}\left(\frac{2f(n)\,\alpha_{\max}}{N}\right) - 1\right), & i \neq j \end{cases} \tag{44}$$

sample perturbing the transmitted signal by the ith transmit antenna at the nth sample[2]. Similarly, the PN perturbing the receiver VCO is modeled by the diagonal matrix $\mathbf{P}^{\text{rx}} \triangleq \text{diag}(\{e^{j\phi_n^{\text{rx}}}\}_{n=0}^{N-1})$. The discrete-time PN model is given by

$$\phi_n^i = \phi_{n-1}^i + \epsilon_n^i \quad \text{and} \quad \phi_n^{\text{rx}} = \phi_{n-1}^{\text{rx}} + \epsilon_n^{\text{rx}}, \tag{46}$$

where $\{\epsilon_n^i\}$ and $\{\epsilon_n^{\text{rx}}\}$ are independent Gaussian distributed random variables with zero means and variances $\left\{\frac{2\pi\beta_i^{\text{tx}}}{N f_{\text{sub}}}\right\}$ and $\left\{\frac{2\pi\beta^{\text{rx}}}{N f_{\text{sub}}}\right\}$ for all n respectively. Without loss of generality, we assume for all i, $\phi_0^i = \phi_0^{\text{rx}} = 0$. The parameters β_i^{tx} and β^{rx} denote the two-sided 3-dB linewidths of the Lorentzian power density spectrums of the VCOs feeding the ith transmit antenna and the receive antenna, respectively [30]. We express $\mathbf{P}_i^{\text{tx}}$ and $\mathbf{P}^{\text{rx}}$ as

$$\mathbf{P}_i^{\text{tx}} = \mathbf{I}_N + \left(\mathbf{P}_i^{\text{tx}} - \mathbf{I}_N\right) \triangleq \mathbf{I}_N + \widetilde{\mathbf{P}}_i^{\text{tx}},$$
$$\mathbf{P}^{\text{rx}} = \mathbf{I}_N + \left(\mathbf{P}^{\text{rx}} - \mathbf{I}_N\right) \triangleq \mathbf{I}_N + \widetilde{\mathbf{P}}^{\text{rx}},$$

and expand the LLSE of $\mathbf{h}$ and the trace of its error autocorrelation matrix, respectively, in Eq. (47) and (48), where the edge effect is ignored. Defining $\mathbf{\Psi} \triangleq \sum_{i=0}^{M-1} \mathbf{H}_i^C\, \widetilde{\mathbf{P}}_i^{\text{tx}}\, \mathbf{s}_i$, we expand t_{pn} in Eq. (49).

Using the properties in (40) in addition to the diagonal structure of $\widetilde{\mathbf{P}}^{\text{rx}}$ and assuming that the channel and the PN parameters are statistically independent, we write

$$t_{\text{pn},1} = \text{Tr}\left(\mathbf{C_h}\mathbf{S}^H\left(\left[\mathbf{S}\mathbf{S}^H\right] \circ \mathbb{E}\left[\tilde{\mathbf{p}}^{\text{rx}}\tilde{\mathbf{p}}^{\text{rx},H}\right]\right)\mathbf{S}\right), \tag{52}$$

where $\tilde{\mathbf{p}}^{\text{rx}}$ is a column vector containing the diagonal elements of $\widetilde{\mathbf{P}}^{\text{rx}}$. Furthermore, we can rewrite $\mathbf{\Psi}$ as

$$\mathbf{\Psi} = \sum_{i=0}^{M-1} \widehat{\mathbf{S}}_i \mathbf{h}_i = \begin{bmatrix}\widehat{\mathbf{S}}_0 & \widehat{\mathbf{S}}_1 & \dots & \widehat{\mathbf{S}}_{M-1}\end{bmatrix}\mathbf{h} \triangleq \widehat{\mathbf{S}}\mathbf{h}, \tag{53}$$

where $\widehat{\mathbf{S}}_i$ is a circulant matrix whose first column is $\widetilde{\mathbf{P}}_i^{\text{tx}}\, \mathbf{s}_i$. Using the above formulation of $\mathbf{\Psi}$ and the properties in (40), we rewrite $t_{\text{pn},2}$ in Eq. (50) where $\mathbf{p}^{\text{rx}}$ is a column vector containing the diagonal elements of $\mathbf{P}^{\text{rx}}$. Inspecting the structure of $\widehat{\mathbf{S}}$, we find that it can be expressed as

$$\widehat{\mathbf{S}} = \begin{bmatrix}\widetilde{\mathbf{P}}_0^C \circ \mathbf{S}_0 & \widetilde{\mathbf{P}}_1^C \circ \mathbf{S}_1 & \dots & \widetilde{\mathbf{P}}_{M-1}^C \circ \mathbf{S}_{M-1}\end{bmatrix},$$

where $\widetilde{\mathbf{P}}_i^C$ is a circulant matrix whose first column is $\tilde{\mathbf{p}}_i^{\text{tx}}$, the column vector containing the diagonal elements of $\widetilde{\mathbf{P}}_i^{\text{tx}}$. Hence,

$$\mathbb{E}\left[\widehat{\mathbf{S}}\right] = \left[\left[\mathbb{E}\left[\widetilde{\mathbf{P}}_0^C\right] \circ \mathbf{S}_0 \quad \mathbb{E}\left[\widetilde{\mathbf{P}}_1^C\right] \circ \mathbf{S}_1 \quad \dots \quad \mathbb{E}\left[\widetilde{\mathbf{P}}_{M-1}^C\right] \circ \mathbf{S}_{M-1}\right]\right.$$

where $\mathbb{E}\left[\widetilde{\mathbf{P}}_i^C\right]$ is formed by a circulant matrix whose first column is $\mathbb{E}\left[\tilde{\mathbf{p}}_i^{\text{tx}}\right]$. Finally, we consider the last term $t_{\text{pn},4}$ and express it in Eq. (51) where we used the fact that $\mathbf{H}_i^C = \mathbf{F}^H\mathbf{H}_i^D\mathbf{F}$ where $\mathbf{H}_i^D$ is a diagonal matrix whose diagonal is formed by the vector $\mathbf{H}_i$ which is the N-point DFT of $\mathbf{h}_i$ as defined in Section II. Furthermore,

$$\mathbb{E}_{\mathbf{h},\mathbf{P}^{\text{rx}},\left\{\widetilde{\mathbf{P}}_i^{\text{tx}}\right\}}\left[\cdot\right] = \mathbb{E}_{\mathbf{P}^{\text{rx}}}\mathbb{E}_{\mathbf{h}|\mathbf{P}^{\text{rx}}}\mathbb{E}_{\left\{\widetilde{\mathbf{P}}_i^{\text{tx}}\right\}|\mathbf{h},\mathbf{P}^{\text{rx}}}\left[\cdot\right] = \mathbb{E}_{\mathbf{P}^{\text{rx}}}\mathbb{E}_{\mathbf{h}}\mathbb{E}_{\left\{\widetilde{\mathbf{P}}_i^{\text{tx}}\right\}}\left[\cdot\right]$$

where $|$ is the statistical conditioning operator and the second equality follows from the fact that $\mathbf{h}$, $\mathbf{P}^{\text{rx}}$, and $\left\{\widetilde{\mathbf{P}}_i^{\text{tx}}\right\}$ are statistically independent. Using the above observations and the fact that $\mathbf{P}^{\text{rx}}$, $\left\{\mathbf{H}_i^D\right\}$, and $\left\{\widetilde{\mathbf{P}}_i^{\text{tx}}\right\}$ are all diagonal matrices, we write

$$t_{\text{pn},4} = \text{Tr}\left(\mathbf{S}^H\left(\mathbf{\Xi} \circ \mathbb{E}\left[\mathbf{p}^{\text{rx}}\mathbf{p}^{\text{rx},H}\right]\right)\mathbf{S}\right) \tag{54}$$

where $\mathbf{\Xi} = \sum_{i=0}^{M-1}\sum_{j=0}^{M-1}\mathbf{F}^H\left(\mathbf{\Upsilon}_{i,j} \circ \mathbb{E}\left[\mathbf{H}_i\mathbf{H}_j^H\right]\right)\mathbf{F}$ and $\mathbf{\Upsilon}_{i,j} = \mathbf{F}\left(\left[\mathbf{s}_i\mathbf{s}_j^H\right] \circ \mathbb{E}\left[\tilde{\mathbf{p}}_i^{\text{tx}}\tilde{\mathbf{p}}_j^{\text{tx},H}\right]\right)\mathbf{F}^H$. Finally, the vector $\mathbb{E}\left[\tilde{\mathbf{p}}_i^{\text{tx}}\right]$ and the correlation matrices $\mathbb{E}\left[\tilde{\mathbf{p}}^{\text{rx}}\tilde{\mathbf{p}}^{\text{rx},H}\right]$, $\mathbb{E}\left[\mathbf{p}^{\text{rx}}\tilde{\mathbf{p}}^{\text{rx},H}\right]$, $\mathbb{E}\left[\mathbf{p}^{\text{rx}}\mathbf{p}^{\text{rx},H}\right]$, and $\mathbb{E}\left[\tilde{\mathbf{p}}_i^{\text{tx}}\tilde{\mathbf{p}}_j^{\text{tx},H}\right]$ are determined by using the relation $\mathbb{E}\left[e^{jx}\right] = e^{-\sigma_x^2/2}$ for the random variable $x \sim N(0,\sigma_x^2)$. In fact, $\mathbb{E}\left[e^{jx}\right]$ is the characteristic function [31] of x, $\psi_x(jt)$, evaluated at $t = 1$. Furthermore, the extension of the above analysis to the general case of $K > 1$ training OFDM symbols is straightforward.

V. Design Trade-offs

A closer inspection of the channel estimate MSE expressions derived in Sections IV-A and IV-B reveals a tradeoff between the PAPR of the training sequences and their robustness to CFO and PN. An intuitive explanation is that the training sequence with low PAPR tends to distribute its energy uniformly among all samples including late ones; however, this makes it less immune to PN and CFO which severely affect late samples. On the other hand, the training sequence that is more robust to CFO and PN, should concentrate its energy in the early samples as explained earlier; however, this will result in increasing PAPR. Towards an analytical explanation, we examine the expression of t_{fo} in (41). To simplify the expression, we fairly assume that the channel responses seen by all transmit antennas are uncorrelated, so the matrix $\mathbf{C_h}$ is block-diagonal. Hence, t_{fo} is affected only by the diagonal blocks of $\mathbf{C}_{\mathbf{S}_\Delta}$, i.e. $\{\mathbf{C}_{\mathbf{S}_\Delta}(i,i),\, 0 \le i \le L-1\}$, which are all partly constructed by the term $\left[\mathbf{S}\mathbf{S}^H\right] \circ \mathbf{C}_{\mathbf{v}}^{i,i}$. Inspecting the structure of the matrix $\mathbf{C}_{\mathbf{v}}^{i,i}$ in (44), we find that the energies of the elements increase as the column and/or row indices increase. Hence, if the training sequence has its energy concentrated in the early samples, then the matrix $\mathbf{S}\mathbf{S}^H$ will have the opposite structure of $\mathbf{C}_{\mathbf{v}}^{i,i}$ and, hence, the Hadamard product will yield small elements. This will eventually result in a small value for t_{fo}, i.e. more immunity to CFO. The same rationale can be applied to the expression of t_{pn} keeping in mind that the variance of the PN sample increases as its index increases according to the model adopted in (46).

[2]Transmit antennas supporting the same user experience the same PN matrix as they are fed by the same VCO.

$$\hat{\mathbf{h}} = \frac{1}{\tilde{c}}\mathbf{S}^H\mathbf{y} = \mathbf{h} + \underbrace{\frac{1}{\tilde{c}}\mathbf{S}^H\widetilde{\mathbf{P}}^{\text{rx}}\,\mathbf{S}\mathbf{h} + \frac{1}{\tilde{c}}\mathbf{S}^H\mathbf{P}^{\text{rx}}\sum_{i=0}^{M-1}\mathbf{H}_i^C\,\widetilde{\mathbf{P}}_i^{\text{tx}}\,\mathbf{s}_i}_{\triangleq\mathbf{e}_{\text{pn}}} + \frac{1}{\tilde{c}}\mathbf{S}^H\mathbf{n} \tag{47}$$

$$t_e = \text{Tr}\left(\mathbb{E}\left[\left(\mathbf{h}-\hat{\mathbf{h}}\right)\left(\mathbf{h}-\hat{\mathbf{h}}\right)^H\right]\right) = \text{Tr}\left(\mathbb{E}\left[\mathbf{e}_{\text{pn}}\mathbf{e}_{\text{pn}}^H\right]\right) + 2\sigma_e^2 \triangleq t_{\text{pn}} + 2\sigma_e^2 \tag{48}$$

$$t_{\text{pn}} = \frac{1}{\tilde{c}^2}\underbrace{\text{Tr}\left(\mathop{\mathbb{E}}_{\mathbf{h},\widetilde{\mathbf{P}}^{\text{rx}}}\left[\mathbf{S}^H\widetilde{\mathbf{P}}^{\text{rx}}\,\mathbf{S}\mathbf{h}\mathbf{h}^H\mathbf{S}^H\widetilde{\mathbf{P}}^{\text{rx},H}\,\mathbf{S}\right]\right)}_{\triangleq t_{\text{pn},1}} + \frac{1}{\tilde{c}^2}\underbrace{\text{Tr}\left(\mathop{\mathbb{E}}_{\mathbf{h},\mathbf{P}^{\text{rx}},\{\widetilde{\mathbf{P}}_i^{\text{tx}}\}}\left[\mathbf{S}^H\widetilde{\mathbf{P}}^{\text{rx}}\,\mathbf{S}\mathbf{h}\boldsymbol{\Psi}^H\mathbf{P}^{\text{rx},H}\,\mathbf{S}\right]\right)}_{\triangleq t_{\text{pn},2}}$$

$$+\frac{1}{\tilde{c}^2}\underbrace{\text{Tr}\left(\mathop{\mathbb{E}}_{\mathbf{h},\mathbf{P}^{\text{rx}},\{\widetilde{\mathbf{P}}_i^{\text{tx}}\}}\left[\mathbf{S}^H\mathbf{P}^{\text{rx}}\,\boldsymbol{\Psi}\mathbf{h}^H\mathbf{S}^H\widetilde{\mathbf{P}}^{\text{rx},H}\,\mathbf{S}\right]\right)}_{\triangleq t_{\text{pn},3}} + \frac{1}{\tilde{c}^2}\underbrace{\text{Tr}\left(\mathop{\mathbb{E}}_{\mathbf{h},\mathbf{P}^{\text{rx}},\{\widetilde{\mathbf{P}}_i^{\text{tx}}\}}\left[\mathbf{S}^H\mathbf{P}^{\text{rx}}\,\boldsymbol{\Psi}\boldsymbol{\Psi}^H\mathbf{P}^{\text{rx},H}\,\mathbf{S}\right]\right)}_{\triangleq t_{\text{pn},4}} \tag{49}$$

$$t_{\text{pn},2} = t_{\text{pn},3}^* = \text{Tr}\left(\mathop{\mathbb{E}}_{\mathbf{h},\mathbf{P}^{\text{rx}},\{\widetilde{\mathbf{P}}_i^{\text{tx}}\}}\left[\mathbf{S}^H\widetilde{\mathbf{P}}^{\text{rx}}\,\mathbf{S}\mathbf{h}\mathbf{h}^H\widehat{\mathbf{S}}^H\mathbf{P}^{\text{rx},H}\,\mathbf{S}\right]\right) = \text{Tr}\left(\mathbf{C}_{\mathbf{h}}\mathbb{E}\left[\widehat{\mathbf{S}}\right]^H\left(\left[\mathbf{S}\mathbf{S}^H\right]\circ\mathbb{E}\left[\mathbf{p}^{\text{rx}}\tilde{\mathbf{p}}^{\text{rx},H}\right]\right)\mathbf{S}\right) \tag{50}$$

$$t_{\text{pn},4} = \text{Tr}\left(\mathbb{E}_{\mathbf{h},\mathbf{P}^{\text{rx}},\{\widetilde{\mathbf{P}}_i^{\text{tx}}\}}\left[\mathbf{S}^H\mathbf{P}^{\text{rx}}\left(\sum_{i=0}^{M-1}\sum_{j=0}^{M-1}\mathbf{F}^H\mathbf{H}_i^D\mathbf{F}\,\widetilde{\mathbf{P}}_i^{\text{tx}}\,\mathbf{s}_i\mathbf{s}_j^H\widetilde{\mathbf{P}}_j^{\text{tx},H}\,\mathbf{F}^H\mathbf{H}_j^{D,H}\mathbf{F}\right)\mathbf{P}^{\text{rx},H}\,\mathbf{S}\right]\right) \tag{51}$$

VI. Simulation Results

We have simulated the performance of an OFDM system with N = 64 and $\nu_{\max}$ = 15 as in [9]. We consider uplink transmission in a Multi-User MIMO system with 2 co-located receive antennas at the BST and 2 users each equipped with 2 transmit antennas over which the Alamouti STBC [32] is employed. Each user employs a non-systematic rate-1/2 convolutional code with octal generator (133,171) and constraint length = 7 as in [9]. Coded bits are Quadrature Phase Shift Keying (QPSK) modulated. All channel paths are assumed to have uncorrelated and identically-distributed CIRs with 8 zero-mean complex Gaussian taps following an exponentially-decaying power delay profile (PDP) with a 3 dB decay per tap. K OFDM training symbols are transmitted over each transmit antenna for the purpose of channel estimation as described in Section III. The CIR estimates are used for detection of the OFDM data symbols through the joint Linear Minimum-Mean-Square-Error (LMMSE) technique which processes the received signals from the 2 receive antennas jointly to detect the two users [33]. The background noise is assumed to be AWGN with a single-sided power spectral density of N_o Watts/Hz. The bit energy is denoted by E_b and the per-user Signal-to-Noise Ratio (SNR) is defined as SNR $= \frac{E_b}{N_o}$.

Using these parameters, the dimensionality condition in (20) is met with $K \geq 1$. In Fig. 2, the bit error rate (BER) performances of three FD training sequences (namely: Chirp, Golay, and flat (TD Impulsive)) with $K = 1$ and 2 are compared with the perfect channel state information (CSI) case. In Fig. 2, all users are assumed to have perfect frequency synchronization with the receiver. All training sequences achieve roughly the same BER performance with SNR losses -compared to the perfect CSI case- of 1.5 and 0.7 dB for $K = 1$ and 2, respectively. For comparison purpose, the performance of a random BPSK sequence *not* satisfying the optimality condition is also shown in Fig. 2 for $K = 1$ and 2. The performance of the random sequence is inferior to that of the other sequences satisfying the optimality condition; especially with $K = 1$ training symbol where the number of equations equals the number of unknowns making the channel estimate unreliable when the optimality condition is not satisfied. From another perspective, our optimally-designed training sequences with $K = 1$ training symbol achieve comparable performance to that of the random sequence with $K = 2$ training symbols, i.e. with 50% less training overhead. This is in addition to the additional complexity needed to invert the matrix $\mathbf{S}^H\mathbf{S}$ which is not a scaled identity in the case of non-optimal sequences.

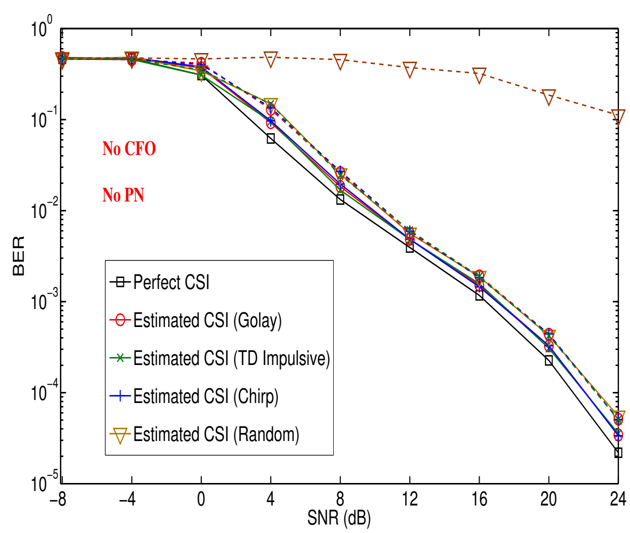


Fig. 2. BER versus SNR for $K = 1$ (dashed) and 2 (solid) training OFDM symbols without CFO.

For the channel PDP described above, we plot t_{fo} derived in (39) versus $\alpha_{\max}$ in Fig. 3 for the proposed training sequences where we observe that the flat FD sequence is more robust to the residual CFOs than the other two sequences thanks to the impulsive nature of its corresponding training sequences where most of power is concentrated in the early samples where the CFO effect is small (CFO effect increases with time). The impact of the residual CFOs on the BER performance is shown in Fig. 4 for $K = 2$ training symbol with $\alpha_{\max} = 0.01$ and 0.05 where the superiority of the flat sequence is observed also in the BER performance. While CFOs with $\alpha_{\max} = 0.01$ do not cause a significant performance degradation, CFOs with $\alpha_{\max} = 0.05$ limit the system performance at high SNR.

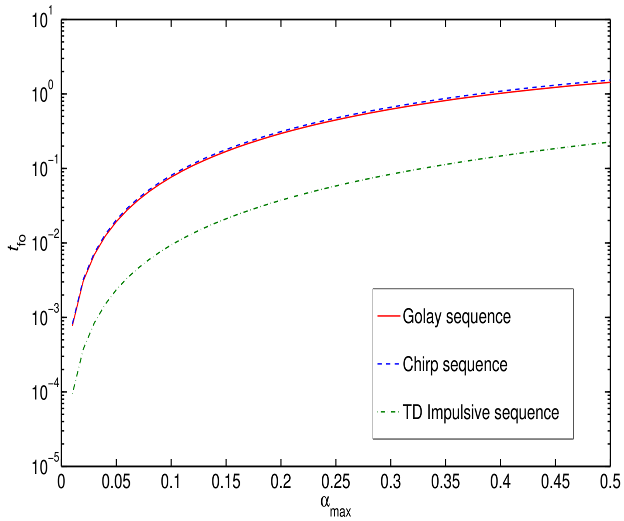


Fig. 3. Plot of t_{fo} in (39) versus $\alpha_{\max}$ with $K = 1$ training OFDM symbol and 2 users each with 1 transmit antenna.

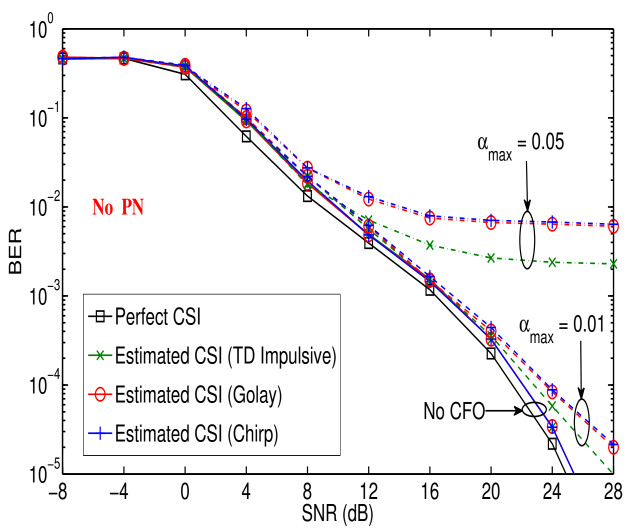


Fig. 4. BER versus SNR without CFOs (solid curves) and with CFOs of $\alpha_{\max} = 0.01$ (dashed curves) and $\alpha_{\max} = 0.05$ (dash-dotted curves). Two training OFDM symbols are used.

Assuming the VCOs feeding all transmit and receive antennas to have the same 3-dB linewidth β, i.e. $\beta_i^{\text{tx}} = \beta^{\text{rx}} = \beta$, $\forall i$, we plot t_{pn} in (49) versus $\sigma_{\text{pn}}^2 \triangleq \frac{2\pi\beta}{N f_{\text{sub}}}$ in Fig. 5 for $K = 1$ training symbols. Like the CFO case, the flat FD training sequence is also more immune to PN than chirp and Golay sequences for the same reason. Fig. 6 depicts the PN impact on the BER performance for different values of σ_{pn}^2 without CFOs, i.e. $\alpha_{\max} = 0$. With $\sigma_{\text{pn}}^2 = 10^{-4}$, PN becomes performance-limiting at high SNR while no significant deterioration is observed for smaller PN variances such as $\sigma_{\text{pn}}^2 = 10^{-5}$.

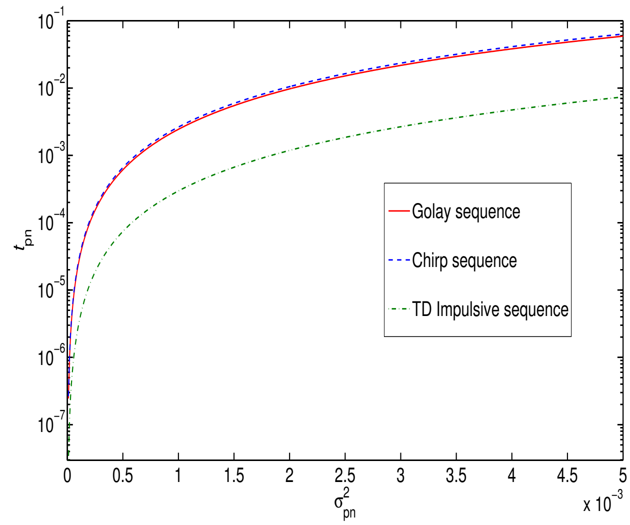


Fig. 5. Plot of t_{pn} in (49) versus σ_{pn}^2 with $K = 1$ training OFDM symbol and 2 users each with 1 transmit antenna.

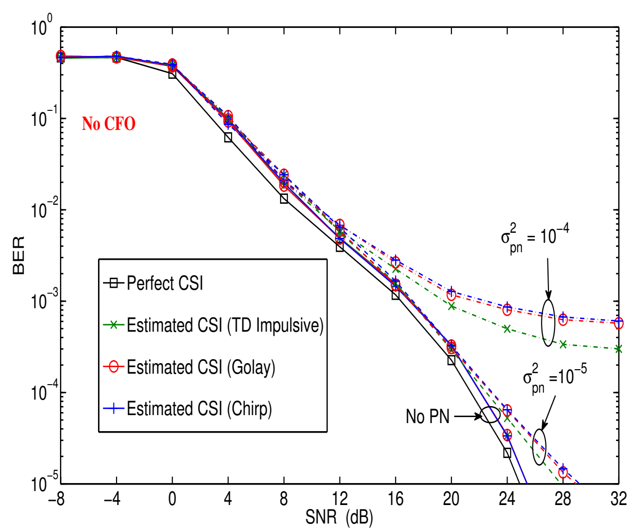


Fig. 6. BER versus SNR for different values of σ_{pn}^2 $\{0$ (solid), 10^{-5} (dashed), and 10^{-4} (dash-dotted)$\}$ with $K = 2$ training OFDM symbols and without CFO.

In Fig. 7, we simulate the BER performance of the training sequences in peak-limited channels under CFO and PN. In peak-limited channels, the received signal power is limited by $P_{\max}$ above which the received signal power is saturated (clipped). In Fig. 7, we show the BER versus $\Delta P \triangleq P_{\text{falt}} - P_{\max}$ (dB) where P_{flat} is the received peak power of the FD flat sequence which is the largest over the three sequences. For small values of ΔP (i.e. high clipping power values), the FD flat sequence outperforms the other two sequences due to its immunity to CFO and PN as discussed before. However, the situation changes at high values of ΔP where the distortion of the FD flat sequence, caused by the peak-limited channel, dominates its immunity to CFO and PN. Hence, we observe that despite achieving the same performance in peak-unlimited channels without CFO or PN, the proposed sequences exhibit different behaviors under these practical impairments.

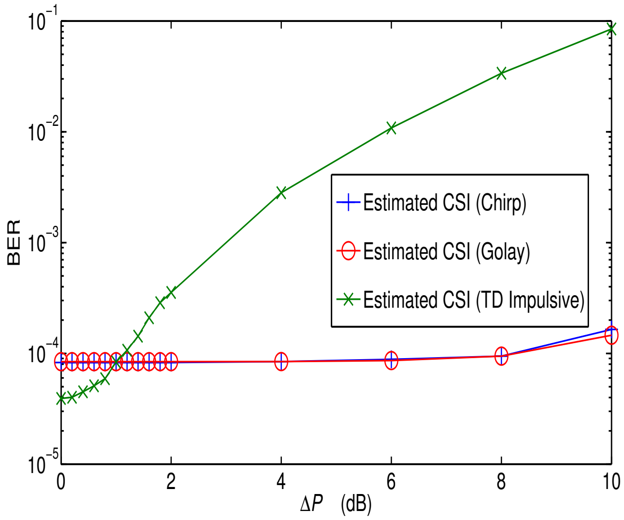


Fig. 7. Comparison of the training sequences performances in peak-limited channels with $\alpha_{\max} = 0.01$, $\sigma^2_{\text{pn}} = 10^{-5}$, SNR = 26dB, and $K = 2$ training OFDM symbols.

Fig. 8 shows the real parts of a single CIR realization along with its estimates in the presence of CFO with $\alpha_{\max} = 0.01$ and PN with $\sigma^2_{\text{pn}} = 10^{-4}$. The real parts of CIR estimates shown in Fig. 9 are with $\alpha_{\max} = 0.1$ and $\sigma^2_{\text{pn}} = 10^{-3}$ where the increased maximum CFO and the PN variance degrade the accuracy of the CIR estimates. The values of $\alpha_{\max}$ and σ^2_{pn} used in our simulations are small since CFO and PN compensations usually precede channel estimation in practical systems. The image parts of CIR realizations are omitted due to space. The trade-off between the PAPR of the training sequences and their immunity to CFO and PN is observed by inspecting Figs. 3 and 5 with Table III.

VII. Conclusions

We derived the MMSE optimality criteria for training sequence designs in Multi-User MIMO-OFDM systems. In spectrally-efficient uplink transmission scenarios where users are separated using spatial processing at the base station, our analysis holds for an arbitrary number of users, OFDM training symbols, transmit antennas per user, and channel delay spread. Within the family of training designs that are MMSE-optimal under ideal conditions, we found that robustness to residual CFO and PN can vary significantly. We also derived analytical expressions for the increase in channel estimation MSE in the presence of CFO and PN. Our analysis includes three detailed case studies; Chirp, Golay, and time-domain impulsive training sequence designs. In each case, we quantified the underlying tradeoff between PAPR and robustness to CFO and PN.

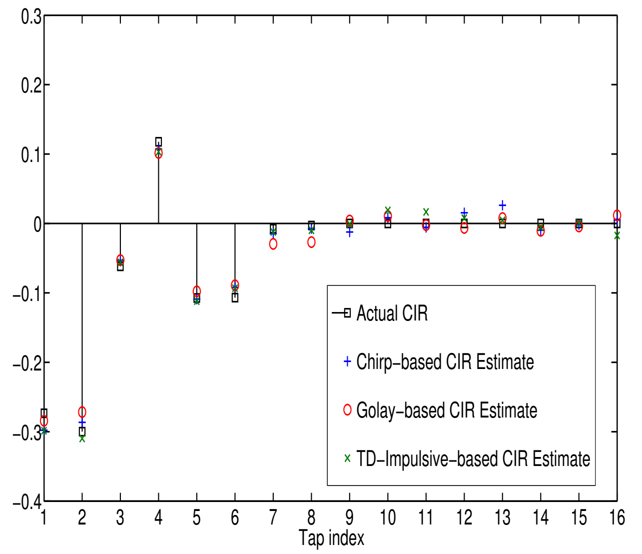


Fig. 8. A realization of the CIR real part and its estimates with $\alpha_{\max} = 0.01$, $\sigma^2_{\text{pn}} = 10^{-4}$, SNR = 15dB, and 2 training OFDM symbols.

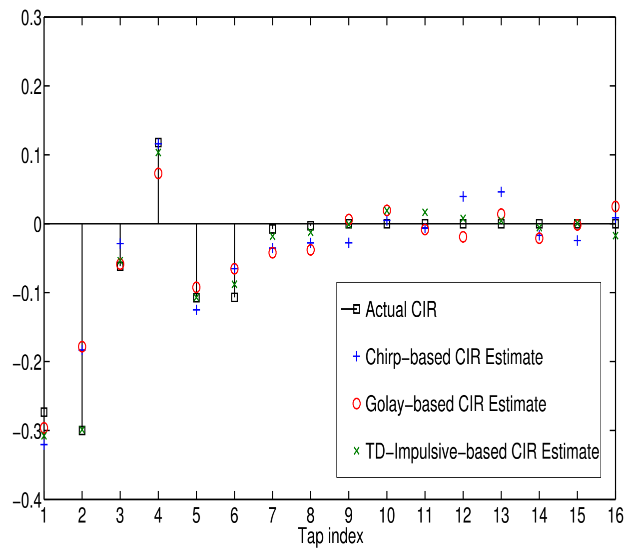


Fig. 9. A realization of the CIR real part and its estimates with $\alpha_{\max} = 0.1$, $\sigma^2_{\text{pn}} = 10^{-3}$, SNR = 15dB, and 2 training OFDM symbols.

References

[1] G. J. Foschini, "Layered space-time architecture for wireless communication in a fading environment when using multi-element antennas," *Bell Labs Technical Journal*, vol. 1, pp. 41–59, 1996.

[2] E. Telatar, "Capacity of multi-antenna gaussian channels," *European Transactions on Telecommunications*, 1999.

[3] V. Tarokh, N. Seshadri, and A. Calderbank, "Space-time codes for high data rate wireless communication: performance criterion and code construction," *IEEE Transactions on Information Theory*, 1998.

[4] A. R. S. Bahai, B. R. Saltzberg, and M. Ergen, *Multi-carrier digital communications theory and applications of OFDM*. New York, NY: Springer, 2004.

[5] D. Tse and P. Viswanath, *Fundamentals of wireless communication*. New York, NY: Cambridge University Press, 2005.

[6] M. Jiang and L. Hanzo, "Multiuser MIMO-OFDM for next-generation wireless systems," *Proceedings of the IEEE*, vol. 95, no. 7, pp. 1430–1469, Jul. 2007.

[7] "IEEE standard for local and metropolitan area networks part 16: Air

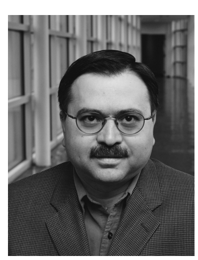
interface for fixed and mobile broadband wireless access systems," *IEEE Std 802.16e-2005*, 2006.
[8] "Physical channels and modulation," *3GPP TS 36.211, Ver 8.6.0*, March 2009.
[9] "IEEE candidate standard 802.11n: Wireless LAN medium access control (MAC) and physical layer (phy) specifications," *[Online]. Available: http : //grouper.ieee.org/groups/802/11/Reports/tgn_update.htm*, 2009.
[10] S. Parkvall, E. Dahlman, A. Furuskar, Y. Jading, M. Olsson, S. Wanstedt, and K. Zangi, "LTE-advanced - evolving LTE towards IMT-advanced," in *IEEE Vehicular Technology Conference*, 2008.
[11] Y. G. Li, "Pilot-symbol-aided channel estimation for OFDM in wireless systems," *IEEE Trans. Veh. Technol.*, vol. 49, pp. 1207–1215, 2000.
[12] L. Deneire, P. Vandenameele, L. van der Perre, B. Gyselinckx, and M. Engels, "A low-complexity ML channel estimator for OFDM," *IEEE Trans. Commun.*, vol. 51, pp. 67 – 75, 2003.
[13] O. Edfors, M. Sandell, J.-J. van de Beek, S. K. Wilson, and P. O. Böesson, "OFDM channel estimation by singular value decomposition," *IEEE Trans. Commun.*, vol. 46, pp. 931–939, 1998.
[14] Y. G. Li, N. Seshadri, and S. Ariyavisitakul, "Channel estimation for OFDM systems with transmitter diversity in mobile wireless channels," *IEEE Journal on Selected Areas in Communications*, vol. 17, pp. 461 – 471, 1999.
[15] H. Minn, D. I. Kim, and V. Bhargava, "A reduced complexity channel estimation for OFDM systems with transmit diversity in mobile wireless channels," *IEEE Transactions on Communications*, 2002.
[16] Z. J. Wang, Z. Han, and K. J. R. Liu, "A MIMO-OFDM channel estimation approach using time of arrivals," *IEEE Transactions On Wireless Communications*, vol. 4, pp. 1207 – 1213, May 2005.
[17] C. Fragouli, N. Al-Dhahir, and W. Turin, "Training-based channel estimation for multiple-antenna broadband transmissions," *IEEE Transactions on Wireless Communications*, 2003.
[18] H. Minn and N. Al-Dhahir, "Optimal training signals for MIMO OFDM channel estimation," *IEEE Trans. Wireless Comm.*, vol. 5, no. 5, pp. 1158–1168, April 2006.
[19] Y. G. Li, "Simplified channel estimation for OFDM systems with multiple transmit antennas," *IEEE Transactions on Wireless Communications*, vol. 1, pp. 67 – 75, 2002.
[20] Y. Zeng, A. R. Leyman, S. Ma, and T.-S. Ng, "Optimal pilot and fast algorithm for MIMO-OFDM channel estimation," in *Proceeding of IEEE International Conference on Information, Communications and Signal Processing (ICICS)*, 2005.
[21] F. Horlin and L. V. der Perre, "Optimal training sequences for low complexity ML multi-channel estimation in multi-user MIMO OFDM-based communications," in *Proceeding of IEEE International Conference on Communications (ICC)*, vol. 4, Jun. 2004.
[22] S. D. Howard, A. R. Calderbank, and W. Moran, "A simple signal processing architecture for instantaneous radar polarimetry," *IEEE Trans. Inform. Theory*, vol. 53, no. 4, pp. 1282–1289, Apr. 2007.
[23] M. D. Zoltowski, T. R. Qureshi, and R. Calderbank, "Complementary codes based channel estimation for MIMO-OFDM systems," in *Proc. 46th Annual Allerton Conf. Communication, Control, and Computing*, Monticello, IL, Sep. 2008, pp. 133–138.
[24] C. C. Tseng, "Signal multiplexing in surface-wave delay lines using orthogonal pairs of Golay's complementary sequences," *IEEE Transactions on Sonics and Ultrasonics*, 1971.
[25] C. Tseng and C. Liu, "Complementary sets of sequences," *IEEE Transactions on Information Theory*, 1972.
[26] S. M. Kay, *Fundamentals Of Statistical Signal Processing: Estimation Theory*. New Jersey: Prentice Hall, 1993.
[27] D. Chu, "Polyphase codes with good periodic correlation properties," *IEEE Trans. on Info. Theory*, vol. IT-18, 1972.
[28] M. J. E. Golay, "Complementary series," *IRE Trans. Inform. Theory*, vol. 7, no. 2, pp. 82–87, April 1961.
[29] R. Gray, *Toeplitz and Circulant Matrices: A review*. Hanover, MA: Now Publishers Inc., 2006.
[30] T. Pollet, M. V. Bladel, and M. Moeneclaey, "BER sensitivity of OFDM systems to carrier frequency offset and wiener phase noise," *IEEE Transcations on Communications*, vol. 43, 1995.
[31] J. G. Proakis, *Digital Communication*, 4th ed. New York, NY: McGraw-Hill, 2001.
[32] S. M. Alamouti, "A simple transmit diversity technique for wireless communications," *IEEE Journal on Selected Areas in Communications*, vol. 16, pp. 1451–1458, 1998.
[33] A. Naguib, N. Seshadri, and A. Calderbank, "Increasing data rate over wireless channels," *IEEE Signal Processing Magazine*, vol. 17, 2000.

Yuejie Chi (S'09) received the B.E. (Hon.) in Electrical Engineering from Tsinghua University, Beijing, China, in 2007. She is currently working towards the Ph.D degree in the department of Electrical Engineering at Princeton University.

Her research interests include compressed sensing, statistical signal processing, machine learning, wireless communications and active sensing.

Ahmad Gomaa (S'09) received his B.S and M.S degrees in electronics and communications engineering from Cairo University, Egypt, in 2005 and 2008, respectively. He is currently pursuing the Ph.D. degree at the University of Texas at Dallas, USA.

His research interests include sparse FIR filters design, RF impairments compensation at the baseband, training sequence design for channel estimation, and compressive sensing applications to digital communications.



Naofal Al-Dhahir earned his PhD degree in Electrical Engineering from Stanford University in 1994. From 1994 to 2003 he was a principal member of the technical staff at GE Research and AT&T Shannon Lab. In 2003, he joined UT-Dallas as an Associate Professor and became a full Professor in 2007, and a Jonsson Distinguished Professor of Engineering in 2010. His current research interests include broadband wireless and wireline transmission, MIMO-OFDM transceivers, baseband processing to mitigate RF/analog impairments and compressive sensing.

He has authored over 220 papers and holds 31 issued US patents. He is co-recipient of the IEEE VTC Fall 2005 best paper award, the 2005 IEEE signal processing society young author best paper award, the 2006 IEEE Donald G. Fink best journal paper award and is an IEEE fellow.

Robert Calderbank (M'89-SM'97-F'98) received the BSc degree in 1975 from Warwick University, England, the MSc degree in 1976 from Oxford University, England, and the PhD degree in 1980 from the California Institute of Technology, all in mathematics.

Dr. Calderbank is Dean of Natural Sciences at Duke University. He was previously Professor of Electrical Engineering and Mathematics at Princeton University where he directed the Program in Applied and Computational Mathematics. Prior to joining Princeton in 2004, he was Vice President for Research at AT&T, responsible for directing the first industrial research lab in the world where the primary focus is data at scale. At the start of his career at Bell Labs, innovations by Dr. Calderbank were incorporated in a progression of voiceband modem standards that moved communications practice close to the Shannon limit. Together with Peter Shor and colleagues at AT&T Labs he showed that good quantum error correcting codes exist and developed the group theoretic framework for quantum error correction. He is a co-inventor of space-time codes for wireless communication, where correlation of signals across different transmit antennas is the key to reliable transmission.

Dr. Calderbank served as Editor in Chief of the IEEE TRANSACTIONS ON INFORMATION THEORY from 1995 to 1998, and as Associate Editor for Coding Techniques from 1986 to 1989. He was a member of the Board of Governors of the IEEE Information Theory Society from 1991 to 1996 and from 2006 to 2008. Dr. Calderbank was honored by the IEEE Information Theory Prize Paper Award in 1995 for his work on the Z4 linearity of Kerdock and Preparata Codes (joint with A.R. Hammons Jr., P.V. Kumar, N.J.A. Sloane, and P. Sole), and again in 1999 for the invention of space-time codes (joint with V.Tarokh and N. Seshadri). He received the 2006 IEEE Donald G. Fink Prize Paper Award and the IEEE Millennium Medal, and was elected to the US National Academy of Engineering in 2005.