Space-Time Interference Cancellation in MIMO-OFDM Systems

Taiwen Tang, *Student Member, IEEE,* and Robert W. Heath, Jr., *Member, IEEE*

***Abstract*— In this paper, a two stage hybrid interference cancellation and equalization framework is proposed for interference cancellation in the uplink of multiple-input multiple-output orthogonal frequency division multiplexing (MIMO-OFDM) systems. The first stage uses time domain equalization to suppress co-channel interference, mitigate asynchronism, and shorten the post-equalization channel response to be no longer than the length of the cyclic prefix. The second stage performs low-complexity single tap equalization and detection in the frequency domain. The framework is developed specifically for spatial multiplexing and is applied to multiuser MIMO-OFDM systems with asynchronism between users as well as to single-user MIMO-OFDM systems. Various equalizer design methods are proposed that determine the coefficients directly from the training data and are compared with methods based on channel estimates. The equalizer coefficients and post-equalization channel response are found by solving a joint optimization that maximizes the signal to interference-plus-noise ratio (SINR) in the frequency domain. Simulations compare various training based methods and show the proposed methods provide good bit error rate (BER) and SINR performance in a variety of interference scenarios.**

***Index Terms*— MIMO, OFDM, interference, cancellation, space-time.**

I. INTRODUCTION

Multiple-input multiple-output orthogonal frequency division multiplexing (MIMO-OFDM) combines the simple equalization of OFDM modulation with the capacity, diversity, and array gain of MIMO communication [1]–[6]. Currently MIMO-OFDM is being considered in various multiuser systems including high-speed wireless local area networks [7] [8] and next generation cellular systems [9]. Communication in multiuser MIMO-OFDM systems, however, requires dealing with a variety of sources of interference besides multiuser interference, including co-antenna interference which depends on the space-time code, interference from jamming, and self interference from the time variation and dispersion of the channel. Transceiver techniques that can flexibly deal with interference are thus an important component of a MIMO-OFDM system.

The dominant sources of interference in a multiuser MIMO-OFDM system are multiple access interference, co-antenna interference which is the interference caused by the signals from multiple transmit antennas of a given user being received on the same receive antenna, and inter-carrier interference (ICI). In particular, ICI has a number of sources. One form of ICI is created when the delay spread of the channel is longer than the cyclic prefix [10] [11], which is used to guard against dispersion from adjacent OFDM symbols. Usually, OFDM systems employ a sufficiently long cyclic prefix to avoid these problems. But in certain harsh propagation environments, for example, the cell edges of clustered urban environments, the channel delay spread can be very large. Further, if the residual timing offset is large in the system, equivalently, this may cause the effective channel response to be longer than the cyclic prefix. More importantly, in multiuser OFDM systems, the asynchronicity between users, which is caused by the different propagation distances between users and the receiver, can create an additional source of ICI when the delay differences between users are significant [12]–[15]. We illustrate this effect in Fig. 1. Other forms of ICI are created when the channel varies during an OFDM symbol period [16]–[18]. The time variation of the channel usually comes from the mobility or residual frequency offsets between transmitters and receivers. In this paper, we assume that the mobility is sufficiently low and the residual frequency offsets are properly compensated. We discuss the problem of frequency offsets in [19].

The issue of ICI suppression in OFDM systems has been studied extensively in [10] [11] [20]–[30]. In general, there are two categories of equalization schemes to handle ICI in OFDM systems. The first one is frequency domain equalization. It has been presented in [12] that due to the significant asynchronicity of the user's signal, the per-tone beamforming scheme cannot completely cancel the ICI introduced by the asynchronicity unless the number of antennas is greater than the number of time taps in the channel, which is not feasible in real systems. Per-tone equalization [29] [30] is another scheme of frequency domain equalization, where each OFDM tone has a different equalizer. This approach can be applied to asynchronous systems, but the design complexity is high, and requires memory to store all the equalizer coefficients [30]. The second approach is time domain equalization with channel shortening [14], [31]–[34]. The channel shortening problem has been studied extensively for Discrete Multitone Systems (DMT) [31] [32], and can be applied to multiuser OFDM systems. Essentially, the asynchronicity in the multiuser system effectively gives a longer tail to the channel response. By applying channel shortening, the channel responses for all the users can be shortened into certain time windows of

Manuscript received May 20, 2004; revised November 3, 2004. This material is based in part upon work supported by the Texas Advanced Technology Program under Grant Nos. 003658-0614-2001 and 003658-0380-2003 and by the National Science Foundation under grant number 0322957.

Taiwen Tang is with the Wireless Networking & Communications Group, Department of Electrical & Computer Engineering, The University of Texas at Austin, Austin, TX 78712, USA (e-mail: twtang@mail.utexas.edu).

Robert W. Heath, Jr. is with the Wireless Networking & Communications Group, Department of Electrical & Computer Engineering, The University of Texas at Austin, Austin, TX 78712, USA (e-mail: rheath@ece.utexas.edu).

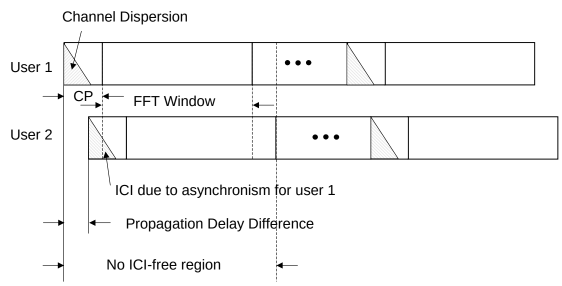


Fig. 1. Intercarrier interference caused by the difference of propagation delays between users. The cyclic prefix (CP) is not long enough to deal with the ICI caused by the asynchronism between users when the propagation distances of users are significantly different.

dimension less than the cyclic prefix to facilitate frequency domain cancellation.

Previous work [14] [33] on time domain equalization for MIMO-OFDM systems required channel estimates to find the optimal equalizers. Channel estimation techniques for general OFDM systems have been discussed in [35] [36], however, obtaining a good channel estimate in the space-time multiuser scenario is a difficult task because the receiver may not be aware of the training information for the interference channel. Additionally, "noise propagation" occurs when the estimated channel is used to find the equalizer since the effect of noise which corrupts the channel estimation might further degrade the design of equalizers. Direct training based equalizer design, which aims to obtain the equalizer coefficients from training directly without estimating the channel [37] [38], however, do not suffer from the aforementioned problems. Therefore, it is of interest to adopt this *direct* approach for equalization in the multiuser MIMO-OFDM scenario.

In this paper, we investigate the problem of mitigating interference in multiuser MIMO-OFDM systems employing spatial multiplexing [39], in which transmit antennas on one or more users send independent data streams in parallel. Spatial multiplexing is important because it leverages simultaneous communication without loss of spectral efficiency, creating multiple communication links to support high data rate communications. We propose a general two-stage interference cancellation framework for this particular system. The first stage uses time domain equalization to suppress the co-channel interference, mitigate asynchronism, and shorten the post-equalization channel response to be no longer than the length of the cyclic prefix. The second stage performs low-complexity per-tone single tap equalization and detection in the frequency domain. A major advantage of the proposed two-stage equalization scheme is that it can handle many sources of interference in the system such as ICI arising from the asynchronism or the channel, CCI coming from multiple users, co-antenna interference, and interference from jammers. Because each data stream is detected independently, it is possible to compensate for the timing error caused by asynchronism with the equalization delay obtained from the training to effectively align different data streams, and thus achieve synchronicity between users. The framework is developed specifically for spatial multiplexing and is applied to multiuser MIMO-OFDM and multiuser SISO-OFDM systems with asynchronism between users as well as to single-user MIMO-OFDM systems.

We propose various equalizer design methods that determine the parameters of the receiver directly from the training data and we compare them with methods that derive the coefficients from channel estimates [13] [14]. We formulate and solve a joint optimization for the space-time equalizer and the shortened post-equalization channel, by maximizing a measure of signal to interference-plus-noise ratio (SINR) in the frequency domain subject to different norm constraints. An advantage of our approach over that in [13] [14] is that it does not require channel estimation and instead derives the coefficients directly from the training data, which is useful when the training of the interferer in the system is unknown.

Our approach is similar to the space-time two-stage hybrid approach in [37] [38] applied to simplify the Viterbi decoder in synchronous single carrier systems. We set up the optimization problem in the frequency domain, however, since performance is measured in the frequency domain in OFDM systems. This also allows flexible selection of subsets of OFDM tones. We propose several variations to the original maximizing SINR formulation, which either reduce computation or improve the performance. Further, our approach deals with asynchronism between users, which [37] [38] do not. This is especially important since OFDM systems employ special block-structured transmission, and are sensitive to such asynchronism.

Compared with conventional approaches for MIMO equalization in the time domain [34] [33] [40], our proposed scheme provides a low computational complexity *per-stream* solution, i.e., it conducts timing synchronization and interference suppression for each transmitted data stream. The conventional approaches [34] [33], though, lead to more complex joint synchronization and detection. Compared with per-tone equalization [30], the time domain equalizer as we considered in this paper requires much lower design complexity and much smaller memory size.

In summary, the purpose of our two-stage framework is to suppress multiple access interference from asynchronous sources and co-antenna interference, to shorten the channel to permit simplified equalization in the frequency domain and to preserve diversity in the channel that can be extracted using coding and interleaving. We present simulation results in a variety of interference scenarios, with and without multiple users, including known and unknown training data to illustrate the performance of our proposed receiver.

The rest of the paper is organized as follows. In Section II, we introduce the MIMO-OFDM multiuser system under consideration, and then describe the space-time data model, the space-frequency model and the detection and decoding scheme in this system. In Section III, we discuss the equalizer design method with channel estimates. In Section IV, we propose a direct training method for maximizing the SINR and assorted variations. We rationalize the cost functions and derive solutions for the optimization problems. Simulation results are presented in Section V, where we illustrate the BER and SINR performance of the schemes in block fading channels. We summarize the major results of this paper in Section VI.

II. Space-time/Space Frequency Data Model

In this section, we will first describe the MIMO-OFDM multiuser system, and then present the channel model and the space-time/space frequency data model. We will discuss the detection and decoding scheme employed in the system in the last part of this section.

A. System Overview

Fig. 2 illustrates the uplink of the multiuser MIMO-OFDM system under consideration. The base station, which has M_r receive antennas, is receiving signals from multiple active mobile units in the system. We number all the active transmit antennas of the mobile units in the system from 1 to M_t, and assume that the number of active antennas in the system, denoted by M_t, is less than or equal to M_r. This creates an effective $M_t \times M_r$ MIMO communication system. The desired antennas are those with coordinated training, i.e., the distinct training data from each desired antenna is known to the receiver, and the rest of the active antennas out of the total M_t antennas are considered as interferers. Such coordination of assigning distinct training sequences to different users can be achieved via the operations of higher level protocols, which we will investigate in future work.

We assume that each user employs spatial multiplexing [39]. Thus the data is demultiplexed into parallel bit streams, coded with forward error correction coding, and then passed to an OFDM modulator. After OFDM modulation, each data stream is transmitted by one transmit antenna. With some additional work, coding and interleaving before demultiplexing or space-time coding can be included, however, we defer this to future research. Each OFDM symbol of the data streams has N tones, and the cyclic prefix length is L_{CP}. Depending on the requirement of spectral shaping, not all tones are loaded with data [41] [42]. To simplify the notation, we use a $N \times N$ diagonal matrix $\mathbf{T}$, whose diagonal elements can only be 0 or 1, to select those tones that are actively used for data transmission.

The proposed receiver with two-stage equalization structure for this system is illustrated in Fig. 3. A space-time equalizer bank is placed at the frontend for channel shortening and interference suppression. The outputs of the space-time equalizer are data streams corresponding to the active transmit antennas. For each output data stream, OFDM demodulation is conducted independently based on the shortened impulse response for each substream. Extending this to joint modulation along the lines of [43]–[45] is of interest but is beyond the scope of this work. After discarding the cyclic prefix, the DFT operation is applied to the equalized signals. Then, in the frequency domain, equalization and detection is conducted for the signals after the DFT using the post-equalization channel response. The coefficients of the space-time filter and the post-equalization channel responses are derived from the training transmitted from the different antennas using algorithms to be described.

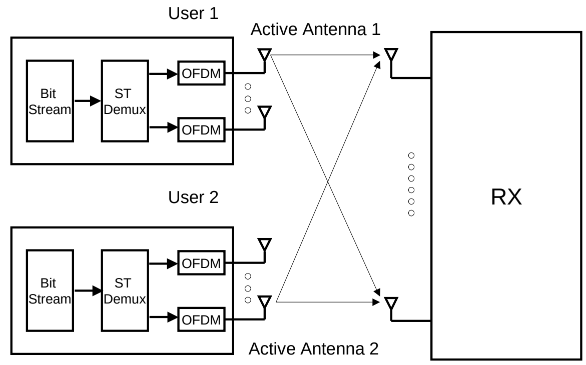


Fig. 2. The MIMO-OFDM multiuser cellular system with spatial multiplexing scheme ($M_t = 2$ and $M_r \geq 2$).

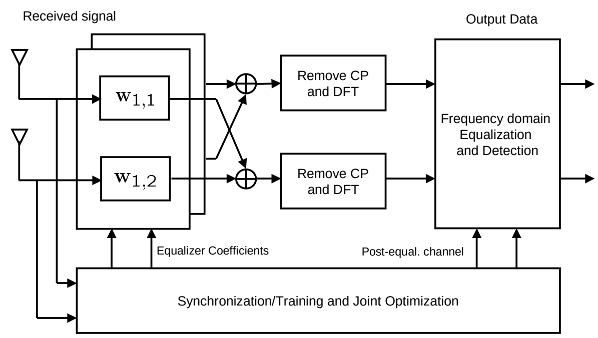


Fig. 3. The block diagram of the receiver structure shown for two RX antennas. The two stage equalization scheme includes time domain equalization and frequency domain equalization and detection with equalizer sets $\mathbf{w}_1 = \{\mathbf{w}_{1,1}, \mathbf{w}_{1,2}\}$ and $\mathbf{w}_2 = \{\mathbf{w}_{2,1}, \mathbf{w}_{2,2}\}$ for the two input data streams.

We denote the l^{th} tap of the space time equalizer at the j^{th} receive antenna for the i^{th} output data stream of the space-time equalizers by $w_{i,j}(l)$, where $l = 0, 1, ..., L$, and L is the order of the equalizer and $i = 1, 2, ..., M_t$. We can write it in a vector form as $\mathbf{w}_{i,j} = [w_{i,j}(L), ..., w_{i,j}(0)]^T$, where $(\cdot)^T$ denotes transpose. The equalizer set $\{\mathbf{w}_{i,1}, ..., \mathbf{w}_{i,M_r}\}$ forms a space-time filter bank for the i^{th} output data stream. Each FIR equalizer $\mathbf{w}_{i,j}$ may have a different equalization decision delay $\Delta_{i,j}$. Thus the equalizer $\mathbf{w}_{i,j}$ "smoothes" the received signal samples from time $k + \Delta_{i,j}$ to $k + \Delta_{i,j} - L$ to output the filtered scalar signal at time k, where k is the common reference time in the system.

The frequency domain signal of the b^{th} OFDM symbol of the m^{th} data stream is denoted by $\mathbf{v}_m(b) = [v_m(0, b), ..., v_m(N-1, b)]^T$. The time domain signal is defined as $\mathbf{s}_m(b) = \mathbf{Q}\mathbf{v}_m(b)$, where $\mathbf{Q}$ is the DFT basis matrix, and $\mathbf{s}_m(b) = [s_m(0, b), ..., s_m(N-1, b)]^T$. For the sample $s_m(p, b)$, the relation between b, p and k is that $k = b(N + L_{CP}) + p + L_{CP}$.

TABLE I

SUMMARY OF KEY SYMBOLS USED IN THE PAPER AND THEIR SIZES

Symbol	Name	Size
L	Equalizer Memory Length	Scalar
L_{CP}	Cyclic Prefix Length	Scalar
ν	Channel Memory Length	Scalar
$\Delta_{i,j}$	Equalization Decision Delay	Scalar
$\mathbf{w}_{i,j}$	Equalizer Vector	$(L+1)\times 1$
$\mathbf{H}_{j,m}$	Channel Matrix	$(L+1)\times(L+\nu+1)$
$\mathbf{y}_i(b)$	Signal Vector After DFT	$N\times 1$
$\mathbf{R}_j(b,\Delta_{i,j})$	Time Domain Signal Matrix with Delay Parameter $\Delta_{i,j}$	$N\times(L+1)$
$\mathbf{V}_m(b)$	Frequency Domain Signal Matrix	$N\times N$
$\mathbf{b}_{i,i}$	Time Domain Shortened Channel	$(L_{CP}+1)\times 1$
$\mathbf{e}_i(b)$	Noise and Interference Vector	$N\times 1$
$\bar{\mathbf{Q}}$	Partial DFT Basis Matrix	$N\times(L_{CP}+1)$
$\mathbf{T}$	Tone Selection Matrix	$N\times N$

B. Channel Model

The channel is assumed to be frequency selective and block-wise time invariant over the duration of multiple OFDM symbols. We denote the channel response by $h_{j,m}(l)$ for the l^{th} tap of the channel between the m^{th} transmit antenna and the j^{th} receive antenna, where $l=0,1,2...,\nu$, $m=1,2,...,M_t$, and $j=1,2,...,M_r$. Note that $\nu+1$ is the length of the channel response between the m^{th} transmit antenna and the j^{th} receive antenna. Also, note that the index m denotes the m^{th} transmit antenna of all M_t active transmit antennas in the system. We denote the absolute propagation delay of the link between the m^{th} transmit antenna and the j^{th} receive antenna by $d_{j,m}$, where the delays between the m^{th} transmit antenna and the j^{th} receive antenna are different in general. Since we consider the case where the transmissions from different users are not coordinated, the propagation delays $d_{j,m}$'s are unknown to the receiver. We model the effects of propagation delays as part of the channel response, therefore, the first $d_{j,m}$ taps of the channel response are all zeros. We make the assumption that ν is less than the number of OFDM tones N, but ν may be less than or greater than L_{CP}.

C. Signal Model

Since the first stage of our receiver will be a time-domain equalization, we will form a stacked channel matrix $\mathbf{H}_{j,m}$ of size $(L+1)\times(L+\nu+1)$ between the m^{th} transmit antenna and the j^{th} receive antenna, which is pre-multiplied by the equalizer $\mathbf{w}_{i,j}$ as in (1), shown at the bottom of the next page.

Let $s_m(k)$ be the signal of the data stream transmitted by the m^{th} antenna at discrete time k. We stack $L+\nu+1$ (the length of the effective channel response after equalization) samples of the transmitted signal into a vector $\mathbf{s}_m(k-L-\nu:k) = [s_m(k-L-\nu),...,s_m(k)]^T$, where $p:q$ denotes the p^{th} to the q^{th} sample of the signal in MATLAB notation. The signal vector at the j^{th} receive antenna, denoted by $\mathbf{r}_j(k-L:k)$, is given as,

$$\mathbf{r}_j(k-L:k) = \sum_{m=1}^{M_t}\mathbf{H}_{j,m}\mathbf{s}_m(k-L-\nu:k) + \mathbf{n}_j(k-L:k) \tag{2}$$

where $\mathbf{n}_j(k-L:k) = [n_j(k-L),...,n_j(k)]^T$, which is the noise vector at the j^{th} receive antenna, and $n_j(k)$'s follow an i.i.d. complex Gaussian distribution with zero mean and variance σ_n^2.

In OFDM systems, training is usually performed by sending known symbols in the frequency domain, consequently, we first introduce the space-time data model, and then derive the space-frequency model based on this space-time data model.

a) Space-Time Model: The k^{th} sample of the i^{th} output data stream after the space-time equalization, denoted by $x_i(k)$, can be written as

$$\begin{aligned} x_i(k) &= \sum_{j=1}^{M_r}\mathbf{w}_{i,j}^T\mathbf{r}_j(k+\Delta_{i,j}-L:k+\Delta_{i,j}) \qquad (3)\\ &= \sum_{j=1}^{M_r}\sum_{m=1}^{M_t}\mathbf{s}_m(k+\Delta_{i,j}-L-\nu:k+\Delta_{i,j})^T\mathbf{H}_{j,m}^T \\ &\times \mathbf{w}_{i,j} + \sum_{j=1}^{M_r}\mathbf{n}_j(k+\Delta_{i,j}-L:k+\Delta_{i,j})^T\mathbf{w}_{i,j} \end{aligned}$$

where $\mathbf{w}_{i,j}$ is the equalizer for the i^{th} output data stream at the j^{th} receive antenna. Note that in general, the equalization decision delay for the i^{th} output data stream can be different for each receive antenna, which we denote by $\Delta_{i,j}$ for the j^{th} receive antenna and the i^{th} output data stream.

b) Space-Frequency Model: The derivation of the space-frequency model based on (3) is given in Appendix I. We summarize the final space-frequency model as follows

$$\begin{aligned} \mathbf{y}_i(b) &= \sum_{j=1}^{M_r}\mathbf{Q}\mathbf{R}_j(b,\Delta_{i,j})\mathbf{w}_{i,j} \\ &= \mathbf{V}_i(b)\underbrace{\bar{\mathbf{Q}}\mathbf{b}_{i,i}}_{\mathbf{f}_{i,i}}+\mathbf{e}_i(b) \qquad (4) \end{aligned}$$

where $\mathbf{y}_i(b)$ is a $N\times 1$ vector, which has the samples of the b^{th} OFDM symbol of the i^{th} output data stream, and $\mathbf{f}_{i,i}$ is the channel response in the frequency domain for the i^{th} data stream. We define $\mathbf{R}_j(b,\Delta_{i,j})$ in (A.I.8), which is the stacked version of the data matrix at the j^{th} receive antenna with delay parameter $\Delta_{i,j}$. We define $\mathbf{V}_i(b) = \mathrm{diag}\{v_i(0,b),...,v_i(N-1,b)\}$, which has the b^{th} frequency domain samples of the data stream transmitted by the i^{th} active transmit antenna. The partial DFT basis is denoted by $\bar{\mathbf{Q}}$, defined in (A.I.12), and $\mathbf{b}_{i,i}$ is the shortened channel response of size $(L_{CP}+1)\times 1$

for the effective channel of the i^{th} output data stream, defined in (A.I.13). The quantity $\mathbf{e}_i(b)$ (defined in A.I.14) is the noise and interference term, which includes the interference due to imperfect channel shortening, residual co-channel interference, and the colored noise.

We list the key scalars, vectors and matrices used in the paper and their sizes in Table I. With these definitions, the objective of this paper is to design the set of equalizers $\{\mathbf{w}_{i,j}\}$ and the set of post-equalization channel responses $\mathbf{b}_{i,i}$ ($i = 1, 2, ..., M_t$, and $j = 1, 2, ..., M_r$) based on the training sequences which are assumed known to the receivers.

D. Detection and Decoding Scheme

We stack the channel matrices between all the active transmit and receive antennas into a matrix $\boldsymbol{\mathcal{H}}$, where

$$\boldsymbol{\mathcal{H}} = \begin{bmatrix} \mathbf{H}_{1,1} & ... & \mathbf{H}_{1,M_t} \\ \mathbf{H}_{2,1} & ... & \mathbf{H}_{2,M_t} \\ \vdots & \ddots & \vdots \\ \mathbf{H}_{M_r,1} & ... & \mathbf{H}_{M_r,M_t} \end{bmatrix}. \tag{5}$$

The equalizers on different antennas are stacked into a vector $\mathbf{w}_i$, where $\mathbf{w}_i = [\mathbf{w}_{i,1}, ..., \mathbf{w}_{i,M_r}]^T$. We also define a matrix $\mathbf{\Gamma}_i$ as the binary diagonal matrix which selects the samples in a certain window of length $L_{CP} + 1$ of the effective channel response after space-time filtering for the i^{th} active transmit antenna.

Given the optimal space-time filter $\mathbf{w}_i^{opt}$, the effective channel response for the i^{th} stream is simply $\mathbf{b}_{i,i}^{opt} = \mathbf{\Gamma}_i^{opt} \boldsymbol{\mathcal{H}}^T \mathbf{w}_i^{opt}$, where $(\cdot)^{opt}$ denotes the optimal solution. With the post-equalization channel response $\mathbf{b}_{i,i}^{opt}$, after the DFT operation, a per frequency tone equalization and detection scheme can be applied. Based on (4), the system equation is given as

$$y_i(k, b) = f_{i,i}(k) v_i(k, b) + e_i(k, b) \tag{6}$$

where (k, b) stands for the k^{th} sample of the b^{th} OFDM symbol. We have $\mathbf{f}_{i,i} = \bar{\mathbf{Q}} \mathbf{b}_{i,i}^{opt}$, and $\mathbf{f}_{i,i} = [f_{i,i}(1), ..., f_{i,i}(N)]^T$, which is the effective channel response in the frequency domain. The error term $e_i(k, b)$ contains the residual signal after the channel shortening and the noise.

For uncoded systems, to simplify the receiver structure, a ML (Maximum likelihood) type detection scheme is adopted here as

$$v_i(k, b)^{opt} = \arg\min_{v_i(k,b) \in \mathcal{A}} \|y_i(k, b) - f_{i,i}(k) v_i(k, b)\|^2 \tag{7}$$

where $\mathcal{A}$ denotes the constellation set. Notice that this method is suboptimal since the error term contains the residual error due to the channel shortening, thus it is not white Gaussian noise. But the detection scheme is simple and offers very good performance in simulations. When convolutional coding across tones is applied, we use the Euclidean metric in (7) to calculate the soft information for Viterbi decoding [13]–[15], [46]–[48].

Let $v_i(u, k, b)$ be the u^{th} bit of the sample at the k^{th} tone of b^{th} OFDM symbol of the i^{th} transmitted data stream. The log likelihood ratio for this bit can be approximated as the following by using the max-log approximation [49]–[51]

$$\begin{aligned} \Lambda_i(u, k, b) &= \min_{\substack{v_i(k,b) \in \mathcal{A}, \\ v_i(u,k,b)=1}} \frac{|y_i(k, b) - f_{i,i}(k) v_i(k, b)|^2}{2\sigma_i(k)^2} \\ &- \min_{\substack{v_i(k,b) \in \mathcal{A}, \\ v_i(u,k,b)=0}} \frac{|y_i(k, b) - f_{i,i}(k) v_i(k, b)|^2}{2\sigma_i(k)^2}. \end{aligned} \tag{8}$$

Let us define a matrix $\mathbf{W}_{i,j}$ of size $N \times (N + L)$ as

$$\mathbf{W}_{i,j} = \begin{bmatrix} w_{i,j}(L) & ... & w_{i,j}(0) & ... & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & ... & w_{i,j}(L) & ... & w_{i,j}(0) \end{bmatrix}. \tag{9}$$

The noise covariance matrix after the space-time equalization and DFT operation can be obtained as

$$\mathbf{G}_i = \sigma_n^2 \sum_{j=1}^{M_r} \mathbf{Q} \mathbf{W}_{i,j} \mathbf{W}_{i,j}^H \mathbf{Q}^H. \tag{10}$$

Hence the noise power on tone k for the i^{th} output data stream can be represented as $\sigma_i(k)^2 = \mathbf{G}_i(k, k)$, i.e., the k^{th} diagonal element of the matrix $\mathbf{G}_i$. The effect of different SINR levels at different OFDM tones has been taken into account in this log likelihood ratio formulation. We use soft-decision Viterbi decoding algorithm to decode the data bits [51] [52]. This is not the optimal decoding algorithm, but has low complexity and performs robustly in our simulations.

III. Channel Estimate Based Equalizer Design

One approach for designing the space-time channel shortening filter is to estimate the channel first and then design the shortening filter with the estimated channel [14] [34] [32]. For the purpose of comparison with the training-based algorithms in simulations, we discuss equalizer design method based on channel estimates in this section.

To estimate the stacked channel matrix in (5), all the training sequences of the active antennas in the system have to be known to the receiver. Based on (A.I.8), we stack all the training sequence matrices of the active antennas into a matrix $\mathbf{S} = [\mathbf{S}_1, ..., \mathbf{S}_{M_t}]$, the received signals at the receive antennas into a matrix $\mathbf{R} = [\mathbf{R}_1, ..., \mathbf{R}_{M_r}]$, and the noise matrices at the receive antennas into a matrix $\mathbf{N} = [\mathbf{N}_1, ..., \mathbf{N}_{M_r}]$. We assume that the training sequences are designed such that the matrix $\mathbf{S}$ be a tall and full-rank matrix, thus all the columns of this matrix are linearly independent [53]. Notice that we drop

$$\mathbf{H}_{j,m} = \begin{bmatrix} h_{j,m}(\nu) & h_{j,m}(\nu-1) & ... & h_{j,m}(0) & 0 & 0 & ... & 0 \\ 0 & h_{j,m}(\nu) & h_{j,m}(\nu-1) & ... & h_{j,m}(0) & 0 & ... & 0 \\ \vdots & \vdots & \ddots & \ddots & \ddots & \ddots & \vdots & \vdots \\ 0 & 0 & ... & 0 & h_{j,m}(\nu) & h_{j,m}(\nu-1) & ... & h_{j,m}(0) \end{bmatrix} \tag{1}$$

the indices b and $\Delta_{i,j}$ here, since the time domain channel estimation is independent of the block structure of the OFDM modulation.

The system equation is given as

$$\mathbf{R} = \mathbf{S}\boldsymbol{\mathcal{H}}^T + \mathbf{N} \tag{11}$$

and the least square channel estimate can be obtained as

$$\hat{\boldsymbol{\mathcal{H}}} = ((\mathbf{S}^H\mathbf{S})^{-1}\mathbf{S}^H\mathbf{R})^T. \tag{12}$$

Least square channel estimation is assumed since it is the maximum likelihood estimation for Gaussian signals in the cases of both white noise and colored noise when the noise covariance is unknown to the receiver [53].

Given the channel estimates in (12), we state the problem of maximizing the signal to interference ratio (henceforth the MaxPwr method introduced in [14]) as

$$\max_{\mathbf{w}_i,\mathbf{\Gamma}_i} \frac{\mathbf{w}_i^H\hat{\boldsymbol{\mathcal{H}}}^*\mathbf{\Gamma}_i^T\mathbf{\Gamma}_i\hat{\boldsymbol{\mathcal{H}}}^T\mathbf{w}_i}{\mathbf{w}_i^H\hat{\boldsymbol{\mathcal{H}}}^*\bar{\mathbf{\Gamma}}_i^T\bar{\mathbf{\Gamma}}_i\hat{\boldsymbol{\mathcal{H}}}^T\mathbf{w}_i} \quad s.t. \quad \|\mathbf{w}_i\|^2 = 1, \tag{13}$$

where the matrix $\mathbf{\Gamma}_i$ is the binary diagonal matrix which selects the samples in a certain window of length $L_{CP}+1$ of the effective channel response after space-time filtering for the i^{th} active transmit antenna. The matrix $\bar{\mathbf{\Gamma}}_i$ selects the samples outside the window, and $(\cdot)^*$ stands for complex conjugate. Thus by performing the optimization over $\mathbf{w}_i$ and $\mathbf{\Gamma}_i$, the energy of the post equalization channel response will be mainly concentrated in the selected window of length $L_{CP}+1$, and the energy outside the selected window is minimized.

IV. Direct Training Based Equalizer Design and Its Variations

To apply the MaxPwr method to design the space-time equalizer, a channel estimate is required. This is difficult when the training sequences are not all known to the receiver. Also, this method may suffer from noise propagation due to imperfect channel estimates. To overcome this problem, we apply a *direct* approach of obtaining the equalizer coefficients and the post-equalization channel response by a joint optimization of maximizing the SINR. We discuss this method and its assorted variations in this section.

A. Maximizing SINR with Independent Norm Constraint (INC)

In the MIMO-OFDM system under consideration, each active antenna transmits an independent data stream, thus we need to design a space-time equalizer for each transmitted stream separately to suppress the interference. We formulate the problem of obtaining the equalizer coefficients, the post equalization channel response, and the equalization decision delay using training sequences as maximizing the SINR subject to some norm constraint on the post-equalization channel response or the equalizer coefficients. The norm constraint is used to avoid degenerate solutions in the optimization. We obtain the equalizer coefficients and the post-equalization channel response jointly through the optimization. A similar idea has been presented in [37], and we extend it to the MIMO-OFDM asynchronous multiuser system here.

Based on the space-frequency model (4), the joint optimization problem is stated as follows.

Problem 1 (Optimization for Independent Norm Constraint): For $i = 1, 2, ..., M_t$, find the optimal solutions for $\{\mathbf{w}_{i,j}\}_{j=1}^{M_r}$, $\mathbf{b}_{i,i}$ and Δ_i (the parameters of the i^{th} output stream of the equalizer bank) by maximizing the cost function

$$\max_{\mathbf{w}_{i,j},\mathbf{b}_{i,i},\Delta_i} \frac{\sum_{b=1}^{K}\|\mathbf{T}\mathbf{V}_i(b)\bar{\mathbf{Q}}\mathbf{b}_{i,i}\|^2}{\sum_{b=1}^{K}\|\mathbf{T}[\sum_{j=1}^{M_r}\mathbf{Q}\mathbf{R}_j(b,\Delta_i)\mathbf{w}_{i,j}-\mathbf{V}_i(b)\bar{\mathbf{Q}}\mathbf{b}_{i,i}]\|^2}$$
$$\text{subject to the norm constraint } \|\mathbf{b}_{i,i}\|^2 = 1 \tag{14}$$

with the known data parameters $\mathbf{T}$ (a binary selection matrix), $\mathbf{V}_i(b)$ and $\mathbf{R}_j(b,\Delta_i)$, where K is the number of training OFDM symbols.

Since the propagation delays between one transmit antenna and the receive antennas at the receiver are very close, in the optimization, we consider the equalization delay $\Delta_{i,j}$ for the i^{th} output data stream at the j^{th} receive antenna to be the same for all receive antennas, i.e., $\Delta_{i,j} = \Delta_i$ for $\forall\ j$ to simplify the computation of the optimization.

B. Rationale for the Cost Function

Problem 1 is based on the idea of maximizing the signal to interference and noise ratio. In this subsection, we provide the rationale to the cost function, and justify the main result that the optimal solution to such a formulation is consistent with the optimal solution of MaxPwr method with perfect channel information under certain statistical assumptions. In the next several subsections, we will propose several variations to this problem, of either maximizing SINR or minimizing mean square error, which can be justified to be consistent with the MaxPwr formulation in a similar fashion, and these variations may have advantages in either complexity or performance.

We make the following statistical assumptions about the transmitted signal and the noise for asymptotic analysis of our proposed algorithms.

Assumption 2: The frequency domain samples of the training data streams sent by the active transmit antennas, denoted by $v_m(p,b)$'s, are independent, identically distributed, and circularly symmetric complex random variables with mean zero and variance σ_v^2. The additive noise at the receive antennas is white Gaussian with zero mean and variance σ_n^2.

Proposition 3: The time domain samples $s_m(p,b)$ of the training data streams (disregarding the cyclic prefix) sent by the active transmit antennas are uncorrelated with zero mean and variance σ_v^2.

Proof: The result follows directly from the fact that $\mathbf{Q}$ is a unitary matrix, and the uncorrelated property of the signal vector is maintained by the unitary transformation. ■

With Assumption 2 and Proposition 3, when the number of training symbols K is large enough, asymptotically summations can be replaced by expectations, and the cost function can be approximated as

$$f_i(\mathbf{w}_{i,j}, \mathbf{b}_{i,i}, \Delta_i) = \frac{E\|\mathbf{T}\mathbf{V}_i(b)\bar{\mathbf{Q}}\mathbf{b}_{i,i}\|^2}{E\|\mathbf{T}[\sum_{j=1}^{M_r}\mathbf{Q}\mathbf{R}_j(b,\Delta_i)\mathbf{w}_{i,j}-\mathbf{V}_i(b)\bar{\mathbf{Q}}\mathbf{b}_{i,i}]\|^2} \tag{15}$$

where E denotes expectation, and the function f_i represents the ratio of the mean signal power and the mean noise and interference power.

The consistency property of the formulation in (15) is summarized in the following proposition.

Proposition 4 (Asymptotic Analysis Result for $\mathbf{T} = \mathbf{I}$*):* The cost function f_i can be simplified as the ratio of

$$\text{the numerator} = \|\mathbf{b}_{i,i}\|^2 \tag{16}$$

and

$$\text{the denominator} = \|\boldsymbol{\mathcal{H}}^T \mathbf{w}_i - \bar{\mathbf{C}}_i \mathbf{b}_{i,i}\|^2 + \frac{\sigma_n^2}{\sigma_v^2}\|\mathbf{w}_i\|^2, \tag{17}$$

where $\bar{\mathbf{C}}_i$ has size $M_t(L+\nu+1) \times (L_{CP}+1)$ and is a binary selection matrix defined as

$$\begin{aligned}\bar{\mathbf{C}}_i \quad = \quad & [\mathbf{0}^T_{(L+\nu+1)(i-1)\times(L_{CP}+1)}, \mathbf{C}_i(\Delta_i)^T, \\ & \mathbf{0}^T_{(L+\nu+1)(M_t-i)\times(L_{CP}+1)}]^T\end{aligned} \tag{18}$$

which selects $L_{CP} + 1$ consecutive columns of the matrix $\mathbf{H}$.

Proof: The proof is given in the Appendix II. ■

With the norm constraint on $\mathbf{b}_{i,i}$, the optimal $\mathbf{w}_i$ and $\mathbf{b}_{i,i}$ according to Proposition 3 can be found by jointly minimizing the denominator of the cost function, such as

$$\begin{aligned}(\mathbf{w}_i^{opt}, \mathbf{b}_{i,i}^{opt}) \quad = \quad & \arg\min_{\mathbf{w}_i, \mathbf{b}_{i,i}, \bar{\mathbf{C}}_i} (\|\boldsymbol{\mathcal{H}}^T \mathbf{w}_i - \bar{\mathbf{C}}_i \mathbf{b}_{i,i}\|^2 + \frac{\sigma_n^2}{\sigma_v^2}\|\mathbf{w}_i\|^2) \\ & \text{s.t.} \quad \|\mathbf{b}_{i,i}\|^2 = 1.\end{aligned} \tag{19}$$

The first term denotes the shortening error and the second term characterizes the ratio of the noise power enhancement to the signal power. *Therefore, asymptotically, this method is the same as the shortening SNR maximization formulation when the channel is known to the receiver.* Thus with enough training, the solution should approach the perfect channel knowledge solution.

Proposition 5 (Asymptotic Analysis Result for $\mathbf{T} \neq \mathbf{I}$*):* The cost function f_i can be simplified as the ratio of

$$\text{the numerator} = \|\mathbf{T}\bar{\mathbf{Q}}\mathbf{b}_{i,i}\|^2, \quad \text{and} \tag{20}$$

$$\begin{aligned}\text{the denominator} \quad = \quad & \|\mathbf{T}\hat{\mathbf{Q}}(\Delta_i)\boldsymbol{\mathcal{H}}^T \mathbf{w}_i - \mathbf{T}\hat{\mathbf{Q}}(\Delta_i)\bar{\mathbf{C}}_i \mathbf{b}_{i,i}\|^2 \\ + \quad & \frac{\sigma_n^2}{\sigma_v^2}\|\hat{\mathbf{T}}\mathbf{Q}_w \mathbf{w}_i\|^2 + \delta(\boldsymbol{\mathcal{H}}^T \mathbf{w}_i, \mathbf{b}_{i,i})\end{aligned} \tag{21}$$

where $\tilde{\mathbf{Q}}(\Delta_{i,j})$ has size $N \times (L+\nu+1)$ and is a partial DFT basis, defined in (22), shown at the bottom of the next page, and $\hat{\mathbf{Q}}(\Delta_i) = [\tilde{\mathbf{Q}}(\Delta_i), ..., \tilde{\mathbf{Q}}(\Delta_i)]$, $\hat{\mathbf{T}} = \mathbf{I}_{M_r} \otimes \mathbf{T}$, where $\otimes$ is the Kronecker product, $\mathbf{Q}_w$ is the stacked partial DFT basis matrix matching the equalizer vector at the receive antennas defined as

$$\mathbf{Q}_w = \mathbf{I}_{M_r} \otimes \begin{bmatrix} 1 & \ldots & 1 \\ e^{-j2\pi L/N} & \ldots & 1 \\ \vdots & \ddots & \vdots \\ e^{-j2\pi L(N-1)/N} & \ldots & 1 \end{bmatrix}, \tag{23}$$

and $\delta(\boldsymbol{\mathcal{H}}^T \mathbf{w}_i, \mathbf{b}_{i,i})$ characterizes the residual error which is defined in (A.II.19).

The sketch of the proof is given in the Appendix II. Essentially, the idea here is still maximizing the shortening SINR, but the difference is that the matrix $\mathbf{T}$ selects the bins of the frequency domain channel responses. The numerator is the signal power in the frequency domain summing over the selected tones, the first term of the denominator is the shortening error and the last term δ characterizes the residual error introduced by the block structure of the OFDM modulation. Note that $\hat{\mathbf{Q}}(\Delta_i)\boldsymbol{\mathcal{H}}^T \mathbf{w}_i$ and $\bar{\mathbf{Q}}\mathbf{b}_{i,i}$ are the frequency domain channel responses.

C. Solution to the Problem of Maximizing SINR with INC

For Problem 1, we introduce the optimal solution in this subsection. Let us define a matrix $\mathbf{D}_i$ as

$$\mathbf{D}_i = \begin{bmatrix} \mathbf{T}\mathbf{V}_i(1)\bar{\mathbf{Q}} \\ \vdots \\ \mathbf{T}\mathbf{V}_i(K)\bar{\mathbf{Q}} \end{bmatrix}. \tag{24}$$

Also define $\mathbf{P}(\Delta_i)$ with parameter Δ_i as

$$\mathbf{P}(\Delta_i) = \begin{bmatrix} \mathbf{TQR}_1(1,\Delta_i) & \ldots & \mathbf{TQR}_{M_r}(1,\Delta_i) \\ \vdots & \ddots & \vdots \\ \mathbf{TQR}_1(K,\Delta_i) & \ldots & \mathbf{TQR}_{M_r}(K,\Delta_i) \end{bmatrix} \tag{25}$$

and $\mathbf{w}_i = [\mathbf{w}_{i,1}^T, ..., \mathbf{w}_{i,M_r}^T]^T$. The problem is restated as

$$\begin{aligned}& \max_{\mathbf{w}_i, \mathbf{b}_{i,i}, \Delta_i} \frac{\|\mathbf{D}_i \mathbf{b}_{i,i}\|^2}{\|\mathbf{P}(\Delta_i)\mathbf{w}_i - \mathbf{D}_i \mathbf{b}_{i,i}\|^2} \\ & \text{s.t.} \quad \|\mathbf{b}_{i,i}\|^2 = 1\,.\end{aligned} \tag{26}$$

Since $\mathbf{w}_i$ is not constrained, given $\mathbf{b}_{i,i}$, $\mathbf{w}_{i,i}^{opt}$ can be obtained by minimizing the denominator of the cost function. Thus the optimal solution for $\mathbf{w}_i^{opt}$ in terms of $\mathbf{b}_{i,i}^{opt}$ is given as follows

$$\mathbf{w}_i^{opt} = (\mathbf{P}(\Delta_i^{opt})^H \mathbf{P}(\Delta_i^{opt}))^{-1}\mathbf{P}(\Delta_i^{opt})^H \mathbf{D}_i \mathbf{b}_{i,i}^{opt}, \tag{27}$$

and $\mathbf{b}_{i,i}^{opt}$ corresponds to the smallest eigenvalue of the matrix $\mathbf{M}_i$, where

$$\begin{aligned}\mathbf{M}_i \quad = \quad & \mathbf{I} - (\mathbf{D}_i^H \mathbf{D}_i)^{-1}\mathbf{D}_i^H \mathbf{P}(\Delta_i^{opt})(\mathbf{P}(\Delta_i^{opt})^H \mathbf{P}(\Delta_i^{opt}))^{-1} \\ \times \quad & \mathbf{P}(\Delta_i^{opt})^H \mathbf{D}_i\,.\end{aligned} \tag{28}$$

We must have $K\text{rank}(\mathbf{T}) \geq M_r(L+1)$ as a necessary condition of the matrix $\mathbf{D}_i^H \mathbf{D}_i$ being invertible.

Note that the eigenvalues of $\mathbf{M}_i$ are all real. The Δ_i^{opt} can be found by an iterative search, and in general this approach improves performance. We should also notice that equalization decision delays are used to compensate for the different propagation delays for all data streams, and by doing so, synchronization can be effectively achieved jointly with the interference suppression and channel shortening.

D. MMSE Method

We can reduce the computation of the INC method by applying the MMSE method which minimizes the power of the noise and interference term with the norm constraint on

$\mathbf{b}_{i,i}$. Based on the rationale of the cost function in (19), the MMSE optimization is stated as

$$\max_{\mathbf{w}_i,\mathbf{b}_{i,i},\Delta_i} \|\mathbf{P}(\Delta_i)\mathbf{w}_i - \mathbf{D}_i\mathbf{b}_{i,i}\|^2 \quad \text{s.t.} \quad \|\mathbf{b}_{i,i}\|^2 = 1\,.$$

The optimal solution can be derived as

$$\mathbf{w}_i^{opt} = (\mathbf{P}(\Delta_i^{opt})^H\mathbf{P}(\Delta_i^{opt}))^{-1}\mathbf{P}(\Delta_i^{opt})^H\mathbf{D}_i\mathbf{b}_{i,i}^{opt}, \tag{29}$$

and $\mathbf{b}_{i,i}^{opt}$ is the normalized eigenvector corresponding to the smallest eigenvalue of the matrix $\mathbf{K}_i$, where

$$\mathbf{K}_i = \mathbf{I} - \mathbf{D}_i^H\mathbf{P}(\Delta_i^{opt})(\mathbf{P}(\Delta_i^{opt})^H\mathbf{P}(\Delta_i^{opt}))^{-1}\mathbf{P}(\Delta_i^{opt})^H\mathbf{D}_i \tag{30}$$

where Δ_i^{opt} can be found by an iterative search. This method is computationally easier to implement than the INC method since computation of the inverse of the matrix $\mathbf{D}_i^H\mathbf{D}_i$ is not required, and it offers similar performance compared to the INC method in simulations. MMSE methods also lead to adaptive solutions, e.g. [10] [54], which is a topic for future research.

E. Independent Norm Constraint on $\mathbf{w}_i$ (WINC)

Since the vector $\mathbf{w}_i$ has a higher dimension than the vector $\mathbf{b}_{i,i}$, i.e., $M_r(L+1) > L_{CP}+1$, more degrees of freedom may be used for interference cancellation than constraining on $\mathbf{b}_{i,i}$. Therefore, with this rationale, we expect better performance under this constraint than the constraint on the shortened channel response $\mathbf{b}_{i,i}$. Therefore, we propose another variation of the maximizing SINR method, the WINC method, which optimizes the SINR with the constraint on the norm of the equalizer coefficients, i.e., $\|\mathbf{w}_i\|^2 = 1$, in this subsection. The optimization problem is stated as

$$\max_{\mathbf{w}_i,\mathbf{b}_{i,i},\Delta_i} \frac{\|\mathbf{P}(\Delta_i)\mathbf{w}_i\|^2}{\|\mathbf{P}(\Delta_i)\mathbf{w}_i - \mathbf{D}_i\mathbf{b}_{i,i}\|^2} \quad s.t. \quad \|\mathbf{w}_i\|^2 = 1\,. \tag{31}$$

The optimal solution for $\mathbf{b}_{i,i}^{opt}$ can be derived as

$$\mathbf{b}_{i,i}^{opt} = (\mathbf{D}_i^H\mathbf{D}_i)^{-1}\mathbf{D}_i^H\mathbf{P}(\Delta_i)\mathbf{w}_i^{opt}. \tag{32}$$

The optimal $\mathbf{w}_i$ is the eigenvector corresponding to the smallest eigenvalue of the matrix

$$\begin{aligned}\mathbf{J}_i &= (\mathbf{P}(\Delta_i^{opt})^H\mathbf{P}(\Delta_i^{opt}))^{-1}\mathbf{D}_i^H\mathbf{P}(\Delta_i^{opt}) \\ &\times (\mathbf{P}(\Delta_i^{opt})^H\mathbf{P}(\Delta_i^{opt}))^{-1}\mathbf{P}(\Delta_i^{opt})^H\mathbf{D}_i.\end{aligned} \tag{33}$$

For the case that $\mathbf{T} = \mathbf{I}$, with the statistical assumptions and assuming that the training sequence is sufficiently long, asymptotically the cost function can be simplified as the ratio of

$$\text{the numerator} = \|\boldsymbol{\mathcal{H}}^T\mathbf{w}_i\|^2 + \frac{\sigma_n^2}{\sigma_v^2}\|\mathbf{w}_i\|^2 \tag{34}$$

and

$$\text{the denominator} = \|\boldsymbol{\mathcal{H}}^T\mathbf{w}_i - \bar{\mathbf{C}}_i\mathbf{b}_{i,i}\|^2 + \frac{\sigma_n^2}{\sigma_v^2}\|\mathbf{w}_i\|^2. \tag{35}$$

Applying the technique of separation of variables, we can write the optimal $\mathbf{b}_{i,i}$ in terms of the $\mathbf{w}_i^{opt}$ as

$$\begin{aligned}\mathbf{b}_{i,i}^{opt} &= (\bar{\mathbf{C}}_i^T\bar{\mathbf{C}}_i)^{-1}\bar{\mathbf{C}}_i^T\boldsymbol{\mathcal{H}}^T\mathbf{w}_i^{opt} \\ &= \bar{\mathbf{C}}_i^T\boldsymbol{\mathcal{H}}^T\mathbf{w}_i^{opt}.\end{aligned} \tag{36}$$

When the SNR goes to infinity, the problem can be stated as

$$\max_{\mathbf{w}_i,\bar{\mathbf{C}}_i} \frac{\mathbf{w}_i^H\boldsymbol{\mathcal{H}}^*\boldsymbol{\mathcal{H}}^T\mathbf{w}_i}{\mathbf{w}_i^H\boldsymbol{\mathcal{H}}^*(\mathbf{I}-\bar{\mathbf{C}}_i\bar{\mathbf{C}}_i^T)\boldsymbol{\mathcal{H}}^T\mathbf{w}_i} \quad s.t. \quad \|\mathbf{w}_i\|^2 = 1, \tag{37}$$

which is equivalent to the following optimization problem

$$\max_{\mathbf{w}_i,\bar{\mathbf{C}}_i} \frac{\mathbf{w}_i^H\boldsymbol{\mathcal{H}}^*\bar{\mathbf{C}}_i\bar{\mathbf{C}}_i^T\boldsymbol{\mathcal{H}}^T\mathbf{w}_i}{\mathbf{w}_i^H\boldsymbol{\mathcal{H}}^*(\mathbf{I}-\bar{\mathbf{C}}_i\bar{\mathbf{C}}_i^T)\boldsymbol{\mathcal{H}}^T\mathbf{w}_i} \quad s.t. \quad \|\mathbf{w}_i\|^2 = 1. \tag{38}$$

This is exactly the shortening SINR formulation.

When $\mathbf{T} \neq \mathbf{I}$, the idea still follows, and the frequency tones corresponding to the 1's of the diagonal terms of the selection matrix $\mathbf{T}$ are chosen in the cost function; the analysis is similar to Proposition 5. Asymptotically, the solution using the training-based algorithm is similar to the solution of the shortening SINR formulation with perfect channel information under tone selection.

F. Algorithms with Noise Variance Information

If the noise variance is known to the receiver [55] [56], it is better to consider directly a cost function derived from the signal to interference ratio (SIR) instead of signal to interference and noise ratio (SINR) in the situations where interference dominates noise. For example, for a low complexity transceiver design with convolutional encoding and Viterbi decoding, the method is more resilient to noise than interference, thus the system performance may be interference limited if interference dominates noise. For this purpose, we can modify the SINR cost function by removing the contribution of the noise variance to get better interference suppression.

We modify the optimization problem under the INC constraint in (26) as

$$\max_{\mathbf{w}_i,\mathbf{b}_{i,i},\Delta_i} \frac{\|\mathbf{D}_i\mathbf{b}_{i,i}\|^2}{\|\mathbf{P}(\Delta_i)\mathbf{w}_i - \mathbf{D}_i\mathbf{b}_{i,i}\|^2 - \sigma_n^2K\|\hat{\mathbf{T}}\mathbf{Q}_w\mathbf{w}_i\|^2} \quad \text{s.t.} \quad \|\mathbf{b}_{i,i}\|^2 = 1\,. \tag{39}$$

For simplicity, we call this method M-INC. The subtraction of the noise variance from the cost function gives a zero-forcing type equalization design, which maximizes the signal

$$\tilde{\mathbf{Q}}(\Delta_{i,j}) = \left[\begin{array}{ccc|ccc} 1 & .. & 1 & 1 & \dots & 1 \\ e^{-j2\pi(L+\nu-\Delta_{i,j})/N} & \dots & 1 & e^{j2\pi/N} & \dots & e^{j2\pi\Delta_{i,j}/N} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ e^{-j2\pi(L+\nu-\Delta_{i,j})(N-1)/N} & \dots & 1 & e^{j2\pi(N-1)/N} & \dots & e^{j2\pi\Delta_{i,j}(N-1)/N} \end{array}\right] \tag{22}$$

to interference ratio (SIR), and is similar to the MaxPwr method with perfect channel information [14] since the noise covariance is subtracted from the cost function.

Similarly, we have the modified MMSE method (M-MMSE) as

$$\max_{\mathbf{w}_i,\mathbf{b}_{i,i},\Delta_i} \|\mathbf{P}(\Delta_i)\mathbf{w}_i - \mathbf{D}_i\mathbf{b}_{i,i}\|^2 - \sigma_n^2 K\|\hat{\mathbf{T}}\mathbf{Q}_w\mathbf{w}_i\|^2 \quad \text{s.t.} \quad \|\mathbf{b}_{i,i}\|^2 = 1\,. \tag{40}$$

We can also modify the cost function for the case of constraining on $\mathbf{w}_i$ in (31) (M-WINC) as

$$\max_{\mathbf{w}_i,\mathbf{b}_{i,i},\Delta_i} \frac{\|\mathbf{P}(\Delta_i)\mathbf{w}_i\|^2 - \sigma_n^2 K\|\hat{\mathbf{T}}\mathbf{Q}_w\mathbf{w}_i\|^2}{\|\mathbf{P}(\Delta_i)\mathbf{w}_i - \mathbf{D}_i\mathbf{b}_{i,i}\|^2 - \sigma_n^2 K\|\hat{\mathbf{T}}\mathbf{Q}_w\mathbf{w}_i\|^2} \quad s.t. \quad \|\mathbf{w}_i\|^2 = 1\,. \tag{41}$$

All these modified problems can be solved by applying the technique of separation of variables and finding the maximal eigenvalue of the desired matrices, which is similar to the solutions to the previous problems. We do not elaborate on the solutions here due to lack of space.

G. Detection Scheme/SINR Definition

The detection scheme in the frequency domain considered for all these training based methods is the same as the detection scheme of the method with channel estimates, i.e., per frequency tone based scalar detection as discussed in Section II-D. For coded systems, we also apply the Viterbi decoding as addressed in Section II-D to decode the data. We define the SINR for stream i as

$$\begin{aligned} SINR_i &= 10\log_{10}(E\|\mathbf{T}\mathbf{V}_i(b)\bar{\mathbf{Q}}\mathbf{b}_{i,i}^{opt}\|^2) \\ &- 10\log_{10}\{E\|\mathbf{T}[\sum_{j=1}^{M_r}\mathbf{Q}\mathbf{R}_j(b,\Delta_i^{opt})\mathbf{w}_{i,j}^{opt} \\ &- \mathbf{V}_i(b)\bar{\mathbf{Q}}\mathbf{b}_{i,i}^{opt}]\|^2\}\,. \end{aligned} \tag{42}$$

The overall SINR is defined as

$$SINR = \frac{1}{M_t}\sum_{i=1}^{M_t} SINR_i \tag{43}$$

which is the average SINR for all the active transmit antennas. This is an additional performance metric that we will evaluate in the simulation section.

V. Simulation Results

A. Simulation Setup

We use QPSK or 16-QAM modulation and $N = 128$ tones. The cyclic prefix length is $L_{CP} = 16$. We use a block fading model in the simulation, namely, the channel is assumed to be fixed over a frame which contains ten OFDM symbols, but for each frame, the channel has an independent realization. All simulation results (BER or SINR), as shown in Fig. 4–9, are averaged over 500 independent channel realizations. The channel coefficients are all independently generated according to a complex Gaussian distribution with zero mean and unit variance. For each channel realization, we need one additional OFDM symbol for the equalizer training ($K = 1$), and we use QPSK modulation for training. In the training stage, the equalizer coefficients and the post-equalization channel response are obtained via the joint optimization. After the training stage, we apply the equalizer obtained at the training stage to equalize the received signal. Convolutional coding with generator polynomial (133,171) and the IEEE 802.11a interleaving scheme are used in our simulations to illustrate the coded BER performance. We use the spatial multiplexing scheme, so there is no interleaving in space. At the receiver, we apply the equalization schemes discussed in the previous sections to obtain the equalizer coefficients and the post equalization channel response. We use soft-decision Viterbi decoding for convolutional decoding, and calculate the decoding metrics based on the Euclidean distance. In the simulation study, we choose the selection matrix $\mathbf{T} = \mathbf{I}$. Throughout the simulations, we assume that there is a common discrete time reference in the system, namely time 0. The propagation delays for the two users considered in the rest of the paper are the absolute time lags with respect to the common time reference.

B. Comparison for the Synchronous Case

Fig. 4 illustrates the comparison of the proposed schemes in the scenario of two active transmit antennas and two receive antennas with QPSK modulation. When the data streams are synchronous we can see that the MaxPwr method based on the channel estimates provides the best BER performance, but the M-WINC, M-MMSE and M-INC method give very similar BER performance to the MaxPwr methods. At 16 dB input SNR, all those four methods achieve the BER performance of less than 10^{-4}. The INC, MMSE and WINC methods perform 1.5 dB worse than the modified methods with noise information in intermediate input SNR regions.

In terms of SINR measure (Fig. 5), the methods without noise subtraction from the cost function generally outperform the modified methods with noise subtraction, because the methods without noise subtraction directly maximize the SINR, whereas the methods with noise subtraction maximize the SIR.

C. Comparison for the Asynchronous Case

To study the asynchronous system performance, we conduct simulations for the case of two active transmit antennas, each of which belongs to a different user. The propagation delays for the two users are 0 and 10 samples respectively, with respect to the absolute reference time in the system. The receiver employs two receive antennas. We assume that the two users are assigned distinct training sequences, but these sequences are known to the receiver. This requires certain coordination to assign training sequences in the system, and it is analogous to assigning distinct spreading sequences to users in CDMA systems. This can be achieved via the operations of higher level protocols in real systems, which will be addressed in our future work. We use 16-QAM modulation in our simulations. From Fig. 6, we see that for the asynchronous case (delays = 0 and 10), the BER of the M-WINC, M-INC and M-MMSE slightly outperform the BER of the MaxPwr

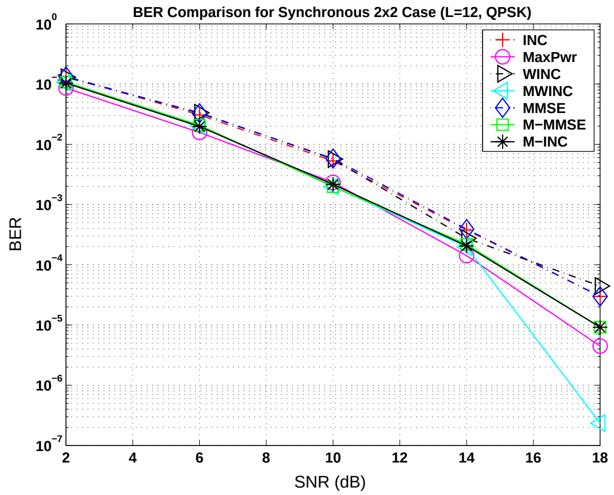


Fig. 4. The plot of BER vs. SNR for the three different schemes with fixed equalizer memory length $L = 12$ and QPSK modulation. The number of channel taps from the first user to the receive antennas is 6, and the number of taps from the second user to the receiver is 8. The propagation delays are both 0.

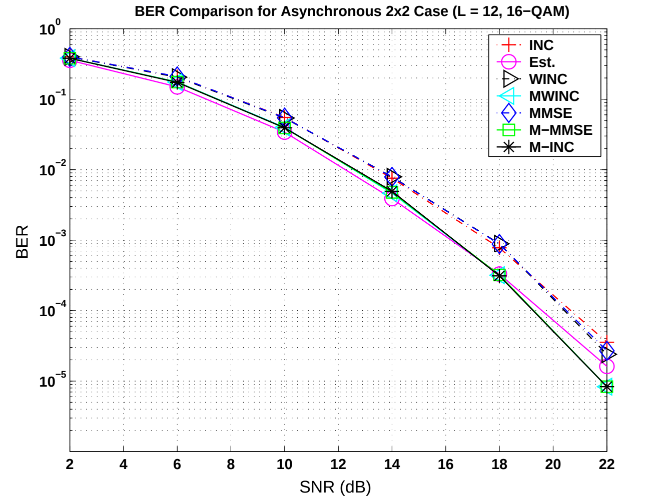


Fig. 6. The plot of BER vs. SNR with fixed equalizer memory length $L = 12$ and 16-QAM modulation. The number of channel taps from the first user to the receive antennas is 6, and the number of taps from the second user to the receiver is 8. The delays are 0 and 10 for the first and second users respectively.

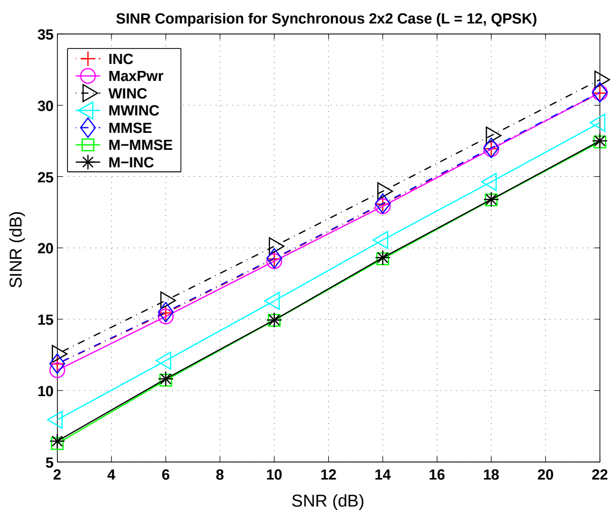


Fig. 5. The plot of SINR vs. SNR for the three different schemes with fixed equalizer memory length $L = 12$ and QPSK modulation. The number of channel taps from the first user to the receive antennas is 6, and the number of taps from the second user to the receiver is 8. The propagation delays are both 0.

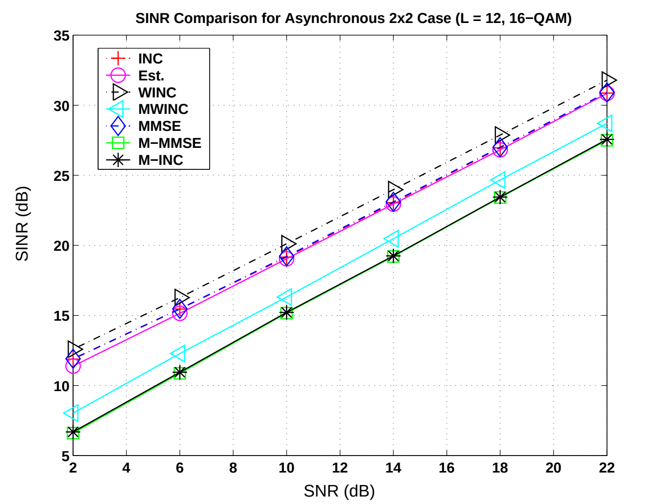


Fig. 7. The plot of SINR vs. SNR with fixed equalizer memory length $L = 12$ and 16-QAM modulation. The number of channel taps from the first user to the receive antennas is 6, and the number of taps from the second user to the receiver is 8. The delays are 0 and 10 for the first and second users respectively.

method with channel estimates in the high SNR region. At 18 dB input SNR, the four schemes all have BER's on the order of 10^{-3} with 16-QAM modulation. The INC, MMSE and WINC methods also perform well in this scenario with only 1.2 dB gap from the BER curves of the modified methods. Similarly, we illustrate the SINR performance in Fig. 7.

D. Comparison in the Presence of Co-channel Interference

To study the effects of co-channel interference on the system performance, we create an effective 4×5 MIMO system. Two of the four active antennas are sending training sequences to the receiver, and the other two active antennas, whose transmit powers are 5 dB less than the other two, are causing co-channel interference. *We assume that the training sequences for the other users are unknown to the receiver*. With this simulation setup, we illustrate the system performance in Fig. 8. It is clear that the MaxPwr method can not handle this scenario since the channels of all the active antennas need to be known at the receiver to conduct the channel estimation. But the other proposed methods give good BER performance. Constraining on the equalizer coefficients gives slightly better BER performance than constraining on the post-equalization channel response. With 16-QAM modulation, at 18 dB input SNR, the modified methods give BER performance of less than 10^{-2}, and the methods without noise subtraction, however, provide BER of less than 10^{-3}.

Notice that in this simulation, we used five receive antennas instead of four to provide better interference suppression. In

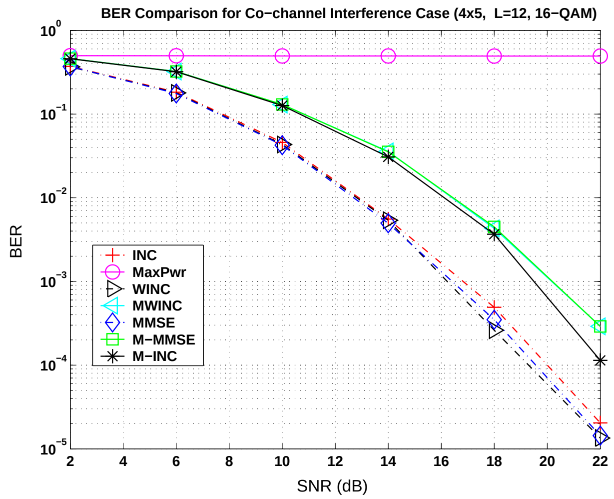


Fig. 8. The plot of BER vs. SNR with fixed equalizer memory length $L = 12$ and 16-QAM modulation. The number of channel taps from the first user to the receive antennas is 6, and the number of taps from the second user to the receiver is 8. The delays are 0 and 10 for the first and second users respectively. The co-channel interference powers are 5 dB less than those for the transmitted data streams.

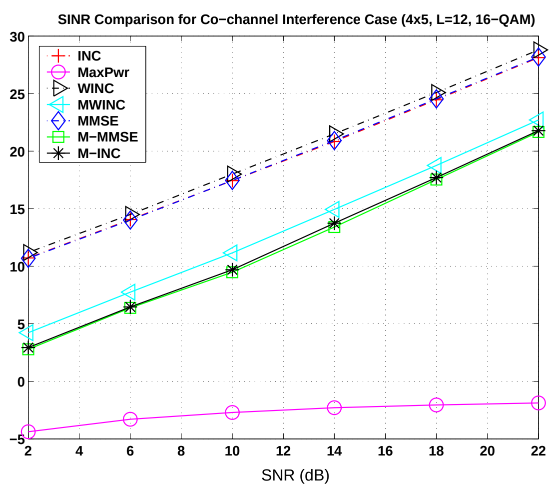


Fig. 9. The plot of BER vs. SNR with fixed equalizer memory length $L = 12$ and 16-QAM modulation. The number of channel taps from the first user to the receive antennas is 6, and the number of taps from the second user to the receiver is 8. The delays are 0 and 10 for the first and second users respectively. The co-channel interference powers are 5 dB less than those for the transmitted data streams.

this scenario, the receiver has more degrees of freedom to cancel interference and thus the BER performance is more noise-dominated rather than interference-dominated. Consequently, the methods based on the rationale of maximizing SINR actually perform better than the methods based on the rationale of maximizing SIR. We also illustrate the SINR versus SNR plot in Fig. 9. The SINR's of the methods with noise subtraction are generally less than those for methods without noise subtraction as expected since maximizing SIR does not necessarily provides the best SINR performance.

VI. Conclusions

In this paper we proposed a general two-stage framework for interference cancellation in MIMO-OFDM multiuser systems. The first stage utilizes time domain equalization to suppress cochannel interference, mitigate asynchronism and shorten the channel to suppress the inter-carrier interference. The second stage consists of a frequency domain per tone scalar equalization and detection to compensate for the channel distortion and detect the signals. We formulated and solved a training-based joint optimization problem of maximizing the signal to noise and interference ratio to find the equalizer coefficients and the post-equalization channel impulse responses. We also proposed several variations to the maximizing SINR formulation, which either reduce computation or improve the performance. Simulations show that our training based methods provide very good BER and SINR performance for both synchronous and asynchronous cases and are robust to co-channel interference.

Appendix I
Derivation of the Space-Frequency Model

Following the space-time data model in (4), let us define a vector $\hat{\mathbf{s}}_m(b, \Delta_{i,j})$ of size $(N+L+\nu)\times 1$ in (A.I.1), shown at the bottom of the next page. The vector $\mathbf{s}_m(p:q, b)$ denotes the p^{th} to the q^{th} samples of the b^{th} OFDM symbol of the m^{th} data stream. Note that $0 \le \Delta_{i,j} \le L+\nu-L_{CP}$ and $L_{CP} < \nu$. When $\Delta_{i,j} = L+\nu-L_{CP}$, there are no samples from the $(b-1)^{th}$ OFDM symbol in the $\hat{\mathbf{s}}_m(b)$. Similarly, when $\Delta_{i,j} = 0$, there are no samples from the $(b+1)^{th}$ OFDM symbol in the $\hat{\mathbf{s}}_m(b)$.

We stack the transmitted signal for the m^{th} data stream in a $N \times (L+\nu+1)$ matrix $\mathbf{S}_m(b, \Delta_{i,j})$, which is generated from the data vector $\hat{\mathbf{s}}_m(b, \Delta_{i,j})$ by sequentially taking $L+\nu+1$ elements of the $\hat{\mathbf{s}}_m(b, \Delta_{i,j})$ as each row of the matrix $\mathbf{S}_m(b, \Delta_{i,j})$. Thus the matrix $\mathbf{S}_m(b, \Delta_{i,j})$ is given in (A.I.2) at the bottom of the next page.
Similarly, we have the stacked noise matrix $\mathbf{N}_j(\Delta_{i,j})$ of size $N \times (L+1)$ generated from the noise vector at the j^{th} receive antenna with the delay parameter $\Delta_{i,j}$, which has size $(N+L)\times 1$.

We define a data matrix $\bar{\mathbf{S}}_m(b)$ of size $N \times (L_{CP}+1)$, which consists of the $(L+\nu-\Delta_{i,j}-L_{CP}+1)^{th}$ to the $(L+\nu-\Delta_{i,j}+1)^{th}$ columns of the data matrix $\mathbf{S}_m(b, \Delta_{i,j})$. Notice that $\bar{\mathbf{S}}_m(b)$ only contains signal samples from the b^{th} OFDM symbol. We also define matrices $\tilde{\mathbf{S}}_{m,1}(b, \Delta_{i,j})$ and $\tilde{\mathbf{S}}_{m,2}(b, \Delta_{i,j})$, such that

$$\mathbf{S}_m(b, \Delta_{i,j}) = [\tilde{\mathbf{S}}_{m,1}(b, \Delta_{i,j}), \bar{\mathbf{S}}_m(b), \tilde{\mathbf{S}}_{m,2}(b, \Delta_{i,j})]. \quad \text{(A.I.3)}$$

To simplify the notation, we define a matrix $\tilde{\mathbf{S}}_m(b, \Delta_{i,j})$ as

$$\tilde{\mathbf{S}}_m(b, \Delta_{i,j}) = [\tilde{\mathbf{S}}_{m,1}(b, \Delta_{i,j}), \mathbf{0}_{N\times(L_{CP}+1)}, \tilde{\mathbf{S}}_{m,2}(b, \Delta_{i,j})]. \quad \text{(A.I.4)}$$

Similarly, we define a channel matrix $\bar{\mathbf{H}}_{j,m}(\Delta_{i,j})$ of size $(L+1)\times(L_{CP}+1)$, which consists of the $(L+\nu-\Delta_{i,j}-L_{CP}+1)^{th}$ to the $(L+\nu-\Delta_{i,j}+1)^{th}$ columns of the channel matrix $\mathbf{H}_{j,m}(\Delta_{i,j})$. We also define two other channel matrices $\tilde{\mathbf{H}}_{j,m,1}(\Delta_{i,j})$ and $\tilde{\mathbf{H}}_{j,m,2}(\Delta_{i,j})$, such that

$$\mathbf{H}_{j,m} = [\tilde{\mathbf{H}}_{j,m,1}(\Delta_{i,j}), \bar{\mathbf{H}}_{j,m}(\Delta_{i,j}), \tilde{\mathbf{H}}_{j,m,2}(\Delta_{i,j})]. \quad \text{(A.I.5)}$$

To simplify the notation, we define a matrix $\tilde{\mathbf{H}}_{j,m}(\Delta_{i,j})$ as

$$\tilde{\mathbf{H}}_{j,m}(\Delta_{i,j}) = [\tilde{\mathbf{H}}_{j,m,1}(\Delta_{i,j}), \mathbf{0}_{(L+1)\times(L_{CP}+1)}, \tilde{\mathbf{H}}_{j,m,2}(\Delta_{i,j})]. \quad \text{(A.I.6)}$$

After time domain equalization and discarding the cyclic prefix, we stack the i^{th} output stream of the space-time equalizer bank in a vector $\mathbf{x}_i(b)$,

$$\mathbf{x}_i(b) = \begin{bmatrix} x_i(0,b) \\ \vdots \\ x_i(N-1,b) \end{bmatrix} \quad \text{(A.I.7)}$$

where the relation between $x_i(p,b)$ and $x_i(k)$ is defined in (3), and $k = b(N+L_{CP}) + p + L_{CP}$. The time domain received signal matrix $\mathbf{R}_j(b,\Delta_{i,j})$ at the j^{th} receive antenna is defined as

$$\mathbf{R}_j(b,\Delta_{i,j}) = \sum_{m=1}^{M_t} \mathbf{S}_m(b,\Delta_{i,j})\mathbf{H}_{j,m}^T + \mathbf{N}_j(\Delta_{i,j}). \quad \text{(A.I.8)}$$

Thus, we have

$$\begin{aligned} \mathbf{x}_i(b) &= \sum_{j=1}^{M_r} \mathbf{R}_j(b,\Delta_{i,j})\mathbf{w}_{i,j} && \text{(A.I.9)} \\ &= \sum_{j=1}^{M_r}\sum_{m=1}^{M_t} \mathbf{S}_m(b,\Delta_{i,j})\mathbf{H}_{j,m}^T\mathbf{w}_{i,j} \\ &+ \sum_{j=1}^{M_r} \mathbf{N}_j(\Delta_{i,j})\mathbf{w}_{i,j}. && \text{(A.I.10)} \end{aligned}$$

Taking the DFT of the stacked data vector $\mathbf{x}_i(b)$, we have the following equations

$$\begin{aligned} \mathbf{y}_i(b) &= \mathbf{Q}\mathbf{x}_i(b) \\ &= \sum_{j=1}^{M_r} \mathbf{Q}\mathbf{R}_j(b,\Delta_{i,j})\mathbf{w}_{i,j} \\ &= \sum_{j=1}^{M_r} \mathbf{Q}\bar{\mathbf{S}}_i(b)\bar{\mathbf{H}}_{j,i}(\Delta_{i,j})^T\mathbf{w}_{i,j} \\ &+ \sum_{j=1}^{M_r} \mathbf{Q}\tilde{\mathbf{S}}_i(b,\Delta_{i,j})\tilde{\mathbf{H}}_{j,i}(\Delta_{i,j})^T\mathbf{w}_{i,j} \\ &+ \sum_{j=1}^{M_r}\sum_{m=1,m\neq i}^{M_t} \mathbf{Q}\mathbf{S}_m(b,\Delta_{i,j})\mathbf{H}_{j,m}^T\mathbf{w}_{i,j} \\ &+ \sum_{j=1}^{M_r} \mathbf{Q}\mathbf{N}_j(\Delta_{i,j})\mathbf{w}_{i,j}. && \text{(A.I.11)} \end{aligned}$$

Notice that $\mathbf{Q}\bar{\mathbf{S}}_i(b) = \mathbf{V}_i(b)\bar{\mathbf{Q}}$, where, $\mathbf{V}_i(b) = \text{diag}\{v_i(0,b),...,v_i(N-1,b)\}$, and the matrix $\bar{\mathbf{Q}}$ which has size $N\times(L_{CP}+1)$ is defined as

$$\bar{\mathbf{Q}} = \begin{bmatrix} 1 & ... & 1 \\ e^{-j2\pi L_{CP}/N} & ... & 1 \\ \vdots & \ddots & \vdots \\ e^{-j2\pi L_{CP}(N-1)/N} & ... & 1 \end{bmatrix}. \quad \text{(A.I.12)}$$

Define

$$\mathbf{b}_{i,i} = \sum_{j=1}^{M_r} \bar{\mathbf{H}}_{j,i}(\Delta_{i,j})^T\mathbf{w}_{i,j}, \quad \text{(A.I.13)}$$

which has size $(L_{CP}+1)\times 1$, and define the noise and

---

$$\begin{aligned} \hat{\mathbf{s}}_m(b,\Delta_{i,j}) = [&\mathbf{s}_m(N+L_{CP}+\Delta_{i,j}-L-\nu : N-1, b-1)^T, \mathbf{s}_m(N-L_{CP}:N-1,b)^T, \mathbf{s}_m(0:N-1,b)^T, \\ &\mathbf{s}_m(N-L_{CP}:N-L_{CP}+\Delta_{i,j}-1,b+1)^T]^T. \end{aligned} \quad \text{(A.I.1)}$$

$$\left[\begin{array}{c|c|c} \mathbf{s}_m(N+L_{CP}+\Delta_{i,j}-L-\nu : N-1,b-1)^T & [\mathbf{s}_m(N-L_{CP}:N-1,b)^T, s_m(0,b)] & \mathbf{s}_m(1:\Delta_{i,j},b)^T \\ [\mathbf{s}_m(N+L_{CP}+\Delta_{i,j}-L-\nu+1 : N-1,b-1)^T, s_m(N-L_{CP},b)] & [\mathbf{s}_m(N-L_{CP}+1:N-1,b)^T, \mathbf{s}_m(0:1,b)^T] & \mathbf{s}_m(2:\Delta_{i,j}+1,b)^T \\ \vdots & \ddots & \vdots \\ \mathbf{s}_m(N+\Delta_{i,j}-L-\nu-1 : N-L_{CP}-2,b)^T & \mathbf{s}_m(N-L_{CP}-1:N-1,b)^T & \mathbf{s}_m(N-L_{CP}:N-L_{CP}+\Delta_{i,j}-1,b+1)^T \end{array}\right] \quad \text{(A.I.2)}$$

interference term $\mathbf{e}_i(b)$ as follows

$$\begin{aligned}\mathbf{e}_i(b) &= \sum_{j=1}^{M_r} \mathbf{Q}\tilde{\mathbf{S}}_i(b,\Delta_{i,j})\tilde{\mathbf{H}}_{j,i}(\Delta_{i,j})^T\mathbf{w}_{i,j} \\ &+ \sum_{j=1}^{M_r}\sum_{m=1,m\neq i}^{M_t} \mathbf{Q}\mathbf{S}_m(b,\Delta_{i,j})\mathbf{H}_{j,m}^T\mathbf{w}_{i,j} \\ &+ \sum_{j=1}^{M_r}\mathbf{Q}\mathbf{N}_j(\Delta_{i,j})\mathbf{w}_{i,j}. \qquad \text{(A.I.14)}\end{aligned}$$

Appendix II
Sketch of the Proof of Proposition 4 and 5

a) Proof of Proposition 4: The proof follows by expanding equation (15) and calculating the auto-correlation and cross-correlation for $\mathbf{S}_m(b,\Delta_i)$ and $\bar{\mathbf{S}}_m(b)$ (both are defined in Appendix I), because $\mathbf{V}_m(b)\bar{\mathbf{Q}} = \mathbf{Q}\bar{\mathbf{S}}_m(b)$.

b) Proof of Proposition 5: Define two matrices $\mathbf{A}_m(b)$ (A.II.15) and $\bar{\mathbf{A}}_m(b)$ (A.II.16) at the bottom of the next page.

Notice the fact that with Assumption 2, we have

$$\begin{aligned}&E(\mathbf{S}_m(b,\Delta_i)^H\mathbf{Q}^H\mathbf{T}\mathbf{V}_m(b)\bar{\mathbf{Q}}) = E(\tilde{\mathbf{Q}}(\Delta_i)^H\mathbf{V}_m(b)^H\mathbf{T} \\ &\times\mathbf{V}_m(b)\bar{\mathbf{Q}}) - E(\mathbf{A}_m(b,\Delta_i)^H\mathbf{Q}^H\mathbf{T}\mathbf{Q}\bar{\mathbf{S}}_m(b))\end{aligned} \qquad \text{(A.II.17)}$$

and

$$\begin{aligned}&E(\mathbf{S}_m(b,\Delta_i)^H\mathbf{Q}^H\mathbf{T}\mathbf{Q}\mathbf{S}_m(b,\Delta_i)) = E(\tilde{\mathbf{Q}}(\Delta_i)^H\mathbf{V}_m(b)^H \\ &\times\mathbf{T}\mathbf{V}_m(b)\tilde{\mathbf{Q}}(\Delta_i)) - E(\mathbf{A}_m(b,\Delta_i)^H\mathbf{Q}^H\mathbf{T}\mathbf{Q}\bar{\mathbf{A}}_m(b,\Delta_i)) \\ &-E(\bar{\mathbf{A}}_m(b,\Delta_i)^H\mathbf{Q}^H\mathbf{T}\mathbf{Q}\mathbf{A}_m(b,\Delta_i)).\end{aligned} \qquad \text{(A.II.18)}$$

Define $\delta(\boldsymbol{\mathcal{H}}^T\mathbf{w}_i,\mathbf{b}_{i,i})$ in (A.II.19) at the bottom of the next page. Notice that the term $\delta(\boldsymbol{\mathcal{H}}^T\mathbf{w}_i,\mathbf{b}_{i,i})$ vanishes when $\mathbf{T}=\mathbf{I}$. The rest of the proof follows straightforward calculations.

References

[1] H. Bolcskei, D. Gesbert, and A. J. Paulraj, "On the capacity of OFDM-based spatial multiplexing systems," *IEEE Trans. Commun.*, vol. 50, pp. 225–234, Feb. 2002.

[2] D. Gesbert, M. Shafi, D.-S. Shiu, P. J. Smith, and A. Naguib, "From theory to practice: An overview of MIMO space-time coded wireless systems," *IEEE Jour. Select. Areas in Commun.*, vol. 21, no. 3, pp. 281–302, April 2003.

[3] A. Goldsmith, S. A. Jafar, N. Jindal, and S. Vishwanath, "Capacity limits of MIMO channels," *IEEE Jour. Select. Areas in Commun.*, vol. 21, no. 5, pp. 684–702, June 2003.

[4] G. J. Foschini and M. J. Gans, "On limits of wireless communications in a fading environment when using multiple antennas," *Wireless Personal Communications*, vol. 6, no. 3, pp. 311–335, March 1998.

[5] L. Zheng and D. N. C. Tse, "Diversity and multiplexing: A fundamental tradeoff in multiple-antenna channels," *IEEE Trans. Inf. Th.*, vol. 49, no. 5, pp. 1073–1096, May 2003.

[6] G. L. Stuber, J. R. Barry, S. W. Mclaughlin, Y. G. Li, M. A. Ingram, and T. G. Pratt, "Broadband MIMO-OFDM wireless communications," in *Proceedings of the IEEE*, Feb. 2004, vol. 92, pp. 271 – 294.

[7] "IEEE 802.11n," http://grouper.ieee.org/groups/802/11.

[8] A. Van Zelst and T. C. W. Schenk, "Implementation of a MIMO OFDM-based wireless LAN system," *IEEE Trans. on Sig. Proc.*, vol. 52, no. 2, pp. 483–494, Feb. 2004.

[9] 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 Comm. Mag.*, vol. 40, pp. 143 – 149, Sep. 2002.

[10] R. K. Martin, J. Balakrishnan, W. A. Sethares, and C. R. Johnson, Jr., "A blind adaptive TEQ for multicarrier systems," *IEEE Sig. Proc. Lett.*, vol. 9, no. 11, pp. 341–343, Nov. 2002.

[11] J. S. Chow, J. M. Cioffi, and J. A. C. Bingham, "Equalizer training algorithms for multicarrier modulation systems," in *Proc. Int. Conf. on Commun.*, May 1993, pp. 761–765.

[12] T. A. Thomas and F. W. Vook, "Asynchronous interference suppression in broadband cyclic-prefix communications," in *IEEE Wireless Communications and Networking Conference*, March 2003, pp. 568–572.

[13] M. B. Breinholt, M. D. Zoltowski, and T. A. Thomas, "Space-time processing for maximum frequency reuse in OFDM cellular communications," in *Proc. of Allerton Conf. on Comm. Cont. and Comp.*, Oct. 2002, pp. 713–722.

[14] M. B. Breinholt, M. D. Zoltowski, and T. A. Thomas, "Space-time equalization and interference cancellation for MIMO-OFDM," in *Proc. of the Asil. Conf. on Sig. Sys. and Comp.*, Nov. 2002, vol. 2, pp. 1688–1693.

[15] M. B. Breinholt and M. D. Zoltowski, "Space-time equalization and interference cancellation with convolutional coding for MIMO-OFDM," in *Proc. Int. Conf. Acoust., Speech and Sig. Proc.*, Apr. 2003, vol. 4, pp. 413–416.

[16] A. Stamoulis, S. N. Diggavi, and N. Al-Dhahir, "Intercarrier interference in MIMO OFDM," *IEEE Trans. on Sig. Proc.*, vol. 50, no. 10, pp. 2451–2464, Oct. 2002.

[17] S. Kapoor, D. J. Marchok, and Y-F. Huang, "Adaptive interference suppression in multiuser wireless OFDM systems using antenna arrays," *IEEE Trans. on Sig. Proc.*, vol. 47, no. 12, pp. 3381–3391, Dec. 1999.

[18] Y. Li and L. J. Cimini, Jr., "Bounds on the interchannel interference of OFDM in time-varying impairments," *IEEE Trans. Commun.*, vol. 49, no. 3, pp. 401–404, Mar. 2001.

[19] T. Tang and R. W. Heath, Jr., "Joint frequency offset estimation and interference cancellation in MIMO-OFDM systems," in *Proc. of IEEE Veh. Technol. Conf.*, Los Angeles, CA, Sept. 2004.

[20] N. Al-Dhahir and J. Cioffi, "A bandwidth-optimized reduced-complexity equalized multicarrier transceiver," *IEEE Trans. Commun.*, vol. 45, no. 8, pp. 948–956, 1997.

[21] T. Starr, J. M. Cioffi, and P. J. Silvermann, *Understanding Digital Subscriber Line Technology*, Prentice Hall PTR, Upper Saddle River, NJ, 1999.

[22] M. Nafie and A. Gatherer, "Time-domain equalizer training for ADSL," in *Proc. Glob. Telecom. Conf.*, June 1997, pp. 1085–1089.

[23] M. V. Bladel and M. Moeneclaey, "Time-domain equalization for multicarrier communication," in *Proc. Glob. Telecom. Conf.*, Nov. 1995, pp. 167–171.

[24] D. Pal, G. N. Iyengar, and J. M. Cioffi, "A new method of channel shortening with applications to discrete multi-tone (DMT) systems," in *Proc. Int. Conf. on Commun.*, June 1998, vol. 2, pp. 763–768.

[25] Yi Sun and Lang Tong, "Channel equalization using one-tap DFE for wireless OFDM systems with ICI and ISI," in *Signal Processing Advances in Wireless Communications, 1999. SPAWC '99. 1999 2nd IEEE Workshop on*, May 1999, pp. 146 – 149.

[26] Xiaodong Cai and G. B. Giannakis, "Low-complexity ICI suppression for OFDM over time- and frequency-selective rayleigh fading channels," in *Signals, Systems and Computers, 2002. Conference Record of the Thirty-Sixth Asilomar Conference on*, Nov. 2002, vol. 2, pp. 1822–1826.

[27] W. Younis and N. Al-Dhahir, "Joint prefiltering and MLSE equalization of space-time coded transmissions over frequency-selective channels," *IEEE Trans. on Veh. Technol.*, vol. 51, no. 1, pp. 144–154, 2002.

[28] J. Kim, R. W. Heath Jr., and E. J. Powers, "Receiver designs for alamouti coded OFDM systems in fast fading channels," to appear in IEEE Trans. on Wireless in 2004.

[29] K. V. Acker, G. Leus, M. Moonen, and T. Pollet, "RLS-based initialization for per-tone equalizers in DMT receivers," *IEEE Trans. Commun.*, vol. 51, no. 6, pp. 885–889, June 2003.

[30] G. Leus and M. Moonen, "Per-tone equalization for MIMO-OFDM systems," *IEEE Trans. on Sig. Proc.*, vol. 51, no. 11, pp. 2965–2975, November 2003.

[31] P. J. W. Melsa, R. C. Younce, and C. E. Rohrs, "Impulse response shortening for discrete multitone transceivers," *IEEE Trans. Commun.*, vol. 44, pp. 1662–1672, Dec. 1996.

[32] G. Arslan, B. L. Evans, and S. Kiaei, "Equalization for discrete multitone transceivers to maximize bit rate," *IEEE Trans. on Sig. Proc.*, vol. 49, no. 12, pp. 3123–3135, Dec. 2001.

[33] N. Al-Dhahir, "FIR channel-shortening equalizers for MIMO ISI channels," *IEEE Trans. Commun.*, vol. 49, no. 2, pp. 213–218, Feb. 2001.

[34] A. Tkacenko and P. P. Vaidyanathan, "Eigenfilter design of MIMO equalizers for channel shortening," in *Proc. Int. Conf. Acoust., Speech and Sig. Proc.*, May 2002, pp. 2361–2364.

[35] Ye (Geoffrey) Li, Nambirajan Seshadri, and Sirikiat Ariyavisitakul, "Channel estimation for OFDM systems with transmitter diversity in mobile wireless channels," *IEEE Jour. Select. Areas in Commun.*, vol. 17, no. 3, pp. 461–471, Mar. 1999.

[36] Ye (Geoffrey) Li and Nelson R. Sollenberger, "Adpative antenna arrays for OFDM systems with cochannel interference," *IEEE Trans. Commun.*, vol. 47, no. 2, pp. 217–229, Feb. 1999.

[37] J.-W. Liang, J.-T. Chen, and A. J. Paulraj, "A two-stage hybrid approach for CCI/ISI reduction with space-time processing," *IEEE Commun. Lett.*, vol. 1, no. 6, pp. 163–165, Nov. 1997.

[38] R. Samanta, R. W. Heath Jr., and B. L. Evans, "Joint space-time interference cancellation and channel shortening," in *Proc. of the Asil. Conf. on Sig. Sys. and Comp.*, Nov. 2003.

[39] A. Paulraj, R. Nabar, and D. Gore, *Introduction to Space-Time Wireless Communications*, Cambridge University Press, 2003.

[40] J. Shen and Z. Ding, "Direct blind MMSE channel equalization based on second-order statistics," *IEEE Trans. on Sig. Proc.*, vol. 48, no. 4, pp. 1015–1022, April 2000.

[41] A. Leke and J. M. Cioffi, "A maximum rate loading algorithm for discrete multitone modulation systems," *Proc. Glob. Telecom. Conf.*, vol. 3, pp. 1514–1518, Nov. 1997.

[42] L. Goldfeld, V. Lyandres, and D. Wulich, "Minimum ber power loading for ofdm in fading channel," *IEEE Trans. Commun.*, vol. 50, pp. 1729 – 1733, Nov. 2002.

[43] V. Tarokh, N. Seshadri, and A. R. Calderbank, "Space-time codes for high data rate wireless communication: Performance criterion and code construction," *IEEE Trans. Inf. Th.*, vol. 44, no. 2, pp. 744–765, March 1998.

[44] Z. Liu and G. B. Giannakis, "Space-time block coded multiple access through frequency-selective fading channels," *IEEE Trans. Commun.*, vol. 49, pp. 1033–1044, June 2001.

[45] S. Zhou and G. B. Giannakis, "Space-time coding with maximum diversity gains over frequency-selective fading channels," in *IEEE Sig. Proc. Lett.*, Oct. 2001, vol. 8, pp. 269–272.

[46] H. H. Zeng, Y. Li, and J. H. Winters, "Improved spatial-temporal equalization for EDGE: A fast selective-direction MMSE timing recovery algorithm and two-stage soft-output equalizer," *IEEE Trans. Commun.*, vol. 49, no. 12, pp. 2124–2134, Dec. 2001.

[47] S. L. Ariyavisitakul, J. H. Winters, and N. R. Sollenberger, "Joint equalization and interference suppression for high data rate wireless systems," *IEEE Jour. Select. Areas in Commun.*, vol. 18, no. 7, pp. 1214–1220, July 2000.

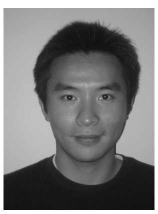
[48] Y. Li, J. H. Winters, and N. R. Sollenberger, "Spatial-temporal equalization for IS-136 TDMA systems with rapid dispersive fading

$$\mathbf{A}_m(b) = \left[\begin{array}{c|c|c} \mathbf{s}_m(N+\Delta_{i,j}-L-\nu : N-L_{CP}-1,b)^T & \mathbf{0}^T_{1\times(L_{CP}+1)} & \mathbf{0}^T_{1\times\Delta_{i,j}} \\ [\mathbf{s}_m(N+\Delta_{i,j}-L-\nu+1 : N-L_{CP}-1,b)^T, 0] & \mathbf{0}^T_{1\times(L_{CP}+1)} & \mathbf{0}^T_{1\times\Delta_{i,j}} \\ \vdots & \ddots & \vdots \\ \mathbf{0}^T_{1\times(L+\nu-\Delta_{i,j}-L_{CP})} & \mathbf{0}^T_{1\times(L_{CP}+1)} & \mathbf{s}_m(N-L_{CP} : N-L_{CP}+\Delta_{i,j}-1,b+1)^T \end{array}\right] \quad \text{(A.II.15)}$$

$$\bar{\mathbf{A}}_m(b) = \left[\begin{array}{c|c|c} \mathbf{0}^T_{1\times(L+\nu-\Delta_{i,j}-L_{CP})} & [\mathbf{s}_m(N+L_{CP}+\Delta_{i,j}-L-\nu+1 : N-1,b-1)^T, s_m(0,b)]^T & \mathbf{s}_m(1:\Delta_{i,j},b)^T \\ [\mathbf{0}^T_{1\times(L+\nu-\Delta_{i,j}-L_{CP}-1)}, s_m(N-L_{CP},b)] & [\mathbf{s}_m(N-L_{CP}+1 : N-1,b)^T, \mathbf{s}_m(0:1,b)^T] & \mathbf{s}_m(2:\Delta_{i,j}+1,b)^T \\ \vdots & \ddots & \vdots \\ \mathbf{s}_m(N+\Delta_{i,j}-L-\nu-1 : N-L_{CP}-2,b)^T & \mathbf{s}_m(N-L_{CP}-1 : N-1,b)^T & \mathbf{0}^T_{1\times\Delta_{i,j}} \end{array}\right] \quad \text{(A.II.16)}$$

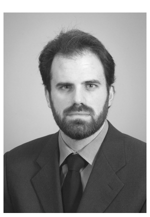
$$\begin{aligned} \delta(\boldsymbol{\mathcal{H}}^T\mathbf{w}_i, \mathbf{b}_{i,i}) &= E\{(\sum_{j=1}^{M_r}\mathbf{w}^H_{i,j}\mathbf{H}^*_{j,i})\mathbf{A}_i(b)^H\mathbf{Q}^H\mathbf{T}\mathbf{Q}(\bar{\mathbf{S}}_i\mathbf{b}_{i,i} - \bar{\mathbf{A}}_i(b)\sum_{j=1}^{M_r}\mathbf{H}^T_{j,i}\mathbf{w}_{i,j}) \\ &+ ((\sum_{j=1}^{M_r}\mathbf{w}^H_{i,j}\mathbf{H}^*_{j,i})\mathbf{A}_i(b)^H\mathbf{Q}^H\mathbf{T}\mathbf{Q}(\bar{\mathbf{S}}_i\mathbf{b}_{i,i} - \bar{\mathbf{A}}_i(b)\sum_{j=1}^{M_r}\mathbf{H}^T_{j,i}\mathbf{w}_{i,j}))^H \\ &- \sum_{m=1,m\neq i}^{M_t}(\sum_{j=1}^{M_r}\mathbf{w}^H_{i,j}\mathbf{H}^*_{j,m})\mathbf{A}_m(b)^H\mathbf{Q}^H\mathbf{T}\mathbf{Q}\bar{\mathbf{A}}_m(b)\sum_{j=1}^{M_r}\mathbf{H}^T_{j,m}\mathbf{w}_{i,j} \\ &- \sum_{m=1,m\neq i}^{M_t}((\sum_{j=1}^{M_r}\mathbf{w}^H_{i,j}\mathbf{H}^*_{j,m})\mathbf{A}_m(b)^H\mathbf{Q}^H\mathbf{T}\mathbf{Q}\bar{\mathbf{A}}_m(b)\sum_{j=1}^{M_r}\mathbf{H}^T_{j,m}\mathbf{w}_{i,j})^H \\ &+ \sum_{j=1}^{M_r}\mathbf{w}^H_{i,j}\mathbf{N}^H_j\mathbf{Q}^H\mathbf{T}\mathbf{Q}\mathbf{N}_j\mathbf{w}_{i,j} - \|\hat{\mathbf{T}}\mathbf{Q}_w\mathbf{w}_i\|^2\} \end{aligned} \quad \text{(A.II.19)}$$

and cochannel interference," *IEEE Trans. on Veh. Technol.*, vol. 48, no. 4, pp. 1182–1194, July 1999.
[49] A. Burr, *Modulation and Coding*, Prentice Hall, 2001.
[50] S. Le Goff, A. Glavieux, and C. Berrou, "Turbo-codes and high spectral efficiency modulation," in *Proc. Int. Conf. on Commun.*, May 1994, vol. 2, pp. 645–649.
[51] K.-B. Song and S. A. Mujtaba, "A low complexity space-frequency BICM MIMO-OFDM system for next generation WLANs," in *Proc. Glob. Telecom. Conf.*, Dec. 2003, vol. 2, pp. 1059–1063.
[52] J. G. Proakis, *Digital Communications*, McGraw-Hill, 2000.
[53] E. G. Larsson and P. Stoica, *Space-Time Block Coding for Wireless Communications*, Cambridge University Press, New York, 2003.
[54] D. D. Falconer and F. R. Magee, "Adaptive channel memory truncation for maximum likelihood sequence estimation," *The Bell Sys. Tech. Jour.*, pp. 1541–1562, 1973.
[55] F.-I. Buchholz, W. Kessel, and F. Melchert, "Noise power measurements and measurement uncertainties," *Instrumentation and Measurement, IEEE Transactions on*, vol. 41, no. 4, pp. 476–481, Aug. 1992.
[56] C.-Y. Lee, W. Panton, G. Granata, and A. Rajkotia, "Measurement of noise figure, G/T, and E_b/N_0 using RSSI," in *Technologies for Wireless Applications, 1999. Digest. 1999 IEEE MTT-S Symposium on*, Feb. 1999, pp. 101–103.



Taiwen Tang (S'04) received the B.E. in electronic engineering from Beijing University of Posts and Telecommunications, Beijing, China in 2001, and the M.S. in electrical and computer engineering from Colorado State University, Fort Collins, CO in 2003. He is currently pursuing the Ph.D. degree in the Department of Electrical and Computer Engineering, The University of Texas at Austin.

From August 2001 to July 2003, he was a member of the Radar and Communications Group in Colorado State University, working on remote sensing with dual polarized meteorological Radar systems. Since August 2003, he has been a research assistant in the Wireless Networking and Communications Group of The University of Texas at Austin. His current research interests are in wireless digital communications and wireless networks with emphasis on MIMO-OFDM communications, scheduling algorithms and ad hoc networks. His industrial experience includes internship in Beceem Communications Inc., Santa Clara, CA, participating in IEEE 802.16 d/e system development and standard contributions, in the summer of 2004.



Robert W. Heath, Jr. (S'96 - M'01) received the B.S. and M.S. degrees from the University of Virginia, Charlottesville, VA, in 1996 and 1997 respectively, and the Ph.D. from Stanford University, Stanford, CA, in 2002, all in electrical engineering.

From 1998 to 1999, he was a Senior Member of the Technical Staff at Iospan Wireless Inc, San Jose, CA where he played a key role in the design and implementation of the physical and link layers of the first commercial MIMO-OFDM communication system. From 1999 to 2001 he served as a Senior Consultant for Iospan Wireless Inc. In 2003 he founded MIMO Wireless Inc, a consulting company dedicated to the advancement of MIMO technology. Since January 2002, he has been with the Department of Electrical and Computer Engineering at The University of Texas at Austin where he serves as an Assistant Professor as part of the Wireless Networking and Communications Group. His research interests include interference management in wireless networks, sequence design, and all aspects of MIMO communication including antenna design, practical receiver architectures, limited feedback techniques, and scheduling algorithms.

Dr. Heath serves as an Associate Editor for the IEEE Transactions on Communications and the IEEE Transactions on Vehicular Technology. He is a member of the SPCOM Technical Committee in the Signal Processing Society.