A New Precoder Design for Blind Channel Estimation in MIMO-OFDM Systems

Song Noh, *Student Member, IEEE*, Youngchul Sung, *Senior Member, IEEE*, and Michael D. Zoltowski, *Fellow, IEEE*

***Abstract*—A new precoder for precoding-based blind channel estimation for MIMO-OFDM systems is proposed. In the proposed scheme, only a small number of data symbols, commensurate with the channel length, are linearly precoded prior to transmission to induce the signal correlation needed for the scheme to blindly estimate the channels. Similar to how pilot symbols are transmitted, the subcarriers carrying these linearly precoded data symbols are equi-spaced across the frequency band. The other subcarriers carry data symbols in the standard way, enabling MLD per subcarrier and also allowing for MIMO linear precoding across antennas at each subcarrier. This is in contrast to previous precoding-based blind channel estimation schemes which precode *all* of the data symbols so that every subcarrier carries a linear combination of symbols, aking the resulting joint MLD problem infeasible. In addition, this also makes it infeasible to employ MIMO precoding per subcarrier across the transmit antennas. The proposed precoder is designed via a multi-stage optimization process that seeks to minimize both channel estimation error and symbol estimation error. For channel estimation purposes, the resulting optimal design offers low-cost features such as sign change and FFT while providing reasonable channel estimation performance for low mobility applications.**

***Index Terms*—Blind channel estimation, linear precoding, OFDM, MIMO, Hadamard matrix.**

I. Introduction

THERE has been a resurgence of interest in blind channel estimation schemes. The resurgence of interest is evidenced by a number of recent papers on this topic [1]–[3]. There are several reasons for the renewed interest in blind channel estimation. In current fourth-generation systems based on orthogonal frequency-division multiplexing (OFDM) technology, symbols are sent in blocks, so block-based, iterative channel estimation is enabled. Thus, for example, one could first employ a blind channel estimate to obtain initial symbol estimates, and then use the preliminary symbol estimates to obtain a higher fidelity channel estimate. The process can then be repeated, with an exchange of soft information, to iteratively enhance both the channel estimates and the data symbol estimates [4], [5]. Another motivating factor for the renewed interest in blind channel estimation is the emerging architecture of heterogenous networks with small cells, such as femtocells that are characterized by low-mobility users. As a result, there is a projected proliferation of low mobility applications, thereby opening up an opportunity for blind channel estimation, which typically requires a relatively large number of samples under quasi-static channel conditions for good performance.

There exist several blind channel estimation techniques for OFDM systems. One is the technique based on the redundancy introduced by the cyclic prefix (CP) [6], [7]. This CP-based method shows relatively good performance at high signal-to-noise ratio (SNR) but requires high computational complexity. Thus, a different framework for blind channel estimation has recently been proposed based on non-redundant precoding for OFDM systems [1], [8]–[15] to achieve good performance at moderate to low SNR at a reasonable computational complexity. In this paper, a new non-redundant, precoding-based blind channel estimation scheme for MIMO-OFDM systems is proposed in which only a small number of subcarriers, commensurate with the channel length, carry pre-coded data symbols for blind channel estimation purposes. The overwhelming majority of carriers transport symbols in a conventional manner, thereby enabling Maximum Likelihood Detection (MLD) of data symbols [16] and also simultaneous use of per-carrier precoding across antennas for data rate enhancement [17]. Previous precoding-based, blind channel estimation schemes [1], [8]–[15] were premised on linear precoding across *all* carriers which has a number of drawbacks. The linearly precoded symbols have to be jointly estimated. Thus, linearly precoding across *all* carriers renders MLD computationally intractable. Even minimum mean square error (MMSE) based symbol estimation is highly computationally burdensome due to the high-dimensional matrix inversion required. In addition, linearly precoding across all carriers (for blind channel estimation purposes) makes it extremely difficult to simultaneously employ per-carrier precoding across antennas for data rate enhancement [17]. Furthermore, such precoding cannot be used in the uplink where OFDMA is typically used, whereas the proposed sparse precoder design using only partial subcarriers can readily be applied to OFDMA-based systems.

In contrast to previous precoder designs (for blind channel estimation,) the underlying foundation for our sparse precoder design is a proof that only using a small number of linearly precoded carriers does not impact channel estimation performance, as long as the number of linearly precoded carriers is greater than the channel length under the equi-spaced sampling

Manuscript received December 17, 2013; revised June 12, 2014 and August 8, 2014; accepted August 20, 2014. Date of publication September 4, 2014; date of current version December 8, 2014. This work was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education under Grant 2013R1A1A2A10060852. The associate editor coordinating the review of this paper and approving it for publication was L. Sanguinetti.

S. Noh and M. D. Zoltowski are with the School of Electrical and Computer Engineering, Purdue University, West Lafayette, IN 47907 USA (e-mail: songnoh@purdue.edu; mikedz@ecn.purdue.edu).

Y. Sung is with the Department of Electrical Engineering, Korea Advanced Institute of Science and Technology (KAIST), Daejeon 305-701, Korea (e-mail: ysung@ee.kaist.ac.kr).

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

Digital Object Identifier 10.1109/TWC.2014.2354342

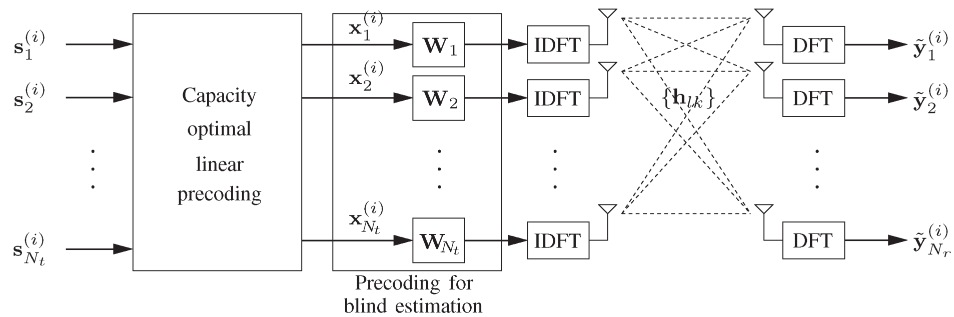


Fig. 1. System model.

structure. Subsequently operating within the framework of a sparse design, we develop an optimal precoder for both channel estimation performance and data symbol decoding performance (as opposed to previous works which didn't address optimality for either case), under a certain mean square error (MSE) criterion. Proofs are provided substantiating the optimality of certain key features of our design. For example, the proposed precoder is designed to be well-conditioned, thereby minimizing noise enhancement that may occur in the data symbol decoding process, again, only for those small number of carriers carrying linearly precoded data symbols. As a result of other key features of our design, the attendant channel estimation and decoding process can be implemented efficiently with low-cost operations such as sign change and fast Fourier Transform (FFT). Numerical results show the efficacy of the proposed design.

Some contents of this paper were presented at several conferences: The optimality of an equal absolute value for the square precoder matrix was presented in [18], and the idea of Hadamard precoder matrices was proposed in [19] and their optimality was proved [20] with the multiple user case. In this paper, the contents of the conference papers are summarized and furthermore the optimal precoder design is extended by relaxing the diagonal dominance constraint on the precoder and the optimality of equi-spaced precoding is proved with a perturbation analysis.

Notation Vectors and matrices are written in boldface with matrices in capitals. All vectors are column vectors. $\mathbf{A}^{\mathrm{T}}$ and $\mathbf{A}^{\mathrm{H}}$ indicate the transpose and conjugate transpose of $\mathbf{A}$, respectively. $\mathbf{A}(\mathcal{I},\mathcal{J})$ is the submatrix of $\mathbf{A}$ with the elements at rows $\mathcal{I}$ and columns $\mathcal{J}$ for some index sets $\mathcal{I}$ and $\mathcal{J}$. $[\mathbf{A}]_{i,j}$ denotes the element of $\mathbf{A}$ at the i-th row and j-th column, and $[a_{i,j}]$ is the matrix composed of element $a_{i,j}$ at the i-th row and j-th column. $\mathrm{diag}(d_1,\ldots,d_n)$ is the diagonal matrix composed of elements $d_1,\ldots,d_n$. $\mathbf{A}^\dagger$, $\|\mathbf{A}\|_F$, $\|\mathbf{A}\|_2$, and $\mathrm{tr}(\mathbf{A})$ denote the pseudo-inverse, Frobenius norm, L_2-norm, and trace of $\mathbf{A}$, respectively. $\otimes$ denotes the Kronecker product, and $\odot$ and $\oslash$ stand for the Hadamard (element-wise) product and division, respectively. $\mathbf{I}_n$ stands for the identity matrix of size n. $\mathbb{E}\{\mathbf{x}\}$ represents the expectation of $\mathbf{x}$. $\mathbf{x}\sim\mathcal{CN}(\boldsymbol{\mu},\boldsymbol{\Sigma})$ means that random vector $\mathbf{x}$ is complex Gaussian distributed with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$.

II. System Model and Background

We consider a $N_r\times N_t$ MIMO-OFDM system, where the transmitter has N_t transmit antennas and the receiver has N_r receive antennas as shown in Fig. 1. At the transmitter, the frequency-domain data vector $\mathbf{x}_k^{(i)}=[x_k^{(i)}(0),\ldots,x_k^{(i)}(N-1)]^{\mathrm{T}}$ for the i-th OFDM symbol assigned to transmit antenna k is linearly precoded by a precoder matrix $\mathbf{W}_k$ of size $N\times N$ for blind channel estimation, where N is the number of OFDM subcarriers, and then OFDM modulated and transmitted, i.e., processed by IDFT, attached by a cyclic prefix that is longer than the channel length, and transmitted. The transmitted signal from transmit antenna k to receive antenna l passes through a finite impulse response (FIR) channel $\mathbf{h}_{lk}=[h_{lk}(0),\ldots,h_{lk}(L-1)]^{\mathrm{T}}$, $1\le l\le N_r$, $1\le k\le N_t$, which is assumed to be time-invariant over a block of N_s successive OFDM symbols. (The sample data covariance matrix required for blind channel estimation will be computed by using the coherent N_s symbols later.) At the receiver, collecting the chips corresponding to the i-th OFDM symbol at receive antenna l and removing the cyclic prefix portion, we have the received time-domain signal vector for symbol time i at receive antenna l, given by

$$\mathbf{y}_l^{(i)}=\sum_{k=1}^{N_t}\mathbf{H}_{lk}\mathbf{F}^{\mathrm{H}}\mathbf{W}_k\mathbf{x}_k^{(i)}+\mathbf{n}_l^{(i)},\quad l=1,\ldots,N_r, \tag{1}$$

where $\mathbf{F}^{\mathrm{H}}$ is the normalized IDFT matrix, $\mathbf{n}^{(i)}\sim\mathcal{CN}(\mathbf{0},\sigma_n^2\mathbf{I})$ is a noise vector, and $\mathbf{H}_{lk}$ is a circulant channel matrix with $[\mathbf{h}_{lk}^{\mathrm{T}},0,\ldots,0]^{\mathrm{T}}$ as its first column. The final received signal in frequency domain is given by the DFT of $\mathbf{y}_l^{(i)}$ as

$$\tilde{\mathbf{y}}_l^{(i)}=\sum_{k=1}^{N_t}\tilde{\mathbf{H}}_{lk}\mathbf{W}_k\mathbf{x}_k^{(i)}+\tilde{\mathbf{n}}_l^{(i)},\quad l=1,\ldots,N_r, \tag{2}$$

where $\tilde{\mathbf{n}}_l^{(i)}=\mathbf{F}\mathbf{n}_l^{(i)}$, and $\tilde{\mathbf{H}}_{lk}=\mathbf{F}\mathbf{H}_{lk}\mathbf{F}^{\mathrm{H}}$ is a diagonal matrix the diagonal elements of which are given by

$$\tilde{\mathbf{h}}_{lk}=\tilde{\mathbf{F}}\mathbf{h}_{lk}. \tag{3}$$

Here, $\tilde{\mathbf{F}}$ is the $N\times L$ skinny DFT matrix, i.e., the matrix consisting of the first L columns of the $N\times N$ normalized DFT matrix $\mathbf{F}$. We assume that the input $\mathbf{x}_k^{(i)}$ to the precoder $\mathbf{W}_k$ is independent of the noise $\mathbf{n}_l^{(i)}$ and is the output of a linear MIMO precoder that is applied to the input data vector $\mathbf{s}_n^{(i)}$ of dimension $D(\le\min\{N_t,N_r\})$, $n=1,\ldots,N$, to enhance the data rate as usual in a typical MIMO system. The considered two-layer precoding strategy will be explained in detail in Section V. At this moment, we assume that $\mathbf{x}_k^{(i)}$ has zero mean with $\mathbb{E}\{\mathbf{x}_k^{(i)}\mathbf{x}_k^{(i)H}\}=\sigma_s^2\mathbf{I}$ and this assumption is satisfied by the proposed two-layer precoding strategy. Without loss of generality, we assume $\sigma_s^2 = 1$ and the superscript (i) representing the symbol time will be omitted if unnecessary.

A. Precoding-Based Blind Estimation for OFDM Systems

In this subsection, we briefly explain the precoding-based blind channel estimation for OFDM systems as the background for our further development in the next sections. The precoding-based blind channel estimation method is based on precoding at the transmitter and second-order statistics of the received signal. To understand the method, first note that the $N \times N$ covariance matrix $\mathbf{R}_l := \mathbb{E}\{\tilde{\mathbf{y}}_l \tilde{\mathbf{y}}_l^{\mathrm{H}}\}$ of the received signal (2) at antenna l at the receiver is given by

$$\begin{aligned} \mathbf{R}_l &= \tilde{\mathbf{H}}_{l1}\mathbf{Q}_1\tilde{\mathbf{H}}_{l1}^{\mathrm{H}} + \cdots + \tilde{\mathbf{H}}_{lN_t}\mathbf{Q}_{N_t}\tilde{\mathbf{H}}_{lN_t}^{\mathrm{H}} + \sigma_n^2\mathbf{I} \\ &= \left(\sum_{k=1}^{N_t} \tilde{\mathbf{h}}_{lk}\tilde{\mathbf{h}}_{lk}^{\mathrm{H}}\right) \odot \mathbf{Q} + \sigma_n^2\mathbf{I}, \end{aligned} \tag{4}$$

under the assumption of $\mathbf{Q}_1 = \cdots = \mathbf{Q}_{N_t} = \mathbf{Q}$, where $\mathbf{Q}_k = \mathbf{W}_k\mathbf{W}_k{}^{\mathrm{H}}$. We will refer to $\mathbf{Q}$ as the *square precoder matrix*. The diagonal and off-diagonal elements of $\mathbf{R}_l$ are given by $[\mathbf{R}_l]_{i,i} = \sum_{k=1}^{N_t} |\tilde{h}_{lk}(i)|^2[\mathbf{Q}]_{i,i} + \sigma_n^2$ and $[\mathbf{R}_l]_{i,j} = \sum_{k=1}^{N_t} \tilde{h}_{lk}(i)\tilde{h}_{lk}^*(j)[\mathbf{Q}]_{i,j}$ for $i \neq j$, respectively. From this, we have

$$\frac{[\mathbf{R}_l]_{i,j} - \delta_{ij}\sigma_n^2}{[\mathbf{Q}]_{i,j}} = \sum_{k=1}^{N_t} \tilde{h}_{lk}(i)\tilde{h}_{lk}^*(j), \tag{5}$$

where δ_{ij} is the Kronecker delta. Most of the precoding-based blind channel estimation methods for OFDM systems are based on (5). That is, from the knowledge of $\mathbf{R}_l$ and $\mathbf{Q}$ (and additionally σ_n^2 in some cases), one can properly construct a vector or a matrix from which the channel can be identified. For example, consider the method by Gao and Nallanathan [11]. In their method, they construct

$$\begin{aligned} \mathbf{v}_j &:= \left[\frac{[\mathbf{R}_l]_{1,j}}{[\mathbf{Q}]_{1,j}}, \ldots, \frac{[\mathbf{R}_l]_{j-1,j}}{[\mathbf{Q}]_{j-1,j}}, \frac{[\mathbf{R}_l]_{j+1,j}}{[\mathbf{Q}]_{j+1,j}}, \ldots, \frac{[\mathbf{R}_l]_{N,j}}{[\mathbf{Q}]_{N,j}}\right]^{\mathrm{T}}, \\ \tilde{\mathbf{F}}_j &:= \begin{bmatrix} \tilde{\mathbf{F}}(1:j-1,:) \\ \tilde{\mathbf{F}}(j+1:N,:) \end{bmatrix} = \begin{bmatrix} \mathbf{F}(1:j-1,1:L) \\ \mathbf{F}(j+1:N,1:L) \end{bmatrix}, \end{aligned} \tag{6}$$

where $\tilde{\mathbf{F}}_j$ is the matrix resulting from removing the j-th row of $\tilde{\mathbf{F}}$. Then, $\mathbf{v}_j = \tilde{\mathbf{F}}_j \sum_{k=1}^{N_t} \mathbf{h}_{lk}\tilde{h}_{lk}^*(j)$ from (3) and (5). Collecting the data from all the columns yields

$$\mathbf{J}_l^G := \left[\tilde{\mathbf{F}}_1^\dagger\mathbf{v}_1 \cdots \tilde{\mathbf{F}}_N^\dagger\mathbf{v}_N\right]\tilde{\mathbf{F}} = \sum_{k=1}^{N_t} \mathbf{h}_{lk}\mathbf{h}_{lk}^{\mathrm{H}} =: \mathbf{J}_l, \tag{7}$$

since $[\tilde{\mathbf{F}}_1^\dagger\mathbf{v}_1 \cdots \tilde{\mathbf{F}}_N^\dagger\mathbf{v}_N] = \sum_k \mathbf{h}_{lk}\tilde{\mathbf{h}}_{lk}^{\mathrm{H}} = (\sum_k \mathbf{h}_{lk}\mathbf{h}_{lk}^{\mathrm{H}})\tilde{\mathbf{F}}^{\mathrm{H}}$ and $\tilde{\mathbf{F}}^{\mathrm{H}}\tilde{\mathbf{F}} = \mathbf{I}_L$. Once $\mathbf{J}_l^G$ is constructed based on $\mathbf{v}_j$ which is in turn constructed from the received signal covariance matrix $\mathbf{R}_l$ and the square precoder matrix $\mathbf{Q}$, the channel can be identified up to rotational ambiguity by low rank decomposition of $\mathbf{J}_l^G$ [21], [22]. (If $\mathbf{Q} = c\mathbf{I}_N$ for some c, the construction of $\mathbf{J}_l$ in (7) is impossible because $\mathbf{J}_l$ is constructed from a set of vectors $\{\mathbf{v}_n\}$ which are obtained by element-wise division of $\mathbf{R}_l$ by $\mathbf{Q}$ as shown in (6).)

The essential point of the precoding-based blind channel estimation method is that the elements of the square precoder matrix $\mathbf{Q}$ corresponding to the data acquisition or sampling points from the data covariance matrix $\mathbf{R}_l$ should be non-zero to exploit (5). For maximal data acquisition for the blind channel estimation, the authors in [9], [11], [12], [14], [15] proposed the following square precoder matrix, $p \neq 0$:

$$\mathbf{Q}_p = \mathrm{diag}(1-p, \ldots, 1-p) + p\mathbf{1}\mathbf{1}^{\mathrm{T}}, \quad -(N-1)^{-1} < p < 1, \tag{8}$$

where $\mathbf{1}$ is the column vector composed of all ones. The matrix $\mathbf{Q}$ in (8) is parameterized by p and the value of p can be optimized by trading-off between channel estimation performance and data decoding performance. However, a dense precoding matrix $\mathbf{W}$ resulting from the square root operation on a fully dense square precoder matrix like $\mathbf{Q}_p$ in (8), mixes the signals of all subcarriers, and thus destroys the very desirable feature of MIMO-OFDM systems that each subcarrier provides an independent flat-fading $N_r \times N_t$ MIMO channel after IDFT and DFT. Note in (2) that the elements of the MIMO-precoded signal vector $\mathbf{x}_k^{(i)}$ are fully mixed by the precoder $\mathbf{W}$. Such linear precoding across *all* carriers makes it difficult to employ MLD of data symbols and per-carrier precoding across antennas for data rate enhancement, as later discussed in Section V.

III. The Proposed Sparse Precoder Structure

In this section, we propose a new precoder that has a sparse structure, aiming at ML detection of data symbols and low-cost operation required for blind channel estimation. We assume that the noise variance is unknown as in [8], [11]–[13], [23]. Let us consider the single-input single-output (SISO) case first and then extend to the MIMO case. In the SISO case, the received signal covariance matrix (4) at the receiver reduces to

$$\mathbf{R} = \tilde{\mathbf{h}}\tilde{\mathbf{h}}^{\mathrm{H}} \odot \mathbf{Q} + \sigma_n^2\mathbf{I}, \tag{9}$$

and thus we have $[\tilde{\mathbf{J}}]_{i,j} = [\mathbf{R}]_{i,j}/[\mathbf{Q}]_{i,j} = \tilde{h}(i)\tilde{h}^*(j), i \neq j$ and $[\tilde{\mathbf{J}}]_{i,i} = ([\mathbf{R}]_{i,i} - \sigma_n^2)/[\mathbf{Q}]_{i,i} = |\tilde{h}(i)|^2$ for $\tilde{\mathbf{J}} := \tilde{\mathbf{h}}\tilde{\mathbf{h}}^{\mathrm{H}}$. (The antenna indices are omitted for notational simplicity in the SISO case).

To reduce the number of subcarriers that are linearly precoded for blind channel estimation, we propose the following sampling scheme to select the elements from $\tilde{\mathbf{J}}$ for the blind identification of the channel. We choose the same T rows at T columns where $L \leq T < N$, but to avoid selecting the diagonal elements corrupted by the noise variance, with the application to the low-SNR case in mind,[1] we require that the indices $\mathcal{I} = \{i_1, \ldots, i_T\} \subset \{1, 2, \ldots, N\}$ of the selected rows are disjoint from those $\mathcal{J} = \{j_1, \ldots, j_T\}$ of the selected columns, i.e., $\mathcal{I} \cap \mathcal{J} = \emptyset$. Indeed, such a sampling scheme facilitates optimal precoder design in Lemmas 1 and 2. The blind channel estimation based on this sampling scheme is then

[1] In [24], the authors showed that at high SNR the diagonal elements are estimated better than the off-diagonal elements. At low SNR, on the other hand, the off-diagonal elements are estimated better.

given as follows. Note first that $[\mathbf{R}]_{i_s,j_t}/[\mathbf{Q}]_{i_s,j_t} = \tilde{h}(i_s)\tilde{h}^*(j_t)$ for $i_s \in \mathcal{I}$ and $j_t \in \mathcal{J}$. Construct

$$\mathbf{v}'_{j_t} := \left[\frac{[\mathbf{R}]_{i_1,j_t}}{[\mathbf{Q}]_{i_1,j_t}}, \ldots, \frac{[\mathbf{R}]_{i_T,j_t}}{[\mathbf{Q}]_{i_T,j_t}}\right]^{\mathrm{T}} = \tilde{\mathbf{F}}(\mathcal{I},:)\mathbf{h}\tilde{h}^*(j_t). \quad (10)$$

Then, we have $\tilde{\mathbf{F}}(\mathcal{I},:)^\dagger \mathbf{v}'_{j_t} = \mathbf{h}\tilde{h}^*(j_t)$ since the matrix $\tilde{\mathbf{F}}(\mathcal{I},:)$ is the matrix composed of the rows $\mathcal{I}$ of the skinny DFT matrix $\tilde{\mathbf{F}}$ and is a $T \times L$ Vandermonde matrix with full column rank if $T \geq L$. (So is $\tilde{\mathbf{F}}(\mathcal{J},:)$ in the below.) Collecting the data from all the columns $\mathcal{J}$ yields

$$\mathbf{J} = \mathbf{J}_{\mathcal{I},\mathcal{J}} := \tilde{\mathbf{F}}(\mathcal{I},:)^\dagger \left[\mathbf{v}'_{j_1} \cdots \mathbf{v}'_{j_T}\right] \left(\tilde{\mathbf{F}}(\mathcal{J},:)^\dagger\right)^{\mathrm{H}} = \mathbf{h}\mathbf{h}^{\mathrm{H}}, \quad (11)$$

then the channel can be obtained from $\mathbf{J}_{\mathcal{I},\mathcal{J}}$ up to scalar ambiguity by subspace decomposition.

Note that $\tilde{\mathbf{F}}(\mathcal{I},:)^\dagger$ in (11) is outside the matrix $[\mathbf{v}'_{j_1} \cdots \mathbf{v}'_{j_T}]$ in the middle, and this makes an implementation of the proposed algorithm easy, as seen in Section V-A. Under the proposed structure, the square precoder matrix $\mathbf{Q}$ has non-zero off-diagonal elements at $\mathcal{I} \times \mathcal{J}$. Due to its symmetric structure by its definition, however, it has non-zero off-diagonal elements at $\mathcal{I} \times \mathcal{J} \cup \mathcal{J} \times \mathcal{I}$, and the minimum number of non-zero off-diagonal elements is $2L^2$ achieved when $T = L$. Under the proposed structure of $\mathbf{Q}$, using $\mathbf{R}(\mathcal{I},\mathcal{J})$ is equivalent to using $\mathbf{R}(\mathcal{J},\mathcal{I})$ for the estimation of $\mathbf{J}$ via $\mathbf{J}_{\mathcal{I},\mathcal{J}}$ or $\mathbf{J}_{\mathcal{J},\mathcal{I}}$ due to its perfect symmetry. An example of the proposed $\mathbf{Q}$ with $N = 8$, $T = 2$, $\mathcal{I} = \{2, 6\}$ and $\mathcal{J} = \{1, 5\}$ is shown in (12) (see equation at the bottom of the page).

Next, consider the MIMO case. We define the overall received vector across the receive antennas as $\tilde{\mathbf{y}} := [\tilde{\mathbf{y}}_1^{\mathrm{T}}, \ldots, \tilde{\mathbf{y}}_{N_r}^{\mathrm{T}}]^{\mathrm{T}}$. From (2), the overall frequency-domain channel matrix $\tilde{\mathbf{H}} \in \mathbb{C}^{NN_r \times NN_t}$ is described as

$$\tilde{\mathbf{H}} = \begin{bmatrix} \tilde{\mathbf{H}}_{11} & \cdots & \tilde{\mathbf{H}}_{1N_t} \\ \vdots & \ddots & \vdots \\ \tilde{\mathbf{H}}_{N_r1} & \cdots & \tilde{\mathbf{H}}_{N_rN_t} \end{bmatrix}. \quad (13)$$

Then, the signal covariance matrix $\mathbf{R}$ is given by

$$\mathbf{R} = \mathbb{E}\{\tilde{\mathbf{y}}\tilde{\mathbf{y}}^{\mathrm{H}}\} = \tilde{\mathbf{H}}(\mathbf{I}_{N_t} \otimes \mathbf{Q})\tilde{\mathbf{H}}^{\mathrm{H}} + \sigma_n^2 \mathbf{I}_{NN_r} = \begin{bmatrix} \mathbf{R}_{11} & \cdots & \mathbf{R}_{1N_r} \\ \vdots & \ddots & \vdots \\ \mathbf{R}_{N_r1} & \cdots & \mathbf{R}_{N_rN_r} \end{bmatrix}, \quad (14)$$

where $\mathbf{R}_{pq} = (\sum_{k=1}^{N_t} \tilde{\mathbf{h}}_{pk}\tilde{\mathbf{h}}_{qk}^{\mathrm{H}}) \odot \mathbf{Q} + \delta_{pq}\sigma_n^2 \mathbf{I}_N$ for $p, q \in \{1, \ldots, N_r\}$. Similarly to (10), we divide the element $[\mathbf{R}_{pq}]_{i_s,j_t}$ with $[\mathbf{Q}]_{i_s,j_t}$ to form $\mathbf{V}_{pq} := [\mathbf{v}_{pq,j_1}, \ldots, \mathbf{v}_{pq,j_T}] \in \mathbb{C}^{T\times T}$, where $i_s \in \mathcal{I}, j_t \in \mathcal{J}$ and $\mathbf{v}_{pq,j_t} = [[\mathbf{R}_{pq}]_{i_1,j_t}/[\mathbf{Q}]_{i_1,j_t}, \ldots, [\mathbf{R}_{pq}]_{i_T,j_t}/[\mathbf{Q}]_{i_T,j_t}]^{\mathrm{T}} = \tilde{\mathbf{F}}(\mathcal{I},:)\sum_{k=1}^{N_t} \mathbf{h}_{pk}\tilde{h}_{qk}^*(j_t)$. Denote by $\mathbf{h}_k = [\mathbf{h}_{1k}^{\mathrm{T}}, \ldots, \mathbf{h}_{N_rk}^{\mathrm{T}}]^{\mathrm{T}}$ the combined channel vector from transmit antenna k to all receive antennas and by $\mathbf{V} \in \mathbb{C}^{N_rT\times N_rT}$ the composite matrix defined by the submatrices $\{\mathbf{V}_{pq}\}$. Then, $\mathbf{J} \in \mathbb{C}^{LN_r\times LN_r}$ is constructed as $\mathbf{J} = (\mathbf{I}_{N_r} \otimes \tilde{\mathbf{F}}(\mathcal{I},:)^\dagger)\mathbf{V}(\mathbf{I}_{N_r} \otimes \tilde{\mathbf{F}}(\mathcal{J},:)^\dagger)^{\mathrm{H}} = \sum_{k=1}^{N_t} \mathbf{h}_k\mathbf{h}_k^{\mathrm{H}}$ and finally a channel estimate can be obtained from $\mathbf{J}$ up to an $N_t \times N_t$ unitary ambiguity matrix by subspace decomposition.

IV. Optimal Sparse Precoder Design

In this section, we consider the optimal precoder design under the proposed precoder structure presented in the previous section. For the optimality criterion, one can consider the Cramér-Rao bound for blind estimation for the data model (2). However, such a standard design method is intractable for the considered problem because the problem is combinatorial, i.e., we should choose the discrete sets ($\mathcal{I}$, $\mathcal{J}$) and the (complex-valued) values of the elements of $\mathbf{Q}$ corresponding to the positions determined by $\mathcal{I}$ and $\mathcal{J}$. To circumvent this difficulty, we exploit the Markov structure of the estimation statistics, i.e., $\tilde{\mathbf{y}}_l^{(i)} \to \hat{\mathbf{R}}_l \overset{(a)}{\to} \hat{\mathbf{J}}_{\mathcal{I},\mathcal{J}} \to [\hat{\mathbf{h}}_{l1}, \ldots, \hat{\mathbf{h}}_{lN_t}]$, where the hat notation denotes the estimated quantity. Note that the design of the precoder (square) matrix impacts the step (a) and the following step in the Markov chain. Thus, our design criterion is to minimize the MSE in the step (a), i.e., the MSE between $\hat{\mathbf{J}}_{\mathcal{I},\mathcal{J}}$ and $\mathbf{J}_{\mathcal{I},\mathcal{J}}(= \mathbf{J})$. This is a valid criterion since the precoding-based blind estimation uses $\hat{\mathbf{R}}_l$ not $\tilde{\mathbf{y}}_l^{(i)}$ directly to construct the estimation statistic $\hat{\mathbf{J}}_{\mathcal{I},\mathcal{J}}$.

Define the difference between the estimated (or sample) data covariance matrix and the true data covariance matrix $\Delta\mathbf{R}_l := \hat{\mathbf{R}}_l - \mathbf{R}_l$ where $\hat{\mathbf{R}}_l = (1/N_s)\sum_{i=1}^{N_s} \tilde{\mathbf{y}}_l^{(i)}\tilde{\mathbf{y}}_l^{(i)\mathrm{H}}$ from (2). Then, the difference between $\hat{\mathbf{v}}'_{j_t}$ and $\mathbf{v}'_{j_t}$ in (10) is given by

$$\Delta\mathbf{v}'_{j_t} = \left[\frac{[\Delta\mathbf{R}_l]_{i_1,j_t}}{[\mathbf{Q}]_{i_1,j_t}}, \ldots, \frac{[\Delta\mathbf{R}_l]_{i_T,j_t}}{[\mathbf{Q}]_{i_T,j_t}}\right]^{\mathrm{T}}, \quad (15)$$

where $i_s \in \mathcal{I}, j_t \in \mathcal{J}$, and the estimation error of $\mathbf{J}_{\mathcal{I},\mathcal{J}}$ is given by

$$\Delta\mathbf{J} := \hat{\mathbf{J}}_{\mathcal{I},\mathcal{J}} - \mathbf{J}_{\mathcal{I},\mathcal{J}} = \tilde{\mathbf{F}}(\mathcal{I},:)^\dagger \left[\Delta\mathbf{v}'_{j_1} \cdots \Delta\mathbf{v}'_{j_T}\right]\left(\tilde{\mathbf{F}}(\mathcal{J},:)^\dagger\right)^{\mathrm{H}}. \quad (16)$$

$$\mathbf{Q} = \begin{bmatrix} q_{1,1} & q_{j_1,i_1}\,(= q_{i_1,j_1}^*) & 0 & 0 & 0 & q_{j_1,i_2}\,(= q_{i_2,j_1}^*) & 0 & 0 \\ q_{i_1,j_1} & q_{2,2} & 0 & 0 & q_{i_1,j_2} & 0 & 0 & 0 \\ 0 & 0 & q_{3,3} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \ddots & 0 & 0 & 0 & 0 \\ 0 & q_{j_2,i_1}\,(= q_{i_1,j_2}^*) & 0 & 0 & \ddots & q_{j_2,i_2}\,(= q_{i_2,j_2}^*) & 0 & 0 \\ q_{i_2,j_1} & 0 & 0 & 0 & q_{i_2,j_2} & \ddots & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & q_{N,N} \end{bmatrix} \quad (12)$$

The Frobenius norm of $\Delta\mathbf{J}$ is bounded by

$$\begin{aligned}\|\Delta\mathbf{J}\|_F &\le \sqrt{L}\|\Delta\mathbf{J}\|_2\\ &\le \sqrt{L}\left\|\tilde{\mathbf{F}}(\mathcal{I},:)^\dagger\right\|_2 \left\|[\Delta\mathbf{v}'_{j_1}\cdots\Delta\mathbf{v}'_{j_T}]\right\|_2\\ &\quad\times\left\|\left(\tilde{\mathbf{F}}(\mathcal{J},:)^\dagger\right)^{\mathrm{H}}\right\|_2.\end{aligned}$$

First, let us consider the design of $\mathcal{I}$ and $\mathcal{J}$. For simplicity, we assume that $T(\ge L)$ is a divisor of N from here on. (Since L is a divisor of N in typical OFDM systems, it will be shown shortly that a big value of T is not required but $T=L$ is sufficient.) When the condition numbers of $\tilde{\mathbf{F}}(\mathcal{I},:)$ and $\tilde{\mathbf{F}}(\mathcal{J},:)$ are small, the enhancement of the perturbation $[\Delta\mathbf{v}'_{j_1}\cdots\Delta\mathbf{v}'_{j_T}]$ in (16) is small. The minimum value of one for the condition numbers of $\tilde{\mathbf{F}}(\mathcal{I},:)$ and $\tilde{\mathbf{F}}(\mathcal{J},:)$ is achieved only if the indices of $\mathcal{I}$ and $\mathcal{J}$ are equi-spaced when T is a divisor of N by the property of the skinny DFT matrix. This implies that we should use equi-spaced index sets for $\mathcal{I}$ and $\mathcal{J}$. (This will be revisited soon.). In this case, the optimal indices for $\mathcal{I}$ and $\mathcal{J}$ are given respectively by

$$i_t=(t-1)\frac{N}{T}+c \text{ and } j_t=(t-1)\frac{N}{T}+d, t\in\{1,\ldots,T\} \quad (17)$$

for some $c,d\in\{1,\ldots,N/T-1\}$ and $c\ne d$. In addition to this equi-spaced index condition, we have the following conditions for $\mathbf{Q}$:

(C.1) $\mathbf{Q}$ is positive semi-definite.

(C.2) The diagonal elements of $\mathbf{Q}$ are one.

(C.3) $\mathbf{Q}$ has non-zero off-diagonal elements only at $\mathcal{I}\times\mathcal{J}$ and $\mathcal{J}\times\mathcal{I}$, where $|\mathcal{I}|=|\mathcal{J}|=T$, $L\le T<N$, and $\mathcal{I}\cap\mathcal{J}=\emptyset$. ($\mathcal{I}$ and $\mathcal{J}$ are equi-spaced.)

Condition *(C.1)* is for $\mathbf{Q}=\mathbf{W}_k\mathbf{W}_k^{\mathrm{H}}$ to be decomposed with a unique square root [21]. Condition *(C.2)* guarantees that there is no power boosting. (There is no reason to boost some subcarriers.) Condition *(C.3)* is the considered sparsity constraint discussed in the previous section. Thus, the optimal precoder design problem is formulated as

$$\min_{\mathbf{Q}}\mathbb{E}\left\{\|\Delta\mathbf{J}\|_F^2\right\}, \text{ subject to } (C.1),\ (C.2), \text{ and } (C.3), \quad (18)$$

where the expectation is over both the noise and channel[2] distributions under the assumption of independent and identically distributed (i.i.d.) Rayleigh fading channels, i.e., $\mathbf{h}\sim\mathcal{CN}(\mathbf{0},\sigma_h^2\mathbf{I}_L)$. The following theorem provides a property of an optimal square precoder matrix.

Theorem 1: Under the constraints *(C.1)*, *(C.2)*, and *(C.3)*, $\mathbb{E}\{\|\Delta\mathbf{J}\|_F^2\}$ is minimized if the absolute values of the off-diagonal elements of $\mathbf{Q}$ at $\mathcal{I}\times\mathcal{J}$ are identical with the value of $1/\sqrt{T}$, under the assumption of i.i.d. Rayleigh fading channels, i.e., $\mathbf{h}\sim\mathcal{CN}(\mathbf{0},\sigma_h^2\mathbf{I}_L)$.

[2]The precoder design for (blind) channel estimation should not assume a specific channel realization. Instead, the channel distribution should be exploited.

Proof: $\mathbb{E}\{\|\Delta\mathbf{J}\|_F^2\}$ in (18) is given by

$$\mathbb{E}\{\|\Delta\mathbf{J}\|_F^2\}=\mathbb{E}\left\{\left\|\tilde{\mathbf{F}}(\mathcal{I},:)^\dagger[\Delta\mathbf{v}'_{j_1}\cdots\Delta\mathbf{v}'_{j_T}]\left(\tilde{\mathbf{F}}(\mathcal{J},:)^\dagger\right)^{\mathrm{H}}\right\|_F^2\right\} \quad (19)$$

$$=\mathbb{E}\left\{\mathrm{tr}\left(\tilde{\mathbf{F}}(\mathcal{J},:)^\dagger\,[\Delta\mathbf{v}'_{j_1}\cdots\Delta\mathbf{v}'_{j_T}]^{\mathrm{H}}\left(\tilde{\mathbf{F}}(\mathcal{I},:)^\dagger\right)^{\mathrm{H}}\times\tilde{\mathbf{F}}(\mathcal{I},:)^\dagger[\Delta\mathbf{v}'_{j_1}\cdots\Delta\mathbf{v}'_{j_T}]\left(\tilde{\mathbf{F}}(\mathcal{J},:)^\dagger\right)^{\mathrm{H}}\right)\right\} \quad (20)$$

$$=\mathbb{E}\left\{\mathrm{tr}\left([\Delta\mathbf{v}'_{j_1}\cdots\Delta\mathbf{v}'_{j_T}]^{\mathrm{H}}\left(\tilde{\mathbf{F}}(\mathcal{I},:)^\dagger\right)^{\mathrm{H}}\tilde{\mathbf{F}}(\mathcal{I},:)^\dagger\times[\Delta\mathbf{v}'_{j_1}\cdots\Delta\mathbf{v}'_{j_T}]\left(\tilde{\mathbf{F}}(\mathcal{J},:)^\dagger\right)^{\mathrm{H}}\tilde{\mathbf{F}}(\mathcal{J},:)^\dagger\right)\right\} \quad (21)$$

$$=\mathrm{tr}\left(\mathbb{E}\left\{[\Delta\mathbf{v}'_{j_1}\cdots\Delta\mathbf{v}'_{j_T}]^{\mathrm{H}}\left(\tilde{\mathbf{F}}(\mathcal{I},:)\tilde{\mathbf{F}}(\mathcal{I},:)^{\mathrm{H}}\right)^\dagger\times[\Delta\mathbf{v}'_{j_1}\cdots\Delta\mathbf{v}'_{j_T}]\right\}\left(\tilde{\mathbf{F}}(\mathcal{J},:)\tilde{\mathbf{F}}(\mathcal{J},:)^{\mathrm{H}}\right)^\dagger\right), \quad (22)$$

where (20) holds because $\|\mathbf{A}\|_F^2=\mathrm{tr}(\mathbf{A}^{\mathrm{H}}\mathbf{A})$ and (21) holds because $\mathrm{tr}(\mathbf{ABC})=\mathrm{tr}(\mathbf{BCA})$. By Lemma 1 in Appendix and the fact that $[(\tilde{\mathbf{F}}(\mathcal{I},:)^\dagger)^{\mathrm{H}}\tilde{\mathbf{F}}(\mathcal{I},:)^\dagger]_{i,i}=NL/T^2$ for any equi-spaced index set $\mathcal{I}$, we reduce (22) to $\mathbb{E}\{\|\Delta\mathbf{J}\|_F^2\}=C\sum_{i_s\in\mathcal{I},j_t\in\mathcal{J}}1/|[\mathbf{Q}]_{i_s,j_t}|^2$ where $C:=N^2L^2/N_sT^4((LN_t/N)\sigma_h^2+\sigma_n^2)^2$. Then, the optimal design problem (18) can be rewritten as

$$\min_{\mathbf{Q}}\quad C\sum_{i_s\in\mathcal{I},j_t\in\mathcal{J}}\frac{1}{|[\mathbf{Q}]_{i_s,j_t}|^2}\quad \text{subject to}\quad (C.1),\ (C.2) \text{ and } (C.3). \quad (23)$$

With some permutation matrix $\mathbf{\Pi}_1$, *(C.1)* is equivalent to $\bar{\mathbf{Q}}\succeq 0$ since

$$\mathbf{\Pi}_1^{\mathrm{T}}\mathbf{Q}\mathbf{\Pi}_1=\begin{bmatrix}\bar{\mathbf{Q}} & \mathbf{0}\\ \mathbf{0} & \mathbf{I}_{N-2T}\end{bmatrix}, \quad (24)$$

where $\bar{\mathbf{Q}}$ contains all the non-zero off-diagonal elements of $\mathbf{Q}$ and some zero elements. It can easily shown that $\bar{\mathbf{Q}}$ can be expressed as

$$\mathbf{\Pi}_2\bar{\mathbf{Q}}\mathbf{\Pi}_2^{\mathrm{T}}=\begin{bmatrix}\mathbf{I}_T & \bar{\bar{\mathbf{Q}}}^{\mathrm{H}}\\ \bar{\bar{\mathbf{Q}}} & \mathbf{I}_T\end{bmatrix}, \quad (25)$$

where

$$\bar{\bar{\mathbf{Q}}}:=\begin{bmatrix}q_{i_1,j_1} & q_{i_1,j_2} & \cdots & q_{i_1,j_T}\\ q_{i_2,j_1} & q_{i_2,j_2} & \cdots & q_{i_2,j_T}\\ \vdots & \vdots & & \vdots\\ q_{i_T,j_1} & q_{i_T,j_2} & \cdots & q_{i_T,j_T}\end{bmatrix}, \quad (26)$$

which contains only all the non-zero off-diagonal elements of $\mathbf{Q}$, and $\mathbf{\Pi}_2$ is another permutation matrix. Then, the eigenvalues of $\mathbf{\Pi}_2\bar{\mathbf{Q}}\mathbf{\Pi}_2^{\mathrm{T}}$ are given by $1\pm\sigma_t$, where $\{\sigma_t, t=1,\ldots,T\}$ are the singular values of $\bar{\bar{\mathbf{Q}}}$. For the positive semi-definiteness of $\mathbf{\Pi}_2\bar{\mathbf{Q}}\mathbf{\Pi}_2^{\mathrm{T}}$ (consequently of $\bar{\mathbf{Q}}$), σ_t should not be greater than one. Thus, the optimization problem (23) can be rewritten as

$$\min_{\mathbf{Q}} \quad C \sum_{i_s\in\mathcal{I}, j_t\in\mathcal{J}} \frac{1}{|[\mathbf{Q}]_{i_s,j_t}|^2}$$
$$\text{subject to} \quad 0<\sigma_t\le 1,\ t=1,\ldots,T,\ (C.2),\ \text{and}\ (C.3). \tag{27}$$

Note from (26) that the constraints $0<\sigma_t\le 1$, $t=1,\ldots,T$ can be replaced with

$$\sum_{i_s\in\mathcal{I}, j_t\in\mathcal{J}} |[\mathbf{Q}]_{i_s,j_t}|^2 = \mathrm{tr}(\bar{\bar{\mathbf{Q}}}\bar{\bar{\mathbf{Q}}}^{\mathrm{H}}) = \sum_{t=1}^{T}\sigma_t^2 \le T. \tag{28}$$

Since the objective function in (27) is a convex function with respect to $|[\mathbf{Q}]_{i_s,j_t}|^2$, the minimum occurs when $|[\mathbf{Q}]_{i_s,j_t}|$ for every (i_s, j_t) is the same as $1/\sqrt{T}$. ■

Corollary 1: With the absolute value $|\mathbf{Q}_{i_s,j_t}| = \rho/T$ for all (i_s, j_t) for some $\rho\le\sqrt{T}$, the minimum MSE $\mathbb{E}\{\|\Delta\mathbf{J}\|_F^2\}$ is given by

$$\mathbb{E}\left\{\|\Delta\mathbf{J}\|_F^2\right\} = CT^2\frac{T^2}{\rho^2} = \frac{N^2L^2}{N_s\rho^2}\left(\frac{LN_t}{N}\sigma_h^2+\sigma_n^2\right)^2, \tag{29}$$

which is independent of T.

The independence of the minimum MSE of T is crucial. We can choose T as the smallest divisor of N larger than or equal to L for the maximum sparsity. When L itself is designed to be a divisor of N as in most practical OFDM systems, simply $T=L$. The sparsity of the proposed square precoder matrix is essential for the compatibility of the blind estimation precoding with MLD of data symbols and per-carrier precoding across antenna for data rate enhancement, which will be discussed in Section V. As noticed by other authors [10], here we also have a non-zero lower bound for the minimum MSE due to the σ_h^2 term in the RHS of (29) as the SNR tends to infinity, i.e., $\sigma_n^2\to 0$, whereas it vanishes as N_s increases.

Now return to the issue of equi-spaced index sets. Here we consider the SISO case for simplicity. The sample data covariance matrix is given by

$$\begin{aligned}\hat{\mathbf{R}} =&\ \tilde{\mathbf{H}}\mathbf{W}\left(\frac{1}{N_s}\sum_{i=1}^{N_s}\mathbf{x}^{(i)}\mathbf{x}^{(i)\mathrm{H}}\right)\mathbf{W}^{\mathrm{H}}\tilde{\mathbf{H}}^{\mathrm{H}}\\ &+\frac{1}{N_s}\sum_{i=1}^{N_s}\left(\tilde{\mathbf{H}}\mathbf{W}\mathbf{x}^{(i)}\tilde{\mathbf{n}}^{(i)\mathrm{H}}+\tilde{\mathbf{n}}^{(i)}\mathbf{x}^{(i)\mathrm{H}}\mathbf{W}^{\mathrm{H}}\tilde{\mathbf{H}}^{\mathrm{H}}\right)\\ &+\frac{1}{N_s}\sum_{i=1}^{N_s}\tilde{\mathbf{n}}^{(i)}\tilde{\mathbf{n}}^{(i)\mathrm{H}}.\end{aligned} \tag{30}$$

The first term in the RHS of (30) is the desired term, the second term is the cross-correlation between the signal and the noise, and the third term is the sample noise covariance matrix. The possible error in the first term is by the signal correlation structure not by the noise. The second and third terms are caused by the noise. At low SNR, the second term is negligible compared to the third term because of the uncorrelatedness of the signal and the noise. Thus, we consider the contribution of the third term, which is given by

$$\Delta\mathbf{J}\left(\frac{1}{N_s}\sum_{i=1}^{N_s}\tilde{\mathbf{n}}^{(i)}\tilde{\mathbf{n}}^{(i)\mathrm{H}}\right) = \tilde{\mathbf{F}}(\mathcal{I},:)^{\dagger}\left[\mathbf{z}'_{j_1}\cdots\mathbf{z}'_{j_T}\right]\left(\tilde{\mathbf{F}}(\mathcal{J},:)^{\dagger}\right)^{\mathrm{H}}, \tag{31}$$

where $\mathbf{z}'_{j_t} = [z_{i_1,j_t},\ldots,z_{i_T,j_t}]^{\mathrm{T}}$ and $z_{i_s,j_t} = (1/N_s[\mathbf{Q}]_{i_s,j_t})\sum_{i=1}^{N_s}\tilde{n}^{(i)}(i_s)\tilde{n}^{(i)*}(j_t)$. The following theorem shows that equi-spaced index sets are indeed optimal for minimizing the effect of the noise.

Theorem 2: For the estimation of $\mathbf{J}$ based on $\hat{\mathbf{J}}_{\mathcal{I},\mathcal{J}}$ constructed from the sample data covariance matrix $\hat{\mathbf{R}}_l$, the MSE due to the sample noise covariance matrix is minimized by equi-spaced sampling from the sample data covariance matrix under the assumption that T divides N exactly and all $|[\mathbf{Q}]_{i_s,j_t}|$ are equal.

Proof: The MSE for estimating $\mathbf{J}$ due to the noise covariance matrix is given by

$$\begin{aligned}&\mathbb{E}\left\{\left\|\tilde{\mathbf{F}}(\mathcal{I},:)^{\dagger}\left[\mathbf{z}'_{j_1}\cdots\mathbf{z}'_{j_T}\right]\left(\tilde{\mathbf{F}}(\mathcal{J},:)^{\dagger}\right)^{\mathrm{H}}\right\|_F^2\right\}\\ &\overset{(a)}{=}\mathbb{E}\left\{\mathrm{tr}\left(\tilde{\mathbf{F}}(\mathcal{J},:)^{\dagger}\left[\mathbf{z}'_{j_1}\cdots\mathbf{z}'_{j_T}\right]^{\mathrm{H}}\left(\tilde{\mathbf{F}}(\mathcal{I},:)^{\dagger}\right)^{\mathrm{H}}\tilde{\mathbf{F}}(\mathcal{I},:)^{\dagger}\left[\mathbf{z}'_{j_1}\cdots\mathbf{z}'_{j_T}\right]\right.\right.\\ &\qquad\left.\left.\times\left(\tilde{\mathbf{F}}(\mathcal{J},:)^{\dagger}\right)^{\mathrm{H}}\right)\right\}\\ &\overset{(b)}{=}\mathrm{tr}\left(\mathbb{E}\left\{\left[\mathbf{z}'_{j_1}\cdots\mathbf{z}'_{j_T}\right]^{\mathrm{H}}\left(\tilde{\mathbf{F}}(\mathcal{I},:)\tilde{\mathbf{F}}(\mathcal{I},:)^{\mathrm{H}}\right)^{\dagger}\left[\mathbf{z}'_{j_1}\cdots\mathbf{z}'_{j_T}\right]\right\}\right.\\ &\qquad\left.\times\left(\tilde{\mathbf{F}}(\mathcal{J},:)\tilde{\mathbf{F}}(\mathcal{J},:)^{\mathrm{H}}\right)^{\dagger}\right)\\ &\overset{(c)}{=}c\sum_{i=1}^{T}\left[\mathbf{G}_{\mathcal{I}}^{\dagger}\right]_{i,i}\sum_{j=1}^{T}\left[\mathbf{G}_{\mathcal{J}}^{\dagger}\right]_{j,j},\end{aligned} \tag{32}$$

for some positive constant c, where the expectation is over the noise distribution, $\mathbf{G}_{\mathcal{I}} = \tilde{\mathbf{F}}(\mathcal{I},:)\tilde{\mathbf{F}}(\mathcal{I},:)^{\mathrm{H}}$ and $\mathbf{G}_{\mathcal{J}} = \tilde{\mathbf{F}}(\mathcal{J},:)\tilde{\mathbf{F}}(\mathcal{J},:)^{\mathrm{H}}$. Here, (a) holds because $\|\mathbf{A}\|_F^2 = \mathrm{tr}(\mathbf{A}^{\mathrm{H}}\mathbf{A})$, (b) holds because $\mathrm{tr}(\mathbf{ABC}) = \mathrm{tr}(\mathbf{BCA})$ and (c) holds by Lemma 2 in Appendix. Since $\tilde{\mathbf{F}}(\mathcal{I},:)$ and $\tilde{\mathbf{F}}(\mathcal{J},:)$ are $T\times L$ Vandermonde matrices, we have $\mathrm{tr}(\mathbf{G}_{\mathcal{I}}) = \mathrm{tr}(\mathbf{G}_{\mathcal{J}}) = TL/N$ for any index set $\mathcal{I}$ and $\mathcal{J}$ such that $|\mathcal{I}| = |\mathcal{J}| = T$. Let the non-zero eigenvalues of $\mathbf{G}_{\mathcal{I}}^{\dagger}$ and $\mathbf{G}_{\mathcal{J}}^{\dagger}$ be $\{\lambda_i^{\mathcal{I}}\}$ and $\{\lambda_j^{\mathcal{J}}\}$, respectively. Then, the non-zero eigenvalues of $\mathbf{G}_{\mathcal{I}}$ and $\mathbf{G}_{\mathcal{J}}$ are $\{1/\lambda_i^{\mathcal{I}}\}$ and $\{1/\lambda_j^{\mathcal{J}}\}$, respectively. Thus, the minimization of the MSE (32) is reformulated as the following optimization problem:

$$\begin{aligned}&\min_{\{\lambda_i^{\mathcal{I}},\lambda_j^{\mathcal{J}}\}} \quad \sum_{i=1}^{L}\lambda_i^{\mathcal{I}}\sum_{j=1}^{L}\lambda_j^{\mathcal{J}}\\ &\text{subject to} \quad \sum_{i=1}^{L}\frac{1}{\lambda_i^{\mathcal{I}}} = \frac{TL}{N},\ \sum_{j=1}^{L}\frac{1}{\lambda_j^{\mathcal{J}}} = \frac{TL}{N},\ \lambda_i^{\mathcal{I}},\ \text{and}\\ &\qquad\qquad \lambda_i^{\mathcal{I}},\lambda_j^{\mathcal{J}}>0.\end{aligned} \tag{33}$$

Due to the separable structure of the problem (33), the above optimization can easily be solved using Karush-Kuhn-Tucker conditions. The minimum occurs when all the non-zero eigenvalues are the same. In the case that T divides N exactly, this occurs only when the indices are equi-spaced from the property of the skinny DFT matrix $\tilde{\mathbf{F}}$. ■

Theorem 2 is complementary to Theorem 1 since it requires the off-diagonal values of $\mathbf{Q}$ have the same absolute value. Still it provides an insight into the impact of the noise on sampling. Negi and Cioffi showed that the equi-spaced pilot tone insertion is optimal for the training-based channel estimation [25]. Interestingly, Theorem 2 shows that the equi-spaced sampling of the sample data covariance matrix is optimal for the precoding-based blind channel estimation if all the off-diagonal elements of the square precoder matrix is equally weighted.

V. Two-Step Linear Precoding

In this section, we propose a two-step linear precoding architecture composed of the blind channel estimation precoding presented in the previous sections and the data rate enhancing MIMO precoding. For the compatibility of the blind precoding with the data rate enhancing precoding, we further exploit the sparsity of our precoder square matrix. Based on the result of the previous section, we set $T = L$, which is assumed to be a divisor of N, and set ρ/T for the absolute value of every element of $\mathbf{Q}(\mathcal{I}, \mathcal{J})$ for some $0 < \rho \leq \sqrt{T}$. ($[\mathbf{Q}]_{i,i} = 1$.) Here, ρ is the control parameter determining the trade-off between channel estimation and data decoding. For all such ρ, $\mathbf{Q}$ is positive semidefinite. We further assume that $T = 2^m$ for some non-negative integer m on practical purpose. Our design goal for the precoder matrix $\mathbf{W}$ is to preserve the sparsity of the square precoder matrix. Our approach to this is not to obtain $\mathbf{W} = \sqrt{\mathbf{Q}}$ by the simple eigen-decomposition of $\mathbf{Q}$ (such a solution is dense in general) but instead to exploit the sparse structure of $\mathbf{Q}$. That is, we first collect the non-zero off-diagonal elements of $\mathbf{Q}$ into a $2L \times 2L$ matrix; obtain the square root of the $2L \times 2L$ matrix; and re-embed the square root into the precoder matrix. To do this, let us define a 2×2 matrix

$$\bar{\mathbf{Q}}_{mn} := \begin{bmatrix} [\mathbf{Q}]_{j_m,j_n} & [\mathbf{Q}]_{j_m,i_n} \\ [\mathbf{Q}]_{i_m,j_n} & [\mathbf{Q}]_{i_m,i_n} \end{bmatrix}, \tag{34}$$

for $m, n \in \{1, \ldots, T\}$ and

$$\bar{\mathbf{Q}} := \begin{bmatrix} \bar{\mathbf{Q}}_{11} & \cdots & \bar{\mathbf{Q}}_{1T} \\ \vdots & \ddots & \vdots \\ \bar{\mathbf{Q}}_{T1} & \cdots & \bar{\mathbf{Q}}_{TT} \end{bmatrix}. \tag{35}$$

Then, $\bar{\mathbf{Q}} \in \mathbb{C}^{2T \times 2T}$ contains all the non-zero off-diagonal elements of $\mathbf{Q} \in \mathbb{C}^{N \times N}$ and some of the zero elements of $\mathbf{Q}$. (This positive-semidefinite $\bar{\mathbf{Q}}$ is the same as the matrix $\bar{\mathbf{Q}}$ defined in (24) and (25).) For the example of $\mathbf{Q}$ in (12), $\bar{\mathbf{Q}}$ is given by

$$\bar{\mathbf{Q}} = \begin{bmatrix} q_{1,1} & q_{1,2} & q_{1,5} & q_{1,6} \\ q_{2,1} & q_{2,2} & q_{2,5} & q_{2,6} \\ q_{5,1} & q_{5,2} & q_{5,5} & q_{5,6} \\ q_{6,1} & q_{6,2} & q_{6,5} & q_{6,6} \end{bmatrix}. \tag{36}$$

(Note that in this case, $q_{5,1} = q_{6,2} = q_{1,5} = q_{2,6} = 0$.) Define the square root of $\bar{\mathbf{Q}}$ and its partition as

$$\bar{\mathbf{W}} = [\bar{w}_{i,j}] = \bar{\mathbf{U}}\sqrt{\bar{\mathbf{\Sigma}}} = \begin{bmatrix} \bar{\mathbf{W}}_{11} & \cdots & \bar{\mathbf{W}}_{1T} \\ \vdots & \ddots & \vdots \\ \bar{\mathbf{W}}_{T1} & \cdots & \bar{\mathbf{W}}_{TT} \end{bmatrix}, \tag{37}$$

where $\bar{\mathbf{Q}} = \bar{\mathbf{U}}\bar{\mathbf{\Sigma}}\bar{\mathbf{U}}^{\mathrm{H}}$ is the eigen-decomposition of $\bar{\mathbf{Q}}$ and $\bar{\mathbf{W}}_{mn}$ is a 2×2 matrix for each (m, n). Now, construct the precoder matrix by embedding $\bar{\mathbf{W}}$ into the precoder matrix as

$$\mathbf{W} = \begin{cases} \begin{bmatrix} [\mathbf{W}]_{j_m,j_n} & [\mathbf{W}]_{j_m,i_n} \\ [\mathbf{W}]_{i_m,j_n} & [\mathbf{W}]_{i_m,i_n} \end{bmatrix} = \bar{\mathbf{W}}_{mn} & \text{for } i_m \in \mathcal{I},\ j_n \in \mathcal{J}, \\ [\mathbf{W}]_{i,i} = 1 & \text{for } i \notin \mathcal{I}, \mathcal{J}, \\ [\mathbf{W}]_{i,j} = 0 & \text{otherwise.} \end{cases} \tag{38}$$

It is easy to verify $\mathbf{Q} = \mathbf{W}\mathbf{W}^{\mathrm{H}}$. For the example of $\mathbf{Q}$ in (12), the structure of precoder matrix $\mathbf{W}$ is given by

$$\mathbf{W} = \begin{bmatrix} \bar{w}_{1,1} & \bar{w}_{1,2} & 0 & 0 & \bar{w}_{1,3} & \bar{w}_{1,4} & 0 & 0 \\ \bar{w}_{2,1} & \bar{w}_{2,2} & 0 & 0 & \bar{w}_{2,3} & \bar{w}_{2,4} & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ \bar{w}_{3,1} & \bar{w}_{3,2} & 0 & 0 & \bar{w}_{3,3} & \bar{w}_{3,4} & 0 & 0 \\ \bar{w}_{4,1} & \bar{w}_{4,2} & 0 & 0 & \bar{w}_{4,3} & \bar{w}_{4,4} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}. \tag{39}$$

Note that $\mathbf{W}$ obtained in this way preserves the sparsity of $\mathbf{Q}$ and only a $2T \times 2T$ submatrix perturbs the structure of the identity matrix. Now, recall the received signal model at antenna l:

$$\tilde{\mathbf{y}}_l = \tilde{\mathbf{H}}_{l1}\mathbf{W}\mathbf{x}_1 + \tilde{\mathbf{H}}_{l2}\mathbf{W}\mathbf{x}_2 + \cdots + \tilde{\mathbf{H}}_{lN_t}\mathbf{W}\mathbf{x}_{N_t} + \tilde{\mathbf{n}}_l,$$

for $l = 1, 2, \ldots, N_r$. Due to the structure of our $\mathbf{W}$, we have

$$\tilde{y}_l(i) = \tilde{H}_{l1}(i)x_1(i) + \tilde{H}_{l2}(i)x_2(i) + \cdots + \tilde{H}_{lM}(i)x_M(i) + \tilde{n}_l(i),$$

for each subcarrier $i \notin (\mathcal{I} \cup \mathcal{J})$, since $\tilde{\mathbf{H}}_{lk}$ is diagonal for each l, where $\tilde{H}_{lk}(i)$ and $x_k(i)$ are $[\tilde{\mathbf{H}}_{lk}]_{i,i}$ and the i-th element of $\mathbf{x}_k$, respectively. Collecting all received signals at subcarrier $i \notin (\mathcal{I} \cup \mathcal{J})$, we have a typical $N_r \times N_t$ MIMO system:

$$\begin{bmatrix} \tilde{y}_1(i) \\ \vdots \\ \tilde{y}_{N_r}(i) \end{bmatrix} = \begin{bmatrix} \tilde{H}_{11}(i) & \cdots & \tilde{H}_{1N_t}(i) \\ \vdots & \ddots & \vdots \\ \tilde{H}_{N_r1}(i) & \cdots & \tilde{H}_{N_rN_t}(i) \end{bmatrix} \begin{bmatrix} x_1(i) \\ \vdots \\ x_{N_t}(i) \end{bmatrix} + \begin{bmatrix} \tilde{n}_1(i) \\ \vdots \\ \tilde{n}_{N_t}(i) \end{bmatrix}.$$

Thus, $N - 2T$ (or $N - 2L$ if $T = L$) subcarriers are intact from the blind estimation precoding, enabling MLD per carrier and also allowing per-carrier precoding across antennas for data rate enhancement [17]. Note that $N \gg 2L$ in most OFDM systems.

Next consider subcarriers i's $\in (\mathcal{I} \cup \mathcal{J})$. These subcarriers are mixed by the proposed precoder for blind estimation. For these subcarriers, we have

$$\begin{aligned} \tilde{\mathbf{y}}_l(\mathcal{I} \cup \mathcal{J}) = {} & \tilde{\mathbf{H}}_{l1}(\mathcal{I} \cup \mathcal{J}, \mathcal{I} \cup \mathcal{J})\bar{\mathbf{W}}\mathbf{x}_1(\mathcal{I} \cup \mathcal{J}) + \cdots \\ & + \tilde{\mathbf{H}}_{lN_t}(\mathcal{I} \cup \mathcal{J}, \mathcal{I} \cup \mathcal{J})\bar{\mathbf{W}}\mathbf{x}_{N_t}(\mathcal{I} \cup \mathcal{J}) + \tilde{\mathbf{n}}_l(\mathcal{I} \cup \mathcal{J}), \end{aligned}$$

for each $l = 1, 2, \ldots, N_r$. Collecting the signals at all the receive antennas yields

$$\begin{aligned}\tilde{\mathbf{y}}^{\mathcal{I},\mathcal{J}} &= \begin{bmatrix} \tilde{\mathbf{H}}_{11}^{\mathcal{I},\mathcal{J}} & \cdots & \tilde{\mathbf{H}}_{1N_t}^{\mathcal{I},\mathcal{J}} \\ \vdots & \cdots & \vdots \\ \tilde{\mathbf{H}}_{N_r1}^{\mathcal{I},\mathcal{J}} & \cdots & \tilde{\mathbf{H}}_{N_rN_t}^{\mathcal{I},\mathcal{J}} \end{bmatrix} \begin{bmatrix} \bar{\mathbf{W}} & \cdots & \mathbf{0} \\ \vdots & \ddots & \vdots \\ \mathbf{0} & \cdots & \bar{\mathbf{W}} \end{bmatrix} \begin{bmatrix} \mathbf{x}_1^{\mathcal{I},\mathcal{J}} \\ \vdots \\ \mathbf{x}_{N_t}^{\mathcal{I},\mathcal{J}} \end{bmatrix} \\ &\quad + \begin{bmatrix} \tilde{\mathbf{n}}_1^{\mathcal{I},\mathcal{J}} \\ \vdots \\ \tilde{\mathbf{n}}_{N_t}^{\mathcal{I},\mathcal{J}} \end{bmatrix} \\ &= \mathcal{H}\mathrm{diag}(\bar{\mathbf{W}}, \ldots, \bar{\mathbf{W}})\mathbf{x}^{\mathcal{I},\mathcal{J}} + \tilde{\mathbf{n}}^{\mathcal{I},\mathcal{J}}, \end{aligned} \tag{40}$$

where $\tilde{\mathbf{y}}^{\mathcal{I},\mathcal{J}} = [\tilde{\mathbf{y}}_1(\mathcal{I} \cup \mathcal{J})^{\mathrm{T}}, \ldots, \tilde{\mathbf{y}}_{N_r}(\mathcal{I} \cup \mathcal{J})^{\mathrm{T}}]^{\mathrm{T}}$, $\tilde{\mathbf{H}}_{lk}^{\mathcal{I},\mathcal{J}} = \tilde{\mathbf{H}}_{lk}(\mathcal{I} \cup \mathcal{J}, \mathcal{I} \cup \mathcal{J})$, $\mathbf{x}_k^{\mathcal{I},\mathcal{J}} = \mathbf{x}_k, (\mathcal{I} \cup \mathcal{J})$, $\tilde{\mathbf{n}}_l^{\mathcal{I},\mathcal{J}} = \tilde{\mathbf{n}}_l(\mathcal{I} \cup \mathcal{J})$, and the definitions of $\mathcal{H}$, $\mathbf{x}^{\mathcal{I},\mathcal{J}}$ and $\tilde{\mathbf{n}}^{\mathcal{I},\mathcal{J}}$ are clear from the equation. Since these subcarriers are mixed by the proposed blind precoder, we need joint processing to decode $\mathbf{x}^{\mathcal{I},\mathcal{J}}$. Note that the proposed blind precoder yields a much smaller system for joint symbol estimation than the existing fully dense precoders [9], [11], [14], [15]. One can apply one of several MIMO encoding and decoding schemes to these subcarriers such as MMSE filtering via the state-space inversion technique [26].

A. Sign Determination of the Square Precoder Matrix

From Theorem 1, we know that the equal absolute value design for the elements of $\mathbf{Q}(\mathcal{I}, \mathcal{J})$ is optimal and the signs of the elements of $\mathbf{Q}(\mathcal{I}, \mathcal{J})$ are irrelevant for the channel estimation performance of the proposed blind scheme since the channel estimation performance depends on $\mathbf{Q}(\mathcal{I}, \mathcal{J})$ only through the absolute values of the elements of $\mathbf{Q}(\mathcal{I}, \mathcal{J})$. However, the determination of the remaining signs for the elements of $\mathbf{Q}(\mathcal{I}, \mathcal{J})$ affects the decoding performance. As seen in the above, the proposed precoding scheme does not affect the decoding of the subcarriers $\{1, 2, \ldots, N\} \setminus (\mathcal{I} \cup \mathcal{J})$, but affects that of the subcarriers $\mathcal{I} \cup \mathcal{J}$, as seen in (40). For the good decoding performance, we should minimize the condition number of $\bar{\mathbf{W}}$ to minimize the noise enhancement. Since the condition number of $\bar{\mathbf{W}}$ is the square root of that of $\bar{\mathbf{Q}}$ from (37), we should design the signs of the elements of $\mathbf{Q}(\mathcal{I}, \mathcal{J})$ to minimize the condition number of $\bar{\mathbf{Q}}$. To this end, first note that we can decompose $\bar{\mathbf{Q}}$ as $\bar{\mathbf{Q}} = [\bar{q}_{i,j}] = \mathbf{I}_{2T} + \bar{\mathbf{Q}}'$ since the diagonal elements of $\mathbf{Q}$ are one in our design. An example of $T = 4$ is shown in (41):

$$\bar{\mathbf{Q}}' = \begin{bmatrix} 0 & \times & 0 & \times & 0 & \times & 0 & \times \\ q_{i_1,j_1} & 0 & q_{i_1,j_2} & 0 & q_{i_1,j_3} & 0 & q_{i_1,j_4} & 0 \\ 0 & \times & 0 & \times & 0 & \times & 0 & \times \\ q_{i_2,j_1} & 0 & q_{i_2,j_2} & 0 & q_{i_2,j_3} & 0 & q_{i_2,j_4} & 0 \\ 0 & \times & 0 & \times & 0 & \times & 0 & \times \\ q_{i_3,j_1} & 0 & q_{i_3,j_2} & 0 & q_{i_3,j_3} & 0 & q_{i_3,j_4} & 0 \\ 0 & \times & 0 & \times & 0 & \times & 0 & \times \\ q_{i_4,j_1} & 0 & q_{i_4,j_2} & 0 & q_{i_4,j_3} & 0 & q_{i_4,j_4} & 0 \end{bmatrix}, \tag{41}$$

where the positions marked $\times$ are for $\mathbf{Q}(\mathcal{J}, \mathcal{I}) (= \mathbf{Q}(\mathcal{I}, \mathcal{J})^{\mathrm{H}})$. $\mathbf{Q}(\mathcal{I}, \mathcal{J})$ and $\mathbf{Q}(\mathcal{J}, \mathcal{I})$ are interlaced in $\bar{\mathbf{Q}}'$ with $\mathcal{I} \cap \mathcal{J} = \emptyset$. Then, an eigenvalue of $\bar{\mathbf{Q}}$ is the sum of one and an eigenvalue of $\bar{\mathbf{Q}}'$. Since the trace of $\bar{\mathbf{Q}}'$ is zero, the sum of the eigenvalues of $\bar{\mathbf{Q}}'$ is zero and thus some of the eigenvalues of $\bar{\mathbf{Q}}'$ are positive and others are negative. Hence, the spectral gap of $\bar{\mathbf{Q}}'$ should be made small to make the condition number of $\bar{\mathbf{Q}}$ small. Let $-1 \leq \lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_{2T}$ be the eigenvalues of $\bar{\mathbf{Q}}'$. Then, the desired optimization problem can be formulated as determining the $2T$ eigenvalues $\lambda_1, \ldots, \lambda_{2T}$ to minimize the condition number $\chi(\bar{\mathbf{Q}}) = (1 + \lambda_{2T})/(1 + \lambda_1)$ of $\bar{\mathbf{Q}}$ under proper constraints, given by

$$\begin{aligned} &\min_{\{\lambda_1, \ldots, \lambda_{2T}\}} \frac{1 + \lambda_{2T}}{1 + \lambda_1} \\ &\quad \text{s.t.} \quad \mathrm{tr}(\bar{\mathbf{Q}}') = \sum_i \lambda_i = 0, \ \mathrm{tr}(\bar{\mathbf{Q}}'\bar{\mathbf{Q}}'^{\mathrm{H}}) = \sum_i \lambda_i^2 = 2\rho^2, \\ &\quad \text{and} \ -1 \leq \lambda_1 \leq \cdots \leq \lambda_{2T}, \end{aligned} \tag{42}$$

where $\mathrm{tr}(\bar{\mathbf{Q}}'\bar{\mathbf{Q}}'^{\mathrm{H}}) = 2\sum_{\mathcal{I},\mathcal{J}} |q_{i_s,j_t}|^2 = 2\rho^2$ under the proposed design $|q_{i_s,j_t}| = \rho/T$.

Theorem 3: Under the assumption that all $|[\mathbf{Q}]_{i_s,j_t}|$, $i_s \in \mathcal{I}$, $j_t \in \mathcal{J}$, are equal, the minimum condition number of $\mathbf{Q}$ is achieved when $\mathbf{Q}(\mathcal{I}, \mathcal{J})$ is a scaled version of a unitary matrix.

Proof: From (25), $\bar{\mathbf{Q}}'$ can be decomposed as

$$\mathbf{\Pi}_2\bar{\mathbf{Q}}'\mathbf{\Pi}_2^{\mathrm{T}} = \begin{bmatrix} \mathbf{0} & \bar{\bar{\mathbf{Q}}}^{\mathrm{H}} \\ \bar{\bar{\mathbf{Q}}} & \mathbf{0} \end{bmatrix}. \tag{43}$$

Thus, the eigenvalues of $\bar{\mathbf{Q}}'$ are given by $\{\pm\sigma_t, t = 1, \ldots, T\}$, where σ_t is a singular value of $\bar{\bar{\mathbf{Q}}}$. Then, the optimal design problem (42) can be rewritten as

$$\min_{\{\sigma_t\}} \frac{1 + \sigma_1}{1 - \sigma_1}, \ \text{s.t.} \ 1 > \sigma_1 \geq \cdots \geq \sigma_T > 0 \ \text{and} \ \sum_{t=1}^{T} \sigma_t^2 = \rho^2. \tag{44}$$

We have $1 > \sigma_1 \geq \rho/\sqrt{T}$ from the squared sum constraint in (44). Note that the objective function in (44) monotonically increases as σ_1 increases within the feasible range $[\rho/\sqrt{T}, 1)$. Hence, the minimum condition number occurs when $\sigma_1 = \rho/\sqrt{T}$. By (44), we then have $\sigma_1 = \cdots = \sigma_T = \rho/\sqrt{T}$ and thus $\bar{\bar{\mathbf{Q}}}$ is a scaled version of a unitary matrix. Finally, $\mathbf{Q}(\mathcal{I}, \mathcal{J}) = \bar{\bar{\mathbf{Q}}}$. (See (26).) ■

Corollary 2: With the optimal solution, $\sigma_1 = \cdots = \sigma_T = \rho/\sqrt{T}$ with $\rho \in (0, \sqrt{T})$, the condition numbers of $\bar{\mathbf{Q}}$ (or $\mathbf{Q}$) and $\bar{\mathbf{W}}$ (or $\mathbf{W}$) are given by

$$\chi(\bar{\mathbf{Q}}) = \frac{\sqrt{T} + \rho}{\sqrt{T} - \rho} \quad \text{and} \quad \chi(\bar{\mathbf{W}}) = \sqrt{\frac{\sqrt{T} + \rho}{\sqrt{T} - \rho}}. \tag{45}$$

Remark 1: Corollaries 1 and 2 describe the trade-off between the blind channel estimation and the data decoding. Note that ρ up to $\sqrt{T}$ is valid for the positive semi-definiteness of $\bar{\mathbf{Q}}$. The channel estimation performance can be improved by increasing ρ up to $\sqrt{T}$, as seen in (29), but the condition number deteriorates, as seen in (45). Thus, the value of $\rho \in (0, \sqrt{T})$ should be properly chosen for optimal performance. Note that in our scheme, $N - 2L$ subcarriers are not even affected by the blind precoding.

TABLE I

THE EIGENVALUES AND THE CONDITION NUMBERS WHEN $N = 64$ AND $T = 4$: THE NUMBERS IN PARENTHESIS ARE EIGENVALUE MULTIPLICITY. (THE NOTATIONS IN THE FIRST COLUMN FOLLOW THE REFERENCES.)

Square precoder matrix $\mathbf{Q}$	The off-diagonal value	Eigenvalues				Condition Number of $\mathbf{Q}$	Condition Number of $\mathbf{W}$
Zhang [12], Gao and Nallanathan [15] $\left(-\frac{1}{N-1} < p < 1\right)$	$p = 0.2$	0.8 (63)	13.6 (1)			17	4.12
	$p = 0.5$	0.5 (63)	32.5 (1)			65	8.06
	$p = 0.8$	0.2 (63)	51.4 (1)			257	16.03
Shin et al. [16] $(0 < \alpha < 1, \delta = 0.05)$	$\alpha = 0.8$	0.75 (31)	0.80 (1)	0.85 (31)	14.4 (1)	19.20	4.38
	$\alpha = 0.5$	0.45 (31)	0.50 (1)	0.55 (31)	34.5 (1)	76.67	8.76
	$\alpha = 0.2$	0.15 (31)	0.2 (1)	0.25 (31)	54.6 (1)	364	19.08
Proposed Design $(0 < \rho < \sqrt{T})$	$\rho = 0.4$	0.8 (4)	1 (56)	1.2 (4)		1.5	1.22
	$\rho = 1$	0.5 (4)	1 (56)	1.5 (4)		3	1.73
	$\rho = 1.6$	0.2 (4)	1 (56)	1.8 (4)		9	3

Theorem 3 states that any properly scaled unitary matrix with elements of an equal absolute value is optimal from the perspective of the condition number and noise enhancement. Well known such matrices include the FFT matrix and the Hadamard matrix, when $T = 2^m$ for some integer m. Especially, the Hadamard matrix $\mathbf{H}_T$ of size T is advantageous since it contains only ± 1 and thus the operation of obtaining $\mathbf{J}_{\mathcal{I},\mathcal{J}}$ from $\mathbf{R}_l$ by using (10) and (11) is simple sign change if necessary. The Hadamard matrix $\mathbf{H}_T$ for $T = 2^m$ is constructed as (for $m \geq 2$)

$$\mathbf{H}_2 = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \text{ and } \mathbf{H}_{2^m} = \begin{bmatrix} \mathbf{H}_{2^{m-1}} & \mathbf{H}_{2^{m-1}} \\ \mathbf{H}_{2^{m-1}} & -\mathbf{H}_{2^{m-1}} \end{bmatrix}.$$

With the design of $\mathbf{Q}(\mathcal{I},\mathcal{J}) = \rho/T\mathbf{H}_T$, $\bar{\mathbf{Q}}$ is given by $\bar{\mathbf{Q}} = \mathbf{I} + (\rho/T)\mathbf{\Pi}(\mathbf{H}_T \otimes \mathbf{I}_2)$, where $\mathbf{\Pi}$ is a permutation matrix exchanging rows $2i-1$ and $2i$ for each $i = 1, 2, \ldots, T$ and $\mathbf{\Pi} = \mathbf{\Pi}^{\mathrm{T}}$. Note that both matrices $1/\sqrt{T}(\mathbf{H}_T \otimes \mathbf{I}_2)$ and $(1/\sqrt{T})\mathbf{\Pi}(\mathbf{H}_T \otimes \mathbf{I}_2)$ are orthonormal and symmetric. Since all the eigenvalues of an orthonormal matrix have modulus one [27], the eigenvalues of $1/\sqrt{T}\mathbf{\Pi}(\mathbf{H}_T \otimes \mathbf{I}_2)$ are ± 1 due to its symmetry. Further, since the trace of $1/\sqrt{T}\mathbf{\Pi}(\mathbf{H}_T \otimes \mathbf{I}_2)$ is zero and the trace is the sum of eigenvalues, there are exactly T ones and T negative ones. Therefore, $\rho/T\mathbf{\Pi}(\mathbf{H}_T \otimes \mathbf{I}_2)$ has T eigenvalues of the value $\rho/\sqrt{T}$ and T eigenvalues of the value $-\rho/\sqrt{T}$, which coincides with the requirement of the optimal precoder derived in Theorem 3. Now, let us compare the condition numbers of several available precoders for the precoding-based blind channel estimation for OFDM systems [8], [9], [11], [13]. Table I shows the eigenvalue and condition number for the square precoder matrix $\mathbf{Q}$ for each design. It is seen that the proposed precoder design is superior to the previous designs. As seen in the table, for our design, there is no effective noise enhancement in the inversion process for data decoding for the subcarriers affected by the blind estimation precoding.

Since all our design is complete now, we finish this section by summarizing the proposed algorithm in a practical setting. We consider the practical case in which $T = L = 2^m$ for some positive integer m, $\mathcal{I}$, and $\mathcal{J}$ are equi-spaced as (17) and $\mathbf{Q}(\mathcal{I},\mathcal{J}) = (\rho/T)\mathbf{H}_T$. In this case, the proposed algorithm can be implemented very efficiently as follows.

Encoding

(1) The $\mathcal{I} \cup \mathcal{J}$ subcarrier signals are multiplied by the precoder matrix $\bar{\mathbf{W}}$.

(2) The precoded $\mathcal{I} \cup \mathcal{J}$ subcarriers are multiplexed into proper positions in the subcarrier domain, and then the signal is OFDM modulated and transmitted.

Decoding

(1) Obtain the elements at $\mathcal{I} \times \mathcal{J}$ from the sample covariance matrix $\hat{\mathbf{R}}_l$, i.e., compute $\hat{\mathbf{R}}_l(\mathcal{I},\mathcal{J})$.

(2) Obtain $\mathbf{\Psi} := \hat{\mathbf{R}}_l(\mathcal{I},\mathcal{J}) \oslash (\rho/T)\mathbf{H}_T$ where $\mathbf{H}_T$ is the Hadamard matrix of size T.

(3) Inverse fast Fourier transform (FFT) $\mathbf{\Psi}$ from the left and then FFT the resulting matrix from the right to obtain $\hat{\mathbf{J}}_{\mathcal{I},\mathcal{J}}$. (This is because $\tilde{\mathbf{F}}(\mathcal{I},:)$ and $\tilde{\mathbf{F}}(\mathcal{J},:)$ are FFT matrices of size T with some scaling for equi-spaced $\mathcal{I}$ and $\mathcal{J}$.)

(4) Obtain the N_t most significant eigenvalues and eigenvectors of $\hat{\mathbf{J}}_{\mathcal{I},\mathcal{J}}$ obtained at step (3) by using an efficient algorithm such as the power iteration method.

(5) Identify the remaining rotational ambiguity.[3]

Note that the encoding and decoding for blind channel estimation are implemented by well-known low-cost operations such as sign change, FFT and the power iteration method. Furthermore, the computational complexity for precoding-based blind channel estimation reduces from $\mathcal{O}(N^2)$ to $\mathcal{O}(T^2)$ by the proposed method, where $|\mathcal{I}| = |\mathcal{J}| = T$ and $T \ll N$. In data decoding, the proposed method leaves $\{1, 2, \ldots, N\} \setminus (\mathcal{I} \cup \mathcal{J})$ subcarriers untouched and the conventional data rate enhancing MIMO precoding can be applied to these subcarriers. Thus, only $2L$ subcarriers require joint symbol estimation at the receiver as in (40). The proposed precoder design can be generalized to the index sets $\mathcal{I}$ and $\mathcal{J}$ of different cardinality (i.e., $|\mathcal{I}| \neq |\mathcal{J}|$), but the design with the index sets with the same cardinality performs better in terms of both channel estimation

[3]An $N_t \times N_t$ unitary ambiguity matrix in practical systems can be resolved by using a minimum $N_p = N_t^2$ pilot signals combined with a least square estimator [22].

performance and data symbol decoding performance. The proof of this is omitted and can be shown based on the proofs of Theorems 1 and 3.

B. Multiple-User Case

In this subsection, we extend our precoder design for blind channel estimation to multiple user case. We consider the case that a single receiver receives signals from multiple transmitters. Such a case occurs when a terminal station is in the cell boundary and receives signals from interfering basestations as well as from the desired basestation or simply when a basestation receives signals from multiple terminals. Although the proposed design does not solve the interference problem fully, the proposed method can separate the statistics for the blind channel estimation entirely for different users. This separation results from our sparse precoder design. Suppose that different users use different sparse precoder square matrices, i.e., for each user $\mathcal{I}$ is disjoint with $\mathcal{I}$ of any other user and so is $\mathcal{J}$. Such a construction is possible since only L subcarriers are required for each of $\mathcal{I}$ and $\mathcal{J}$. There are N subcarriers in total; $N/T-1$ such constructions are available. Under the assumption that signals from different users are uncorrelated, we have the following data model in the K user case:

$$\mathbf{R}_l = \left(\sum_k \tilde{\mathbf{h}}_{lk}^{(1)}\tilde{\mathbf{h}}_{lk}^{(1)\mathrm{H}}\right) \odot \mathbf{Q}^{(1)} + \cdots + \left(\sum_k \tilde{\mathbf{h}}_{lk}^{(K)}\tilde{\mathbf{h}}_{lk}^{(K)\mathrm{H}}\right) \odot \mathbf{Q}^{(K)} + \sigma_n^2\mathbf{I}, \quad (46)$$

where $\{\tilde{\mathbf{h}}_{lk}^{(j)}\}$ is the channel from antenna k to l of user j in frequency domain, and $\mathbf{Q}^{(j)}$ is the square precoder matrix for user j. When $\{\mathcal{I},\mathcal{J}\}$ is disjoint to each other for different users, the statistic for blind channel estimation is completely separated. Of course, there remains some interference from other users if the estimated sample covariance matrix is used instead of the true covariance matrix. However, the interference can be suppressed in case of slow fading since the sample size N_s for obtaining the sample covariance matrix can be large.

VI. Numerical Results

Simulations were performed to evaluate the proposed precoder design and the corresponding algorithm. First, we validated the analytic results in the previous sections by considering a SISO-OFDM system with 64 subcarriers ($N=64$), 4-tap (time-domain) channel ($L=4$), and $\mathbf{h}\sim\mathcal{CN}(\mathbf{0},\mathbf{I}_L)$. Since the effect of symbol constellation is negligible [11], i.i.d. binary phase-shift keying (BPSK) symbols were used. The channel estimation performance was measured by the normalized mean square error (NMSE) defined as NMSE $= (1/N_{MC})\sum_{n=1}^{N_{MC}}(\|\hat{\mathbf{h}}_n-\mathbf{h}\|_F^2/\|\mathbf{h}\|_F^2)$, where $\hat{\mathbf{h}}_n$ denotes the n-th run estimate of $\mathbf{h}$ and N_{MC} is the number of total Monte Carlo runs ($N_{MC}=1,000$). We used $N_s=300$ symbols to construct the sample data covariance matrix.

Throughout the simulation, for the proposed method and other considered methods, we do not adopt the impractical assumption of perfectly known rotational ambiguity at the receiver. To resolve the rotational ambiguity matrix, we used the minimum number of pilot tones with 10 dB power boosting (only $N_p=N_t^2$ pilot tones for the entire N_s symbol period) and applied the least squares method in Proposition 2 of [22]. Fig. 2 shows the NMSE of the proposed design versus SNR, averaged over random channel realizations for different values of T and ρ: Fig. 2(a) for an equal power delay profile given by $\mathbf{h}\sim\mathcal{CN}(\mathbf{0},\sigma_h^2\mathbf{I})$ and Fig. 2(b) for an exponential power delay profile given by $\sigma_{h,l}^2=\mathbb{E}\{|h(l)|^2\}=e^{-2l/5}$ for $0\le l<L$. Here, the SNR is defined as $\mathbb{E}\|\mathbf{h}\|^2/N\sigma_n^2$. As predicted by Theorem 1 and Corollary 1, the estimation performance improves as ρ increases, and the performance difference caused by different T is negligible even in Fig. 2(b). Also, it is seen that there exists a performance floor at high SNR for given N_s, as predicted by Corollary 1. However, the performance floor for $\rho\ge1$ is below -20 dB which is already sufficient for coherent decoding. We investigated the behavior of the estimation performance floor at high SNR (e.g., SNR $=40$ dB) as the block size N_s increases. Fig. 3 shows the corresponding result. As expected from Corollary 1, the performance improves as N_s increases according to $\Theta(1/N_s)$. We performed the

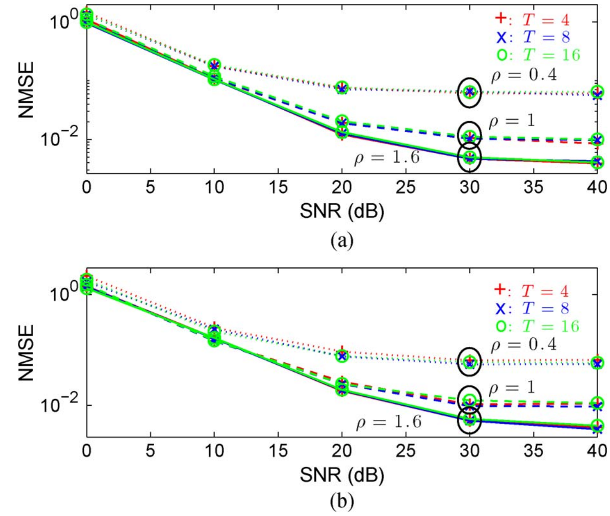


Fig. 2. NMSE versus SNR for different T and ρ in a SISO case where $N_s=300$: (a) equal power delay profile and (b) exponential power delay profile.

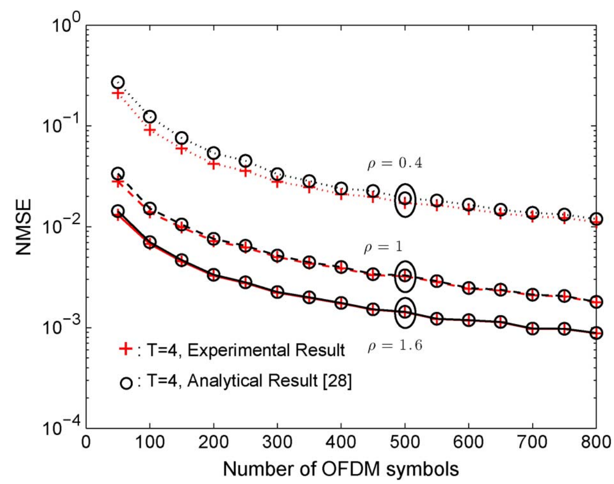
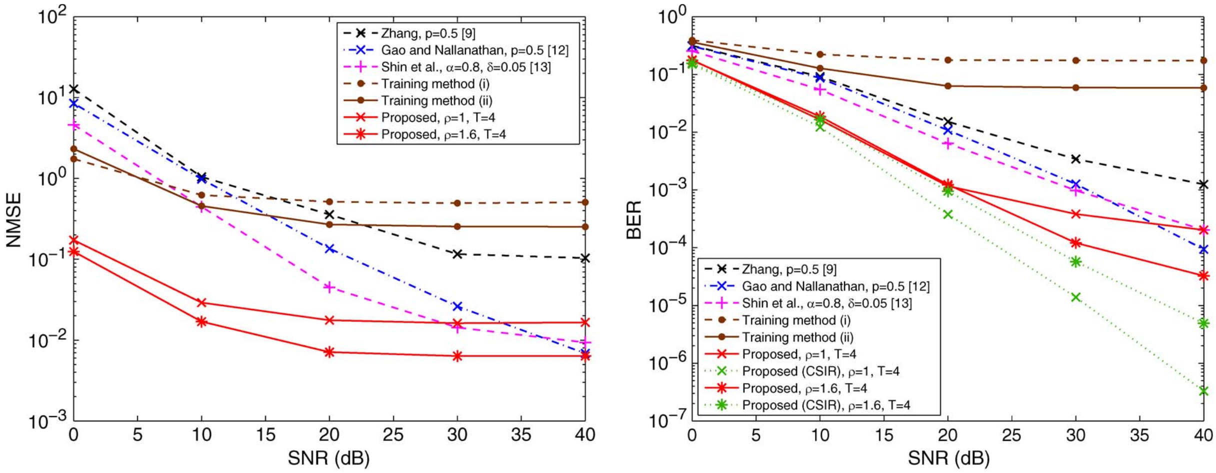


Fig. 3. NMSE versus the block size N_s in the SISO case where SNR 40 dB (assuming known phase ambiguity).[4]



(a) (b)

Fig. 4. Comparison with the previous designs in a 2 × 2 MIMO: $N_s = 500$ and $\sigma^2_{h,l} = \mathbb{E}\{|h(l)|^2\} = e^{-l/5}$ for $0 \leq l < L$. (a) Channel estimation performance. (b) Data decoding performance.

perturbation analysis in [28] and our simulation result matches the analytic result very well.[4]

With the verification of the analytic results, we then investigated the performance of the proposed precoder in the MIMO case of $N_t = N_r = 2$. In the MIMO case, we used $N_s = 500$ and the channel model with an exponential power delay profile given by $\sigma^2_{h,l} = \mathbb{E}\{|h(l)|^2\} = e^{-l/5}$ for $0 \leq l < L$. We first compared the performance of the proposed method to those of several existing non-redundant precoding-based blind methods [9], [11], [13] capable of MIMO channel estimation. Fig. 4(a) shows the channel estimation performance. Because of our equi-spaced sampling, $\tilde{\mathbf{F}}(\mathcal{I},:)$ and $\tilde{\mathbf{F}}(\mathcal{J},:)$ in (11) has the condition number of one and thus there is no noise enhancement during the construction of $\mathbf{J}_{\mathcal{I},\mathcal{J}}$. Hence, the proposed algorithm yields a good performance at the low SNR range. However, it saturates quickly than previous methods that use all subcarriers (i.e., more samples for one symbol time), as expected. In Fig. 4, we also plotted the performance of a training-based least square (LS) channel estimator that uses (i) $N_p = N_t^2$ and (ii) $N_p = (N_t + 1)N_t$ pilot symbols for channel estimation. (Note that all the considered blind channel estimation methods used $N_p = N_t^2$ pilot symbols for resolving the ambiguity matrix for the entire N_s symbol block.) Indeed, with the same amount of pilot tones, the training-based method is not comparable to the blind methods. Fig. 4(b) shows the BER performance of the considered designs for a set of randomly generated channel vectors. Here, since the proposed precoder does not affect $N - 2L$ subcarriers at all, we used MLD approach to detect the two transmitted bits for each of the unaffected subcarriers for the proposed design. (This is the benefit of the proposed design.) For the remaining $2L$ subcarrier channels, we used the MMSE inversion to the effective channel based on the true noise variance. For all other methods that linearly precode all $N = 64$ subcarriers, we constructed a $2N$ dimension data model to apply linear MMSE filtering for data

[4]The analytical result in [28] gives expressions for the errors incurred in the data covariance estimation due to the finite sample size without considering the error in the phase ambiguity estimation. For the purpose of comparison, the phase information is assume to be known perfectly for the proposed estimation method in Fig. 3.

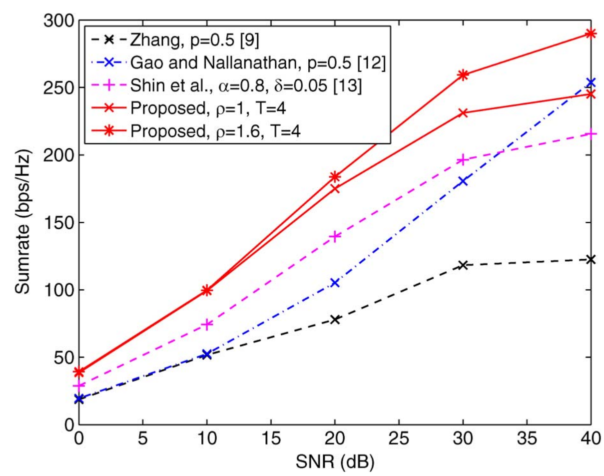


Fig. 5. Sum-rate comparison in a 2 × 2 MIMO where $N_s = 500$.

symbol detection. Note that for this big system the overall MLD approach is computationally impossible. It is shown that the proposed precoder is superior to the other precoders in the BER performance. This is because for the proposed precoder, $N - 2L$ subcarriers are not even affected and the mixed $2L$ subcarrier channels are unmixed by the inverse of the well-conditioned precoder matrix.

Finally, we investigated the sum-rate performance of the proposed precoder in the 2 × 2 MIMO cases, and the result is shown in Fig. 5. Here, the sum rate computation is based on the typical training-based rate computation method, explained as follows. For simplicity, in the SISO case, we have the data model $\tilde{\mathbf{y}} = \mathbf{H}\tilde{\mathbf{W}}\mathbf{x} + \tilde{\mathbf{n}}$ from (2). Since $\tilde{\mathbf{H}} = \hat{\tilde{\mathbf{H}}} + \Delta\tilde{\mathbf{H}}$ with the estimated channel, we have a new data model $\tilde{\mathbf{y}} = \hat{\tilde{\mathbf{H}}}\mathbf{W}\mathbf{x} + (\tilde{\mathbf{n}} + \Delta\mathbf{H}\tilde{\mathbf{W}}\mathbf{x})$. However, for the previous dense precoded designs, achieving the capacity of this new channel is not realizable in practice. Thus, we used a further processed practical model $\mathcal{F}_{\hat{\tilde{\mathbf{H}}}\mathbf{W}}\tilde{\mathbf{y}} = \mathbf{x} + \mathcal{F}_{\hat{\tilde{\mathbf{H}}}\mathbf{W}}(\tilde{\mathbf{n}} + \Delta\mathbf{H}\tilde{\mathbf{W}}\mathbf{x})$, where $\mathcal{F}_{\hat{\tilde{\mathbf{H}}}\mathbf{W}}$ is the MMSE filter based on $\hat{\tilde{\mathbf{H}}}\mathbf{W}$. Then, we assumed an independent stream with equal power for each subcarrier and computed the rate for each subcarrier based on the simulation result on the MSE of $\Delta\tilde{\mathbf{H}}$. For the MIMO case, we simply extended the SISO case accordingly. For each of the other methods, the design parameter was swept over a feasible range, and the best one among the swept values was chosen. It is seen that the proposed precoder is also superior to the existing designs relative to sum rate performance, as well as BER.

VII. Conclusion

We have considered the precoder design for the precoding-based blind channel estimation for MIMO-OFDM systems. We have proposed a new precoder based on a sparse structure. The main advantage of the proposed precoder is that it is compatible with the data rate enhancing MIMO precoding by using minimal subcarriers to introduce signal correlation necessary for blind channel estimation and leaving most subcarriers unaffected. The proposed precoder is well conditioned and the corresponding blind estimation algorithm can be implemented very efficiently with low cost operations such as sign change and FFT. The proposed precoder design can be generalized by considering the case of partial channel state information at the transmitter via limited feedback and this scenario is practically useful. In addition, the pilot signal required for resolving the ambiguity matrix can be minimized effectively or jointly used for more accurate semi-blind channel estimation [29]–[31].

Appendix

Lemma 1:

$$\mathbb{E}\left\{\left[\Delta\mathbf{v}'_{j_1}\cdots\Delta\mathbf{v}'_{j_T}\right]^{\mathrm{H}}\left(\tilde{\mathbf{F}}(\mathcal{I},:)\tilde{\mathbf{F}}(\mathcal{I},:)^{\mathrm{H}}\right)^{\dagger}\left[\Delta\mathbf{v}'_{j_1}\cdots\Delta\mathbf{v}'_{j_T}\right]\right\} = \frac{NL}{N_sT^2}\left(\frac{LN_t}{N}\sigma_h^2+\sigma_n^2\right)^2\begin{bmatrix}\sum_{t=1}^{T}\frac{1}{\left|[\mathbf{Q}]_{i_t,j_1}\right|^2} & \cdots & 0\\ \vdots & \ddots & \vdots\\ 0 & \cdots & \sum_{t=1}^{T}\frac{1}{\left|[\mathbf{Q}]_{i_t,j_T}\right|^2}\end{bmatrix}.$$

Proof of Lemma 1: Define $\mathbf{C}:=[\Delta\mathbf{v}'_{j_1}\cdots\Delta\mathbf{v}'_{j_T}]^{\mathrm{H}}\mathbf{G}_{\mathcal{I}}^{\dagger}[\Delta\mathbf{v}'_{j_1}\cdots\Delta\mathbf{v}'_{j_T}]$, where $\mathbf{G}_{\mathcal{I}}=\tilde{\mathbf{F}}(\mathcal{I},:)\tilde{\mathbf{F}}(\mathcal{I},:)^{\mathrm{H}}$. Then, the diagonal elements of $\mathbf{C}$ are given by (see equation at the bottom of the page). The following is known from [32]:

$$\mathbb{E}_n\left\{\left|[\Delta\mathbf{R}_l]_{i_t,j_m}\right|^2|\mathbf{h}_{lk}\right\}=\frac{1}{N_s}[\mathbf{R}_l]_{i_t,i_t}[\mathbf{R}_l]_{j_m,j_m}, \quad (50)$$

$$\mathbb{E}_n\left\{[\Delta\mathbf{R}_l]_{i_t,j_m}[\Delta\mathbf{R}_l]^*_{i_u,j_n}|\mathbf{h}_{lk}\right\}=\frac{1}{N_s}[\mathbf{R}_l]_{i_t,i_u}[\mathbf{R}_l]_{j_n,j_m}, \quad (51)$$

where $\mathbb{E}_n\{\cdot|\mathbf{h}_{lk}\}$ is the expectation over the noise distribution for given channel realization. Remember $\mathbf{R}_l:=\mathbb{E}_n\{\tilde{\mathbf{y}}_l\tilde{\mathbf{y}}_l^{\mathrm{H}}|\mathbf{h}_{lk}\}=(\sum_k\tilde{\mathbf{h}}_{lk}\tilde{\mathbf{h}}_{lk}^{\mathrm{H}})\odot\mathbf{Q}+\sigma_n^2\mathbf{I}$. So, any $[\mathbf{R}_l]_{i_t,i_u}[\mathbf{R}_l]_{j_n,j_m}=0$ for $t\neq u$ or $m\neq n$ because of the proposed sparse structure of $\mathbf{Q}$, i.e., $[\mathbf{Q}]_{i,j}=0$ for $(i,j)\notin\mathcal{I}\times\mathcal{J}\cup\mathcal{J}\times\mathcal{I}$ and $\mathcal{I}\cap\mathcal{J}=\emptyset$. Thus, the term in (51) is zero and the second terms in (47)–(49) are all zero. Also, we have

$$[\mathbf{R}_l]_{i_t,i_t}=\left(\sum_{k=1}^{N_t}\left|\tilde{h}_{lk}(i_t)\right|^2\right)[\mathbf{Q}]_{i_t,i_t}+\sigma_n^2 \text{ and}$$
$$[\mathbf{R}_l]_{j_m,j_m}=\left(\sum_{k=1}^{N_t}\left|\tilde{h}_{lk}(j_m)\right|^2\right)[\mathbf{Q}]_{j_m,j_m}+\sigma_n^2, \quad (52)$$

with $[\mathbf{Q}]_{i,i}=1$ for all $i\in\{1,\ldots,N\}$ by Constraint *(C.2)*. Now taking expectation of (50) over channel distribution yields

$$\frac{1}{N_s}\mathbb{E}_h\left\{[\mathbf{R}_l]_{i_t,i_t}[\mathbf{R}_l]_{j_m,j_m}\right\}=\frac{1}{N_s}\left(\frac{LN_t}{N}\sigma_h^2+\sigma_n^2\right)^2,$$

where the fact $\tilde{\mathbf{h}}_{lk}=\tilde{\mathbf{F}}\mathbf{h}_{lk}$ and the assumption of the zero-mean proper complex Gaussian distribution of $\mathbf{h}_{lk}$ are used. Therefore, the expectation of the diagonal elements of $\mathbf{C}$ over

$$\begin{aligned}
[\mathbf{C}]_{m,m}
&=\left(\sum_{t=1}^{T}\left[\mathbf{G}_{\mathcal{I}}^{\dagger}\right]_{t,1}\frac{[\Delta\mathbf{R}_l]^*_{i_t,j_m}}{[\mathbf{Q}]^*_{i_t,j_m}}\right)\frac{[\Delta\mathbf{R}_l]_{i_1,j_m}}{[\mathbf{Q}]_{i_1,j_m}}\\
&\quad+\left(\sum_{t=1}^{T}\left[\mathbf{G}_{\mathcal{I}}^{\dagger}\right]_{t,2}\frac{[\Delta\mathbf{R}_l]^*_{i_t,j_m}}{[\mathbf{Q}]^*_{i_t,j_m}}\right)\frac{[\Delta\mathbf{R}_l]_{i_2,j_m}}{[\mathbf{Q}]_{i_2,j_m}}+\cdots\\
&\quad+\left(\sum_{t=1}^{T}\left[\mathbf{G}_{\mathcal{I}}^{\dagger}\right]_{t,T}\frac{[\Delta\mathbf{R}_l]^*_{i_t,j_m}}{[\mathbf{Q}]^*_{i_t,j_m}}\right)\frac{[\Delta\mathbf{R}_l]_{i_T,j_m}}{[\mathbf{Q}]_{i_T,j_m}}\\
&=\left[\mathbf{G}_{\mathcal{I}}^{\dagger}\right]_{1,1}\frac{\left|[\Delta\mathbf{R}_l]_{i_1,j_m}\right|^2}{\left|[\mathbf{Q}]_{i_1,j_m}\right|^2}+\left(\sum_{t\neq1}^{T}\left[\mathbf{G}_{\mathcal{I}}^{\dagger}\right]_{t,1}\frac{[\Delta\mathbf{R}_l]^*_{i_t,j_m}}{[\mathbf{Q}]^*_{i_t,j_m}}\right)\frac{[\Delta\mathbf{R}_l]_{i_1,j_m}}{[\mathbf{Q}]_{i_1,j_m}} && (47)\\
&\quad+\left[\mathbf{G}_{\mathcal{I}}^{\dagger}\right]_{2,2}\frac{\left|[\Delta\mathbf{R}_l]_{i_2,j_m}\right|^2}{\left|[\mathbf{Q}]_{i_2,j_m}\right|^2}+\left(\sum_{t\neq2}^{T}\left[\mathbf{G}_{\mathcal{I}}^{\dagger}\right]_{t,2}\frac{[\Delta\mathbf{R}_l]^*_{i_t,j_m}}{[\mathbf{Q}]^*_{i_t,j_m}}\right)\frac{[\Delta\mathbf{R}_l]_{i_2,j_m}}{[\mathbf{Q}]_{i_2,j_m}}\\
&\quad+\cdots && (48)\\
&\quad+\left[\mathbf{G}_{\mathcal{I}}^{\dagger}\right]_{T,T}\frac{\left|[\Delta\mathbf{R}_l]_{i_T,j_m}\right|^2}{\left|[\mathbf{Q}]_{i_T,j_m}\right|^2}+\left(\sum_{t\neq T}^{T}\left[\mathbf{G}_{\mathcal{I}}^{\dagger}\right]_{t,T}\frac{[\Delta\mathbf{R}_l]^*_{i_t,j_m}}{[\mathbf{Q}]^*_{i_t,j_m}}\right)\frac{[\Delta\mathbf{R}_l]_{i_T,j_m}}{[\mathbf{Q}]_{i_T,j_m}} && (49)
\end{aligned}$$

the noise and channel distributions are given by the theorem of iterated expectation as

$$\begin{aligned}\mathbb{E}\{[\mathbf{C}]_{m,m}\} &= \left[\mathbf{G}_{\mathcal{I}}^{\dagger}\right]_{1,1} \frac{\mathbb{E}_h\left\{[\mathbf{R}_l]_{i_1,i_1}[\mathbf{R}_l]_{j_m,j_m}\right\}}{N_s\left|[\mathbf{Q}]_{i_1,j_m}\right|^2} \\ &\quad + \left[\mathbf{G}_{\mathcal{I}}^{\dagger}\right]_{2,2} \frac{\mathbb{E}_h\left\{[\mathbf{R}_l]_{i_2,i_2}[\mathbf{R}_l]_{j_m,j_m}\right\}}{N_s\left|[\mathbf{Q}]_{i_2,j_m}\right|^2} \\ &\quad + \cdots + \left[\mathbf{G}_{\mathcal{I}}^{\dagger}\right]_{T,T} \frac{\mathbb{E}_h\left\{[\mathbf{R}_l]_{i_T,i_T}[\mathbf{R}_l]_{j_m,j_m}\right\}}{N_s\left|[\mathbf{Q}]_{i_T,j_m}\right|^2} \\ &= \frac{NL}{N_sT^2}\left(\frac{LN_t}{N}\sigma_h^2+\sigma_n^2\right)^2 \sum_{t=1}^{T} \frac{1}{\left|[\mathbf{Q}]_{i_t,j_m}\right|^2}, \end{aligned} \tag{53}$$

since $[\mathbf{G}_{\mathcal{I}}^{\dagger}]_{i,i} = NL/T^2$ for all i for equi-spaced indices. Now consider the off-diagonal elements of $\mathbf{C}$, given by

$$\begin{aligned}[\mathbf{C}]_{m,n} &= \left(\sum_{t=1}^{T}\left[\mathbf{G}_{\mathcal{I}}^{\dagger}\right]_{t,1}\frac{[\Delta\mathbf{R}_l]_{i_t,j_m}^*}{[\mathbf{Q}]_{i_t,j_m}^*}\right)\frac{[\Delta\mathbf{R}_l]_{i_1,j_n}}{[\mathbf{Q}]_{i_1,j_n}} \\ &\quad + \left(\sum_{t=1}^{T}\left[\mathbf{G}_{\mathcal{I}}^{\dagger}\right]_{t,2}\frac{[\Delta\mathbf{R}_l]_{i_t,j_m}^*}{[\mathbf{Q}]_{i_t,j_m}^*}\right)\frac{[\Delta\mathbf{R}_l]_{i_2,j_n}}{[\mathbf{Q}]_{i_2,j_n}} + \cdots \\ &\quad + \left(\sum_{t=1}^{T}\left[\mathbf{G}_{\mathcal{I}}^{\dagger}\right]_{t,T}\frac{[\Delta\mathbf{R}_l]_{i_t,j_m}^*}{[\mathbf{Q}]_{i_t,j_m}^*}\right)\frac{[\Delta\mathbf{R}_l]_{i_T,j_n}}{[\mathbf{Q}]_{i_T,j_n}}. \end{aligned} \tag{54}$$

Note that j_m and j_n are not equal. By (51) and the sparsity of the proposed $\mathbf{Q}$, all the off-diagonal elements have expectation zero. Hence, the claim follows. ■

Lemma 2:

$$\begin{aligned}&\mathrm{tr}\left(\mathbb{E}\left\{\left[\mathbf{z}'_{j_1}\cdots\mathbf{z}'_{j_T}\right]^{\mathrm{H}}\left(\tilde{\mathbf{F}}(\mathcal{I},:)\tilde{\mathbf{F}}(\mathcal{I},:)^{\mathrm{H}}\right)^{\dagger}\left[\mathbf{z}'_{j_1}\cdots\mathbf{z}'_{j_T}\right]\right\}\right. \\ &\left.\times\left(\tilde{\mathbf{F}}(\mathcal{J},:)\tilde{\mathbf{F}}(\mathcal{J},:)^{\mathrm{H}}\right)^{\dagger}\right) = \frac{\sigma_n^4}{N_sq^2}\sum_{i=1}^{T}\left[\mathbf{G}_{\mathcal{I}}^{\dagger}\right]_{i,i}\sum_{j=1}^{T}\left[\mathbf{G}_{\mathcal{J}}^{\dagger}\right]_{j,j},\end{aligned}$$

where $\mathbf{G}_{\mathcal{I}} = \tilde{\mathbf{F}}(\mathcal{I},:)\tilde{\mathbf{F}}(\mathcal{I},:)^{\mathrm{H}}$ and q^2 is the absolute value of off-diagonal elements of $\mathbf{Q}$.

Proof of Lemma 2: Let $\mathbf{D} := [\mathbf{z}'_{j_1}\cdots\mathbf{z}'_{j_T}]^{\mathrm{H}}\mathbf{G}_{\mathcal{I}}^{\dagger}[\mathbf{z}'_{j_1}\cdots\mathbf{z}'_{j_T}]$. Then, the diagonal elements of $\mathbf{D}$ are given by

$$\begin{aligned}[\mathbf{D}]_{m,m} &= \left[\sum_{t=1}^{T}\left[\mathbf{G}_{\mathcal{I}}^{\dagger}\right]_{t,1}\left(\frac{1}{N_s[\mathbf{Q}]_{i_t,j_m}^*}\sum_{i=1}^{N_s}\tilde{n}^{(i)*}(i_t)\tilde{n}^{(i)}(j_m)\right)\right] \\ &\quad\times\left(\frac{1}{N_s[\mathbf{Q}]_{i_1,j_m}}\sum_{i=1}^{N_s}\tilde{n}^{(i)}(i_1)\tilde{n}^{(i)*}(j_m)\right) \\ &\quad+\left[\sum_{t=1}^{T}\left[\mathbf{G}_{\mathcal{I}}^{\dagger}\right]_{t,2}\left(\frac{1}{N_s[\mathbf{Q}]_{i_t,j_m}^*}\sum_{i=1}^{N_s}\tilde{n}^{(i)*}(i_t)\tilde{n}^{(i)}(j_m)\right)\right] \\ &\quad\times\left(\frac{1}{N_s[\mathbf{Q}]_{i_2,j_m}}\sum_{i=1}^{N_s}\tilde{n}^{(i)}(i_2)\tilde{n}^{(i)*}(j_m)\right)+\cdots \\ &\quad+\left[\sum_{t=1}^{T}\left[\mathbf{G}_{\mathcal{I}}^{\dagger}\right]_{t,T}\left(\frac{1}{N_s[\mathbf{Q}]_{i_t,j_m}^*}\sum_{i=1}^{N_s}\tilde{n}^{(i)*}(i_t)\tilde{n}^{(i)}(j_m)\right)\right] \\ &\quad\times\left(\frac{1}{N_s[\mathbf{Q}]_{i_T,j_m}}\sum_{i=1}^{N_s}\tilde{n}^{(i)}(i_T)\tilde{n}^{(i)*}(j_m)\right), \end{aligned} \tag{55}$$

and $[\mathbf{D}]_{m,m}$ is rewritten as

$$\begin{aligned}[\mathbf{D}]_{m,m} &= \left[\mathbf{G}_{\mathcal{I}}^{\dagger}\right]_{1,1}\left|\frac{1}{N_s[\mathbf{Q}]_{i_1,j_m}}\sum_{i=1}^{N_s}\tilde{n}^{(i)}(i_1)\tilde{n}^{(i)*}(j_m)\right|^2 \\ &\quad+\left[\sum_{t\neq 1}^{T}\left[\mathbf{G}_{\mathcal{I}}^{\dagger}\right]_{t,1}\left(\frac{1}{N_s[\mathbf{Q}]_{i_t,j_m}^*}\sum_{i=1}^{N_s}\tilde{n}^{(i)*}(i_t)\tilde{n}^{(i)}(j_m)\right)\right] \\ &\quad\times\left(\frac{1}{N_s[\mathbf{Q}]_{i_1,j_m}}\sum_{i=1}^{N_s}\tilde{n}^{(i)}(i_1)\tilde{n}^{(i)*}(j_m)\right) \\ &\quad+\left[\mathbf{G}_{\mathcal{I}}^{\dagger}\right]_{2,2}\left|\frac{1}{N_s[\mathbf{Q}]_{i_2,j_m}}\sum_{i=1}^{N_s}\tilde{n}^{(i)}(i_2)\tilde{n}^{(i)*}(j_m)\right|^2 \\ &\quad+\left[\sum_{t\neq 2}^{T}\left[\mathbf{G}_{\mathcal{I}}^{\dagger}\right]_{t,2}\left(\frac{1}{N_s[\mathbf{Q}]_{i_t,j_m}^*}\sum_{i=1}^{N_s}\tilde{n}^{(i)*}(i_t)\tilde{n}^{(i)}(j_m)\right)\right] \\ &\quad\times\left(\frac{1}{N_s[\mathbf{Q}]_{i_2,j_m}}\sum_{i=1}^{N_s}\tilde{n}^{(i)}(i_2)\tilde{n}^{(i)*}(j_m)\right)+\cdots \\ &\quad+\left[\mathbf{G}_{\mathcal{I}}^{\dagger}\right]_{T,T}\left|\frac{1}{N_s[\mathbf{Q}]_{i_T,j_m}}\sum_{i=1}^{N_s}\tilde{n}^{(i)}(i_T)\tilde{n}^{(i)*}(j_m)\right|^2 \\ &\quad+\left[\sum_{t\neq T}^{T}\left[\mathbf{G}_{\mathcal{I}}^{\dagger}\right]_{t,T}\left(\frac{1}{N_s[\mathbf{Q}]_{i_t,j_m}^*}\sum_{i=1}^{N_s}\tilde{n}^{(i)*}(i_t)\tilde{n}^{(i)}(j_m)\right)\right] \\ &\quad\times\left(\frac{1}{N_s[\mathbf{Q}]_{i_T,j_m}}\sum_{i=1}^{N_s}\tilde{n}^{(i)}(i_T)\tilde{n}^{(i)*}(j_m)\right). \end{aligned}$$

Since $\mathcal{I}\cap\mathcal{J}=\emptyset$, the expectation of $\tilde{n}^{(i)}(i_m)\tilde{n}^{(i)}(j_n)$ is zero for all m,n and so is the expectation of $\tilde{n}^{(i)}(j_m)\tilde{n}^{(i)}(j_n)$ for $m\neq n$. Thus, we have

$$\begin{aligned}&\mathbb{E}\left\{[\mathbf{D}]_{m,m}\right\} \\ &= \frac{\sigma_n^4\left[\mathbf{G}_{\mathcal{I}}^{\dagger}\right]_{1,1}}{N_s\left|[\mathbf{Q}]_{i_m,j_1}\right|^2}+\frac{\sigma_n^4\left[\mathbf{G}_{\mathcal{I}}^{\dagger}\right]_{2,2}}{N_s\left|[\mathbf{Q}]_{i_m,j_2}\right|^2}+\cdots+\frac{\sigma_n^4\left[\mathbf{G}_{\mathcal{I}}^{\dagger}\right]_{T,T}}{N_s\left|[\mathbf{Q}]_{i_m,j_T}\right|^2} \\ &\overset{(a)}{=}\frac{\sigma_n^4}{N_sq^2}\sum_{t=1}^{T}\left[\mathbf{G}_{\mathcal{I}}^{\dagger}\right]_{t,t}, \end{aligned} \tag{56}$$

where step (a) is by the assumption of $q^2 = |[\mathbf{Q}]_{i_m,j_t}|^2$ for $1\leq t\leq T$. The off-diagonal elements of $\mathbf{D}$ are all zero after the expectation since $\tilde{n}^{(i)}(i_m)\tilde{n}^{(i)}(i_n)=0, m\neq n$. Thus, we have

$$\mathbb{E}\left\{\left[\mathbf{z}'_{j_1}\cdots\mathbf{z}'_{j_T}\right]^{\mathrm{H}}\mathbf{G}_{\mathcal{I}}^{\dagger}\left[\mathbf{z}'_{j_1}\cdots\mathbf{z}'_{j_T}\right]\right\} = \frac{\sigma_n^4}{N_sq^2}\left(\sum_{t=1}^{T}\left[\mathbf{G}_{\mathcal{I}}^{\dagger}\right]_{t,t}\right)\mathbf{I}_T,$$

$$\begin{aligned}&\mathrm{tr}\left(\mathbb{E}\{\mathbf{D}\}\left(\tilde{\mathbf{F}}(\mathcal{J},:)\tilde{\mathbf{F}}(\mathcal{J},:)^{\mathrm{H}}\right)^{\dagger}\right) \\ &= \frac{\sigma_n^4}{N_sq^2}\left(\sum_{t=1}^{T}\left[\mathbf{G}_{\mathcal{I}}^{\dagger}\right]_{t,t}\right)\left(\sum_{j=1}^{T}\left[\left(\tilde{\mathbf{F}}(\mathcal{J},:)\tilde{\mathbf{F}}(\mathcal{J},:)^{\mathrm{H}}\right)^{\dagger}\right]_{j,j}\right).\end{aligned}$$

Hence, the claim follows. ■

References

[1] C.-C. Tu and B. Champagne, "Blind recursive subspace-based identification of time-varying sideband MIMO channels," *IEEE Trans. Veh. Technol.*, vol. 61, no. 2, pp. 662–674, Feb. 2012.

[2] H. Ngo and E. G. Larsson, "EVD-based channel estimation in multi cell multiuser MIMO systems with very large antenna arrays," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, Kyoto, Japan, Mar. 2012, pp. 3249–3252.

[3] R. Müller, L. Cottatellucci, and M. Vehkaperä, "Blind pilot decontamination," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 773–786, Oct. 2014

[4] M.-A. Khalighi and J. J. Boutros, "Semi-blind channel estimation using the EM algorithm in iterative MIMO APP detectors," *IEEE Trans. Wireless Commun.*, vol. 5, no. 11, pp. 3165–3173, Nov. 2006.

[5] M. Nicoli, S. Ferrara, and U. Spagnolini, "Soft-iterative channel estimation: Methods and performance analysis," *IEEE Trans. Signal Process.*, vol. 55, no. 6, pp. 2993–3006, Jun. 2007.

[6] B. Muquet, M. de Courville, and P. Duhamel, "Subspace based blind and semi blind channel estimation for OFDM systems," *IEEE Trans. Signal Process.*, vol. 50, no. 7, pp. 1699–1712, Jul. 2002.

[7] C. Shin, Jr., R. W. Heath, and E. J. Powers, "Blind channel estimation for MIMO-OFDM systems," *IEEE Trans. Veh. Technol.*, vol. 56, no. 2, pp. 670–685, Mar. 2007.

[8] A. P. Petropulu and R. Zhang, "Blind OFDM channel estimation through simple linear precoding," *IEEE Trans. Wireless Commun.*, vol. 3, no. 2, pp. 647–655, Mar. 2004.

[9] R. Zhang, "Blind channel estimation for precoded OFDM system," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, Philadelphia, PA, USA, Mar. 2005, pp. iii/469–iii/472.

[10] R. Lin and A. P. Petropulu, "Linear precoding assisted blind channel estimation for OFDM systems," *IEEE Trans. Veh. Technol.*, vol. 54, no. 3, pp. 983–995, May 2005.

[11] F. Gao and A. Nallanathan, "Blind channel estimaton for OFDM systems via generalized precoding," *IEEE Trans. Veh. Technol.*, vol. 56, no. 3, pp. 1155–1164, May 2007.

[12] F. Gao and A. Nallanathan, "Blind channel estimation for MIMO OFDM systems via nonredundant linear precoding," *IEEE Trans. Signal Process.*, vol. 55, no. 2, pp. 784–789, Jan. 2007.

[13] C. Shin, R. W. Heath, Jr., and E. J. Powers, "Non-redundant precoding-based blind and semi-blind channel estimation for MIMO block transmission with a cyclic prefix," *IEEE Trans. Signal Process.*, vol. 56, no. 6, pp. 2509–2523, Jun. 2008.

[14] C.-C. Tu and B. Champagne, "Subspace tracking of fast time-varying channels in precoded MIMO-OFDM systems," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, Taipei, Taiwan, Apr. 2009, pp. 2565–2568.

[15] X. Liao, L. Fan, and F. Gao, "Blind channel estimation for OFDM modulated two-way relay network," in *Proc. IEEE Wireless Commun. Netw. Conf.*, Sydney, NSW, Australia, Apr. 2010, pp. 1–5.

[16] D. Tse and P. Viswanath, *Fundamentals of Wireless Communications*. Cambridge, U.K.: Cambridge Univ., 2005.

[17] M. Vu and A. Paulraj, "MIMO wireless linear precoding," *IEEE Signal Process. Mag.*, vol. 24, no. 5, pp. 86–105, Sep. 2007.

[18] S. Noh and Y. Sung, "Precoder design for blind channel estimation in multiple OFDM systems," in *Proc. Asia Pacific Signal Inf. Process. Assoc.*, Biopolis, Singapore, Dec. 2010, pp. 137–140.

[19] S. Noh and Y. Sung, "Linear precoder design for blind channel estimation for MIMO-OFDM systems under a sparse structure," in *Proc. IEEK*, Jeju, Korea, Jun. 2011, pp. 33–36.

[20] S. Noh and M. D. Zoltowski, "A new precoder design for precoding-based blind channel estimation for MIMO-OFDM systems," in *Proc. IEEE Global Commun. Conf.*, Atlanta, GA, USA, Dec. 2013, pp. 3306–3311.

[21] G. H. Golub and C. F. Van Loan, *Matrix Computations*, 3rd ed. Baltimore, MD, USA: The Johns Hopkins Univ. Press, 1996.

[22] Y. Sung, L. Tong, and A. Swami, "Blind channel estimation for space-time coded WCDMA," *EURASIP J. Wireless Commun. Netw.*, vol. 2004, no. 2, pp. 322–334, Dec. 2004.

[23] C.-C. Tu and B. Champagne, "Subspace-based blind channel estimation for MIMO-OFDM systems with reduced time averaging," *IEEE Trans. Veh. Technol.*, vol. 59, no. 3, pp. 1539–1544, Mar. 2010.

[24] F. J. Simois, J. J. M. Fuentes, R. B. Tortosa, and L. Salamanca, "Near the Cramér-Rao bound precoding algorithms for OFDM blind channel estimation," *IEEE Trans. Veh. Technol.*, vol. 61, no. 2, pp. 651–661, Feb. 2012.

[25] R. Negi and J. Cioffi, "Pilot tone selection for channel estimation in a mobile OFDM system," *IEEE Trans. Consum. Electron.*, vol. 44, no. 7, pp. 1112–1128, Aug. 1998.

[26] Y. Sung, Y. Lim, L. Tong, and A.-J. van der Veen, "Signal processing advances for 3G WCDMA: From rake receivers to blind techniques," *IEEE Commun. Mag.*, vol. 47, no. 1, pp. 48–54, Jan. 2009.

[27] P. Lancaster and M. Tismenetsky, *The Theory of Matrices: With Applications*, 2nd ed. San Diego, CA, USA: Academic, 1985.

[28] Z. Xu, "Perturbation analysis for subspace decomposition with applications in subspace-based algorithms," *IEEE Trans. Signal Process.*, vol. 50, no. 11, pp. 2820–2830, Nov. 2002.

[29] F. Gao and A. Nallanathan, "Resolving multidimensional ambiguity in blind channel estimation of MIMO-FIR systems via block precoding," *IEEE Trans. Veh. Technol.*, vol. 57, no. 1, pp. 11–21, Jan. 2008.

[30] S. Noh and M. D. Zoltowski, "Blind separation for precoding-based blind channel estimation for MIMO-OFDM systems," in *Proc. Asilomar Conf. Signal, Syst. Comput.*, Pacific Grove, CA, USA, Nov. 2013, pp. 1263–1267.

[31] S. Chen and L. Hanzo, "Fast converging semi-blind space-time equalisation for dispersive QAM MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 8, no. 8, pp. 3969–3974, Aug. 2009.

[32] Z. Xu, "On the second-order statistics of the weighted sample covariance matrix," *IEEE Trans. Signal Process.*, vol. 51, no. 2, pp. 527–534, Feb. 2003.

Song Noh (S'08) received the B.S. degree from Soongsil University, Seoul, Korea, in 2008 and the M.S. degree from Korea Advanced Institute of Science and Technology, Daejeon, Korea, in 2010, both in electrical engineering. He is currently working toward the Ph.D. degree in electrical and computer engineering at Purdue University, West Lafayette, IN, USA. His research interests are in the design and analysis of large scale antenna systems and adaptive signal processing for wireless communications.

Youngchul Sung (S'92–M'93–SM'09) received the B.S. and M.S. degrees from Seoul National University, Seoul, Korea, in electronics engineering in 1993 and 1995, respectively. After working at LG Electronics, Ltd., Seoul, Korea, from 1995 to 2000, he joined the Ph.D. program and received the Ph.D. degree in electrical and computer engineering from Cornell University, Ithaca, NY, USA, in 2005. From 2005 until 2007, he was a Senior Engineer in the Corporate R&D Center of Qualcomm, Inc., San Diego, CA, USA, and participated in design of WCDMA base station modem. Since 2007, he has been on the faculty of the Department of Electrical Engineering in KAIST, Daejeon, Korea. His research interests include signal processing for communications, statistical signal processing, asymptotic statistics, and information geometry.

Dr. Sung is currently a member of Signal and Information Processing Theory and Methods (SIPTM) Technical Committee of Asia-Pacific Signal and Information Processing Association (APSIPA), Vice-Chair of IEEE ComSoc Asia-Pacific Board ISC, and was an Associate Editor of the IEEE SIGNAL PROCESSING LETTERS from 2012 to 2014.

Michael D. Zoltowski (F'13) joined the faculty of Purdue University in Fall 1986 where he was named a University Faculty Scholar in 2001. He currently an Endowed Chaired position as the "Thomas J. and Wendy Engibous Professor of Electrical and Computer Engineering," In this capacity he has won several teaching awards, most recently the 2012 Engineering Distance Education Award. He received the 2002 Technical Achievement Award from the IEEE Signal Processing Society (SPS). He served as a 2003 Distinguished Lecturer for the IEEE SPS. Dr. Zoltowski is a co-recipient of the IEEE Communications Society 2001 Leonard G. Abraham Prize Paper Award, the IEEE SPS 1991 Paper Award, "The 1998 Fred Ellersick MILCOM Award for Best Paper," and a Best Paper Award at the 2000 IEEE International Symposium on Spread Spectrum Techniques and Applications.

From 1998 to 2001, Dr. Zoltowski served as an elected Member-at-Large of the Board of Governors and Secretary of the IEEE Signal Processing Society (SPS.) From 2003–2005, he served on the Awards Board of the IEEE SPS. and also served as the Area Editor in charge of Feature Articles for the IEEE Signal Processing Magazine. He has served on various SPS Technical Committees. From 2005–2006, he served as Chair of the Technical Committee on Sensor and Multichannel (SAM) Processing. He has served as an Associate Editor for both the IEEE Transactions on Signal Processing and the IEEE Communications Letters. He was Technical Chair for the 2006 IEEE Sensor Array and Multichannel Workshop. He served as Vice-President for Awards and Membership for the IEEE SPS for 2008–2010. His research interests in lie in all areas of MIMO wireless communications, MIMO radar, and GPS.