Subspace-Based (Semi-) Blind Channel Estimation for Block Precoded Space-Time OFDM

Shengli Zhou, *Student Member, IEEE*, Bertrand Muquet, and Georgios B. Giannakis, *Fellow, IEEE*

***Abstract*—Space time coding has by now been well documented as an attractive means of achieving high data rate transmissions with diversity and coding gains, provided that the underlying propagation channels can be accounted for. In this paper, we rely on redundant linear precoding to develop a (semi-)blind channel estimation algorithm for space time (ST) orthogonal frequency division multiplexing (OFDM) transmissions with Alamouti's block code applied on each subcarrier. We establish that multichannel identifiability is guaranteed up to one or two scalar ambiguities, regardless of the channel zero locations and the underlying signal constellations, when distinct or identical precoders are employed for even and odd indexed symbol blocks. With known pilots inserted either before or after precoding, we resolve the residual scalar ambiguities and show that distinct precoders require half the number of pilots than identical precoders to achieve the same channel estimation accuracy. Simulation results confirm our theoretical analysis and illustrate that the proposed semi-blind algorithm is capable of tracking slow channel variations and improving the overall system performance relative to competing differential ST alternatives.**

***Index Terms*—Block precoded transmissions, OFDM, semi-blind, space-time coding, subspace based channel estimation.**

I. Introduction

NEW applications such as high-speed internet access and wireless digital television call for very high data rate transmissions. Using multiple transmit and receive antennas has the potential to increase the channel capacity and, thus, the maximum achievable rate by an order of magnitude or more [7]. Equipped with space-time coding (STC) at the transmitter and intelligent signal processing at the receiver, multiantenna transceivers offer also diversity and coding advantages over single antenna systems (see [18], [20] for tutorial treatments). However, all these enhancements in capacity, diversity, and coding gains can be realized if the underlying channels (flat or frequency-selective) can be acquired at the receiver.

For flat-fading channels, unitary or differential STC approaches dispense with channel estimation at the expense of performance or power loss relative to coherent detection [11]–[13], [22]. For frequency-selective channels, delay diversity space time orthogonal frequency division multiplexing (ST-OFDM) treats the links between multiple transmit antennas and each receive antenna as a single channel and, thus, reduces channel estimation to a single-input single-output (SISO) problem [14]. However, for most STC transceivers, *multichannel* estimation algorithms are needed. Training symbols are transmitted periodically in [10] for the receiver to acquire the multi-input multi-output (MIMO) frequency-flat channels (see also [15] for training-based estimation of frequency-selective channels in ST-OFDM). However, training sequences consume bandwidth and, thereby, incur spectral efficiency (and thus capacity) loss, especially in rapidly varying environments and low signal-to-noise ratios (SNRs)[1] [10]. For this reason, blind channel estimation methods receive growing attention, especially for estimating the MIMO channels corresponding to multiple transmit and receive antennas; see e.g., [9, ch. 3–7] and references therein. However, only a few works have been reported so far on blind MIMO and multi-input single-output (MISO) channel estimation that exploits the unique features of ST codes. Relying on nonredundant and nonconstant modulus precoding, blind channel estimation and equalization for OFDM-based multi-antenna systems has been proposed in [2] using cyclostationary statistics. For ST-OFDM, a deterministic blind channel estimator was derived in [17] when the channel transfer functions are coprime (no common zeros) and the transmitted signals have constant-modulus (CM).

Owing to size and power limitations, mobile units (MUs) usually cannot afford more than two antennas; thus, deployment of one or two transmit antennas is of practical interest in the uplink. Even in the downlink, where more transmit antennas are allowed at the base station (BS), the configuration of two transmit-antennas and one receive antenna is very popular [23].

In this paper, we deal with a linearly precoded ST-OFDM system with two transmit antennas and derive (semi-)blind channel identification algorithms for frequency-selective FIR channels. Due to frequency-selectivity, subcarriers on (or near) the common channel nulls (fades) may be severely attenuated, which leads to significant performance loss. To robustify the performance against channel nulls, linear redundant precoders were introduced in [16] and [18] for ST-OFDM to guarantee symbol recovery, regardless of the channel zero locations. In this paper, we establish that linear precoding also enables (semi-) blind subspace-based channel estimation for ST-OFDM systems. Recall that without precoding, blind channel identifia-

Manuscript received February 12, 2001; revised January 28, 2002. This work was supported by the NSF Wireless Initiative under Grant 99-79443, NSF Grant 0105612, and by the ARL/CTA under Grant DAAD19-01-2-011. This work was presented in part at the 34th Asilomar Conference on Signals and Systems, Pacific Grove, CA, October 29–November 1, 2000, and at the 11th IEEE Workshop on Statistical Signal Processing, Singapore, August 6–8, 2001. The associate editor coordinating the review of this paper and approving it for publication was Dr. Helmut Bölcskei.

S. Zhou and G. B. Giannakis are with Department of Electrical and Computer Engineering, University of Minnesota, Minneapolis, MN 55455 USA (e-mail: szhou@ece.umn.edu; georgios@ece.umn.edu).

B. Muquet was with Motorola Labs Paris, Gif-sur-Yvette Cedex, France. He is now with Stepmind, Boulogne-Billancourt, France (e-mail: bertrand.muquet@stepmind.com).

Publisher Item Identifier S 1053-587X(02)03153-7.

[1]See also [29], where reduced-capacity claims are quantified when channel estimation is accounted for.

bility is impossible for ST-OFDM without imposing constraints on the transmitted signal constellation [17]. However, even with CM constellations, the approach in [17] is only applicable to channel pairs that do not share common zeros on the FFT grid, which cannot be checked at the transmitter. However, with properly designed redundant precoders, we will show here that subspace-based blind channel estimation is possible and possesses the following three attractive features.

i) It can be applied to arbitrary signal constellations.
ii) It guarantees channel identifiability, regardless of the underlying channel zero locations.
iii) It can estimate multiple channels simultaneously up to one or two scalar ambiguities.

To enable channel equalization, we also demonstrate how to resolve the residual scalar ambiguities using a minimal number of pilots that can be inserted either before or after precoding. By reducing two scalar ambiguities to one, distinct precoders will turn out to outperform identical precoders and will require half the number of pilots to achieve the same channel estimation accuracy. Being capable of tracking slow channel variations, the (semi-) blind channel estimator herein outperforms the training based approach in time varying setups and dispenses with frequent training. It also leads to the performance improvement of linearly precoded ST-OFDM (that can take advantage of coherent detection) relative to differential ST alternatives that bypass channel estimation but are limited to semi-coherent demodulation.

The rest of this paper is organized as follows. After presenting the system model in Section II, we develop our novel algorithm in Section III and address blind identifiability issues in Section IV. Section V is devoted to the scalar ambiguity determination, whereas Section VI demonstrates the advantage of distinct over identical precoders. Section VII presents simulation results, and Section VIII gathers our conclusions.

Notation: Bold uppercase letters denote matrices, and bold lowercase letters stand for column vectors. $(\cdot)^*, (\cdot)^T$ and $(\cdot)^{\mathcal{H}}$ denote conjugate, transpose, and Hermitian transpose; $E\{\cdot\}$ stands for expectation; $\|\cdot\|$ stands for the norm of a vector; $\delta(\cdot)$ stands for the Kronecker's delta; $\mathbf{I}_K$ denotes the identity matrix of size K, and $\mathbf{0}_{K\times P}$ denotes an all-zero matrix of size $K \times P$; diag($\mathbf{x}$) will stand for a diagonal matrix with $\mathbf{x}$ on its diagonal; $[\cdot]_i$ denotes the ith entry of a vector, and $[\cdot]_{i,j}$ denotes the (i,j)th entry of a matrix; Matlab's notation $\mathbf{A}(:, i:j)$ is used to denote a submatrix of $\mathbf{A}$ constructed from columns i to j.

II. System Description

Fig. 1 depicts the wireless system considered in this paper, where the ST transceiver is equipped with two transmit antennas as in [1] and [18]. Without loss of generality, we here focus on one receive antenna since channel estimation on each receive antenna can be carried out separately. Prior to transmission, the information bearing symbols are first grouped into blocks $\mathbf{s}(n)$ of size $K \times 1$. Two different linear block precoders denoted by the tall $J \times K$ matrices $\mathbf{\Theta}_1$ and $\mathbf{\Theta}_2$ (one for the even block indices $2n$ and one for the odd indices $2n+1$) are used to introduce redundancy ($J > K$). The corresponding $J \times 1$ precoded blocks

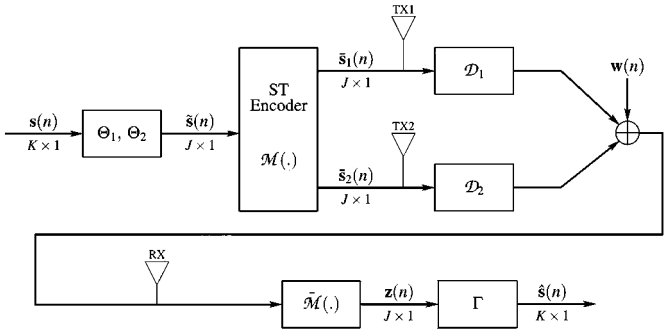

Fig. 1. Block precoded ST-OFDM transceiver model.

$$\tilde{\mathbf{s}}(2n) := \mathbf{\Theta}_1 \mathbf{s}(2n) \quad \text{and} \quad \tilde{\mathbf{s}}(2n+1) := \mathbf{\Theta}_2 \mathbf{s}(2n+1) \tag{1}$$

are fed to the ST encoder $\mathcal{M}(\cdot)$. As in [16] and [18], the redundancy introduced by these block precoders will turn out to be instrumental for intersymbol interference (ISI) elimination and symbol recovery, regardless of the underlying frequency-selective FIR channels. However, as we will show in this paper, redundant precoding also enables (semi-) blind identification of the multiple channels. The ST encoder takes as input two consecutive precoded blocks $\tilde{\mathbf{s}}(2n)$ and $\tilde{\mathbf{s}}(2n+1)$ to output the following $2J \times 2$ code matrix:

$$\begin{bmatrix} \bar{\mathbf{s}}_1(2n) & \bar{\mathbf{s}}_1(2n+1) \\ \bar{\mathbf{s}}_2(2n) & \bar{\mathbf{s}}_2(2n+1) \end{bmatrix} := \begin{bmatrix} \tilde{\mathbf{s}}(2n) & -\tilde{\mathbf{s}}^*(2n+1) \\ \tilde{\mathbf{s}}(2n+1) & \tilde{\mathbf{s}}^*(2n) \end{bmatrix} \tag{2}$$

where each block column is transmitted over successive time intervals with the blocks $\bar{\mathbf{s}}_1(n)$ and $\bar{\mathbf{s}}_2(n)$ sent through transmit-antennas 1 and 2, respectively. Note that without precoding and symbol blocking ($J = 1$), this code matrix reduces to the well known Alamouti's block STC [1].

We assume in what follows that the channels between the two transmit antennas and the receive antenna are frequency selective and that their baseband equivalent effect in discrete time is captured by an FIR linear time-invariant filter with impulse response vectors $\mathbf{h}_\mu := [h_\mu(0), \ldots, h_\mu(L)], \mu = 1, 2$, where L is an upper bound for the channel orders of $\mathbf{h}_1$ and $\mathbf{h}_2$, i.e., $L \geq \max(L_1, L_2)$ if L_1 is the channel order for $\mathbf{h}_1$ and L_2 for $\mathbf{h}_2$. Moreover, we assume that an OFDM modulator at the transmitter together with the corresponding demodulator at the receiver is deployed to convert the FIR channels to a set of parallel flat faded subchannels (see, e.g., [25] for detailed derivations). Let $\mathcal{D}_1$ and $\mathcal{D}_2$ be the diagonal matrices corresponding to subchannels $\mathcal{D}_\mu := \text{diag}(H_\mu(0) \ldots H_\mu(J-1))$, where $H_\mu(k) := \sum_{l=0}^{L} h_\mu(l) e^{-j(2\pi/J)lk}$. Considering two successive received blocks $\bar{\mathbf{y}}(2n)$ and $\bar{\mathbf{y}}(2n+1)$, let us define the super blocks $\check{\mathbf{y}}(n)$ and $\check{\mathbf{s}}(n)$ as $\check{\mathbf{y}}(n) := [\bar{\mathbf{y}}^T(2n), \bar{\mathbf{y}}^{\mathcal{H}}(2n+1)]^T$ and $\check{\mathbf{s}}(n) := [\mathbf{s}^T(2n), \mathbf{s}^T(2n+1)]^T$. In addition, let $\check{\mathbf{w}}(n)$ be the noise added to the noise-free version of $\check{\mathbf{y}}(n)$ that is denoted by $\check{\mathbf{x}}(n)$. The received noisy block $\check{\mathbf{y}}(n)$ can then be expressed as (see also [18] for further details)

$$\begin{aligned} \check{\mathbf{y}}(n) &= \mathcal{D}\mathbf{\Phi}_{12}\check{\mathbf{s}}(n) + \check{\mathbf{w}}(n) \\ &:= \mathcal{H}\check{\mathbf{s}}(n) + \check{\mathbf{w}}(n) := \check{\mathbf{x}}(n) + \check{\mathbf{w}}(n) \end{aligned} \tag{3}$$

where $\mathcal{D}, \Phi_{12}, \mathcal{H}$ are defined, respectively, as

$$\mathcal{D}:=\begin{bmatrix}\mathcal{D}_1 & \mathcal{D}_2\\ \mathcal{D}_2^* & -\mathcal{D}_1^*\end{bmatrix},\quad \Phi_{12}:=\begin{bmatrix}\Theta_1 & 0\\ 0 & \Theta_2\end{bmatrix},\quad \mathcal{H}:=\mathcal{D}\Phi_{12}. \tag{4}$$

When the channel matrices $\mathcal{D}_1$ and $\mathcal{D}_2$ become available at the receiver, it is possible to demodulate $\check{\mathbf{y}}(n)$ with diversity gains by a simple matrix multiplication

$$\check{\mathbf{z}}(n)=\mathcal{D}^{\mathcal{H}}\check{\mathbf{y}}(n)=\begin{bmatrix}\bar{\mathcal{D}}_{12}\Theta_1 & 0\\ 0 & \bar{\mathcal{D}}_{12}\Theta_2\end{bmatrix}\check{\mathbf{s}}(n)+\mathcal{D}^{\mathcal{H}}\check{\mathbf{w}}(n) \tag{5}$$

where $\bar{\mathcal{D}}_{12}:=\mathcal{D}_1^*\mathcal{D}_1+\mathcal{D}_2^*\mathcal{D}_2$. We infer that multi-antenna diversity of order two has been achieved because $\bar{\mathcal{D}}_{12}=\text{diag}(\sum_{\mu=1}^2|H_\mu(0)|^2,\ldots,\sum_{\mu=1}^2|H_\mu(J-1)|^2)$.

Equation (5) reveals that zero-forcing recovery of $\check{\mathbf{s}}(n)$ from $\check{\mathbf{z}}(n)$ requires the matrices $\bar{\mathcal{D}}_{12}\Theta_\mu$, $\mu\in[1,2]$ to be full column rank. Because the channels have maximum order L, $\bar{\mathcal{D}}_{12}$ has at most L zero diagonal entries. Hence, the full rank of $\bar{\mathcal{D}}_{12}\Theta_\mu$ can be always assured if we adopt the following design conditions on the block lengths and the linear precoders:[2]

a1) $J>K+L$.

a2) $\Theta_\mu, \mu\in\{1,2\}$, is designed so that any K rows of Θ_μ are linearly independent.

Based on a1) and a2), our objective in this paper is to develop subspace-based (semi-) blind multichannel estimation algorithms and show that channel identifiability is guaranteed, regardless of the channel zero locations. We will present the algorithms first in Section III and prove the identifiability results in Section IV. Furthermore, we will study channel identifiability even when a1) is violated.

III. Subspace-Based Multichannel Estimation

Before addressing the noisy case, we will start from the noiseless vectors $\check{\mathbf{x}}(n)$ in (3). To estimate the channels $\{\mathbf{h}_\mu\}_{\mu=1}^2$ [or, equivalently, $\mathcal{H}$ in (3)], the receiver collects N blocks of $\check{\mathbf{x}}(n)$ in a $2J\times N$ matrix $\mathbf{X}_N:=[\check{\mathbf{x}}(0),\ldots,\check{\mathbf{x}}(N-1)]$ and forms $\mathbf{X}_N\mathbf{X}_N^{\mathcal{H}}=\mathcal{H}\mathbf{S}_N\mathbf{S}_N^{\mathcal{H}}\mathcal{H}^{\mathcal{H}}$, where $\mathbf{S}_N:=[\check{\mathbf{s}}(0),\ldots,\check{\mathbf{s}}(N-1)]$. At the receiver end, we also select the following.

a3) The number of blocks N is large enough ($\geq 2K$) so that $\mathbf{S}_N\mathbf{S}_N^{\mathcal{H}}$ has full rank $2K$.

Condition a3) expresses the standard "persistence of excitation" assumption that is satisfied by all signal constellations for N sufficiently large ($N\geq 2K$ would suffice in most cases).

Under a1) and a2), we maintain that matrix $\mathcal{H}$ in (3) has always full column rank. To verify this, it suffices to show that $\mathcal{D}^{\mathcal{H}}\mathcal{H}$ has full column rank (since premultiplication by $\mathcal{D}^{\mathcal{H}}$ can only reduce the rank). Note that $\mathcal{D}^{\mathcal{H}}\mathcal{H}=[\mathbf{I}_2\otimes\bar{\mathcal{D}}_{12}]\Phi_{12}$, where $\otimes$ stands for the Kronecker product. Since $\bar{\mathcal{D}}_{12}$ has at most L zero diagonal entries and any K rows of Θ_μ are linearly independent, each diagonal block submatrix $\bar{\mathcal{D}}_{12}\Theta_\mu$ of $\mathcal{D}^{\mathcal{H}}\mathcal{H}$ has full column rank. Thus, $\mathcal{D}^{\mathcal{H}}\mathcal{H}$, and consequently, $\mathcal{H}$ have full column rank. Therefore, a1) and a2), together with a3), imply that $\text{rank}(\mathbf{X}_N\mathbf{X}_N^{\mathcal{H}})=2K$, and the range space $\mathcal{R}(\mathbf{X}_N\mathbf{X}_N^{\mathcal{H}})=\mathcal{R}(\mathcal{H})$. Hence, the nullity of $\mathbf{X}_N\mathbf{X}_N^{\mathcal{H}}$ is $\nu(\mathbf{X}_N\mathbf{X}_N^{\mathcal{H}})=2J-2K$. Further, the eigendecomposition

$$\mathbf{X}_N\mathbf{X}_N^{\mathcal{H}}=[\mathbf{U}\quad\tilde{\mathbf{U}}]\begin{bmatrix}\Sigma & 0\\ 0 & 0\end{bmatrix}\begin{bmatrix}\mathbf{U}^{\mathcal{H}}\\ \tilde{\mathbf{U}}^{\mathcal{H}}\end{bmatrix} \tag{6}$$

where Σ is a diagonal matrix of size $2K\times 2K$ with nonzero diagonal entries, yields the $2J\times(2J-2K)$ matrix $\tilde{\mathbf{U}}$, whose columns span the null space $\mathcal{N}(\mathbf{X}_N\mathbf{X}_N^{\mathcal{H}})$. Because the latter is orthogonal to $\mathcal{R}(\mathbf{X}_N\mathbf{X}_N^{\mathcal{H}})=\mathcal{R}(\mathcal{H})$, it follows that $\tilde{\mathbf{u}}_k^{\mathcal{H}}\mathcal{H}=\mathbf{0}_{2K\times 1}^T$ for $k\in[1,2J-2K]$, where $\tilde{\mathbf{u}}_k$ stands for the kth column of $\tilde{\mathbf{U}}$.

Let us now split the vector $\tilde{\mathbf{u}}_k$ into its upper and lower parts as $\tilde{\mathbf{u}}_k=[\hat{\mathbf{u}}_k^T,\check{\mathbf{u}}_k^T]^T$, where $\hat{\mathbf{u}}_k$ and $\check{\mathbf{u}}_k$ are $J\times 1$ vectors. Using (4), we can factor $\tilde{\mathbf{u}}_k^{\mathcal{H}}\mathcal{H}$ as

$$[\hat{\mathbf{u}}_k^{\mathcal{H}},\check{\mathbf{u}}_k^{\mathcal{H}}]\begin{bmatrix}\mathcal{D}_1 & \mathcal{D}_2\\ \mathcal{D}_2^* & -\mathcal{D}_1^*\end{bmatrix}\begin{bmatrix}\Theta_1 & 0\\ 0 & \Theta_2\end{bmatrix}=\mathbf{0}^T. \tag{7}$$

Define $\tilde{\mathbf{h}}_\mu:=[H_\mu(0),\ldots,H_\mu(J-1)]^T$, and let $\mathbf{D}(\mathbf{v})$ stand for the diagonal matrix with the vector $\mathbf{v}$ on its diagonal. Since for any $J\times 1$ vectors $\mathbf{a}$ and $\mathbf{b}$ it holds that $\mathbf{a}^{\mathcal{H}}\mathbf{D}(\mathbf{b})=\mathbf{b}^T\mathbf{D}(\mathbf{a}^*)$, (7) can be rewritten as

$$\left[\tilde{\mathbf{h}}_1^T,\tilde{\mathbf{h}}_2^{\mathcal{H}}\right]\begin{bmatrix}\mathbf{D}(\hat{\mathbf{u}}_k^*) & -\mathbf{D}(\check{\mathbf{u}}_k)\\ \mathbf{D}(\check{\mathbf{u}}_k^*) & \mathbf{D}(\hat{\mathbf{u}}_k)\end{bmatrix}\underbrace{\begin{bmatrix}\Theta_1 & 0\\ 0 & \Theta_2^*\end{bmatrix}}_{:=\Psi}=\mathbf{0}^T. \tag{8}$$

With $\mathbf{V}$ denoting the $J\times(L+1)$ Vandermonde matrix with $(p+1,q+1)$st entry $e^{-j(2\pi/J)pq}$, one can write the FFT in matrix form and express each channel's transfer function vector as $\tilde{\mathbf{h}}_\mu=\mathbf{V}\mathbf{h}_\mu$. Plugging the latter into (8), we have

$$\left[\mathbf{h}_1^T,\mathbf{h}_2^{\mathcal{H}}\right]\underbrace{\begin{bmatrix}\mathbf{V}^T & 0\\ 0 & \mathbf{V}^{\mathcal{H}}\end{bmatrix}}_{:=\mathcal{F}}\underbrace{\begin{bmatrix}\mathbf{D}(\hat{\mathbf{u}}_k^*) & -\mathbf{D}(\check{\mathbf{u}}_k)\\ \mathbf{D}(\check{\mathbf{u}}_k^*) & \mathbf{D}(\hat{\mathbf{u}}_k)\end{bmatrix}}_{:=\mathcal{D}(\tilde{\mathbf{u}}_k)}\Psi=\mathbf{0}^T. \tag{9}$$

Stacking (9) for each $\tilde{\mathbf{u}}_k$ with $k\in[1,2J-2K]$, we obtain

$$\left[\mathbf{h}_1^T,\mathbf{h}_2^{\mathcal{H}}\right]\underbrace{\mathcal{F}[\mathcal{D}(\tilde{\mathbf{u}}_1)\Psi,\ldots,\mathcal{D}(\tilde{\mathbf{u}}_{2J-2K})\Psi]}_{:=\mathbf{Q}}=\mathbf{0}^T \tag{10}$$

from which one can solve for $[\mathbf{h}_1^T,\mathbf{h}_2^{\mathcal{H}}]^T$. Uniqueness of the solution will be addressed in the next section.

In the presence of white noise with variance σ_w^2, we replace $\mathbf{X}_N\mathbf{X}_N^{\mathcal{H}}$ in (6) by the received data covariance matrix $\mathbf{R}_{\check{\mathbf{y}}\check{\mathbf{y}}}:=E\{\check{\mathbf{y}}(n)\check{\mathbf{y}}^{\mathcal{H}}(n)\}$, whose SVD has the following form:

$$\mathbf{R}_{\check{\mathbf{y}}\check{\mathbf{y}}}=\begin{bmatrix}\mathbf{U}_x & \tilde{\mathbf{U}}_w\end{bmatrix}\begin{bmatrix}\Sigma_x & 0\\ 0 & \Sigma_w\end{bmatrix}\begin{bmatrix}\mathbf{U}_x^{\mathcal{H}}\\ \tilde{\mathbf{U}}_w^{\mathcal{H}}\end{bmatrix} \tag{11}$$

where $\Sigma_x=\text{diag}(\sigma_1^2,\ldots,\sigma_{2K}^2)$, $\Sigma_w=\text{diag}(\sigma_w^2,\ldots,\sigma_w^2)$ with $\sigma_1^2\geq\cdots\geq\sigma_{2K}^2>\sigma_w^2$. Since $\mathbf{U}_x$ and $\tilde{\mathbf{U}}_w$ span the same column space as $\mathbf{U}$ and $\tilde{\mathbf{U}}$, respectively, using $\tilde{\mathbf{U}}_w$ instead of $\tilde{\mathbf{U}}$ to build the system of (10) will yield exactly the same solution.

In practice, the ensemble correlation matrix is replaced by the sample average based on finite (say, N) blocks $\mathbf{R}_{\check{\mathbf{y}}\check{\mathbf{y}}}^{(N)}=(1/N)\sum_{n=0}^{N-1}\check{\mathbf{y}}(n)\check{\mathbf{y}}^{\mathcal{H}}(n)$, which converges in the mean square sense to the true correlation matrix $\mathbf{R}_{\check{\mathbf{y}}\check{\mathbf{y}}}$ since $\check{\mathbf{x}}(n)$ has finite moments. The latter guarantees the consistency of the proposed *blind* channel estimation algorithm that we summarize in the following steps.

[2]These conditions were originally adopted in [25] to guarantee symbol detectability in a single-antenna generalized multicarrier CDMA system without STC and more recently in a linearly precoded OFDM scheme for coding and maximum diversity gains [26].

Step 1) Collect the received data blocks $\check{\mathbf{y}}(n)$, and compute $\mathbf{R}_{\check{\mathbf{y}}\check{\mathbf{y}}}^{(N)} = (1/N)\sum_{n=0}^{N-1} \check{\mathbf{y}}(n)\check{\mathbf{y}}^{\mathcal{H}}(n)$.

Step 2) Determine the eigenvectors $\tilde{\mathbf{u}}_k, k = 1, \ldots, 2J-2K$ corresponding to the smallest $2J-2K$ eigenvalues of the matrix $\mathbf{R}_{\check{\mathbf{y}}\check{\mathbf{y}}}^{(N)}$.

Step 3) From these eigenvectors, estimate $[\mathbf{h}_1^T, \mathbf{h}_2^{\mathcal{H}}]^T$ as the nontrivial solution of (10) by determining the left eigenvector corresponding to the smallest eigenvalue of $\mathbf{Q}$, where $\mathbf{Q}$ is defined in (10).

An inherent problem to all subspace-based estimators is their relatively slow convergence with the number of data required. Note that our algorithm needs to collect at least $2K$ blocks, per a3). With a short record of data, a semi-blind implementation of the subspace based method can be devised by capitalizing on training sequences, which are anyways present for synchronization and rapid channel acquisition in practical systems. On the other hand, adaptive estimation of the channel correlation matrix enables tracking of slow channel variations. Proceeding as in [19], the *semi-blind* implementation for our algorithm is outlined next.

1) Obtain initial channel estimates $\hat{\mathbf{h}}_1$ and $\hat{\mathbf{h}}_2$ (and, thus, $\hat{\boldsymbol{\mathcal{H}}}$) through training (using, e.g., [15]), and estimate the autocorrelation matrix $\mathbf{R}_{\check{\mathbf{y}}\check{\mathbf{y}}}$ as $\mathbf{R}_{\check{\mathbf{y}}\check{\mathbf{y}}}^{(1)} = \sigma_s^2 \hat{\boldsymbol{\mathcal{H}}}\hat{\boldsymbol{\mathcal{H}}}^{\mathcal{H}}$, where $\sigma_s^2 := E\{s(n)s^*(n)\}$ denotes symbol energy.
2) Refine iteratively the autocorrelation matrix each time a new symbol block $\check{\mathbf{y}}(\bar{N})$ becomes available using a rectangular sliding window with window length W in (12), shown at the bottom of the page.
3) Perform the subspace algorithm based on $\mathbf{R}_{\check{\mathbf{y}}\check{\mathbf{y}}}^{(\bar{N}+1)}$.

When $\bar{N} > W$, $\mathbf{R}_{\check{\mathbf{y}}\check{\mathbf{y}}}^{(\bar{N}+1)}$ in (12) is actually the sample averaged estimate of $\mathbf{R}_{\check{\mathbf{y}}\check{\mathbf{y}}}$ based on the most recent W blocks and no longer depends on the training-based estimate $\mathbf{R}_{\check{\mathbf{y}}\check{\mathbf{y}}}^{(1)}$. Training helps only when the number of received blocks is not large enough, i.e., when $\bar{N} \leq W$.

The complexity of all subspace-based methods comes mainly from the eigendecomposition in (6), which is $\mathcal{O}((2J)^3)$ in our case. To reduce the complexity, however, it is possible to compute adaptively the left eigenvectors required by our subspace-based channel estimator using the subspace tracking approaches of [4] and [28]. Derivation, analysis, and on-line implementation of such estimators go beyond the scope of this paper.

IV. Channel Identifiability

The key question here is whether the solution of (10) is unique. We will explore channel identifiability in this section for two precoder choices: identical precoders and distinct precoders.

A. Identical Precoders

We first study the case where the same precoding matrix is used for both even and odd indexed blocks of symbols, i.e., $\boldsymbol{\Theta}_1 = \boldsymbol{\Theta}_2 = \boldsymbol{\Theta}$. It turns out that under this condition, the channels cannot be identified up to one scalar ambiguity, as stated in the following result.

Result 1: Suppose a1), a2), and a3) hold true; if $\boldsymbol{\Theta}_1 = \boldsymbol{\Theta}_2 = \boldsymbol{\Theta}$, the matrix $\mathbf{Q}$ defined in (10) loses row rank by two, and the solution of $[\mathbf{h}_3^T, \mathbf{h}_4^{\mathcal{H}}]\mathbf{Q} = \mathbf{0}^T$ belongs to a two-dimensional (2-D) vector space that is spanned by $\mathbf{h}_{12} = [\mathbf{h}_1^T, \mathbf{h}_2^{\mathcal{H}}]^T$ and $\mathbf{h}_{21} = [\mathbf{h}_2^T, -\mathbf{h}_1^{\mathcal{H}}]^T$.

Proof: Since $(\mathbf{h}_3, \mathbf{h}_4)$ is a pair of channels satisfying (10), it has the same signal subspace as $(\mathbf{h}_1, \mathbf{h}_2)$. There exists a full rank $2K \times 2K$ matrix $\mathbf{A}$ such that $\boldsymbol{\mathcal{H}}(\mathbf{h}_3, \mathbf{h}_4) = \boldsymbol{\mathcal{H}}(\mathbf{h}_1, \mathbf{h}_2)\mathbf{A}$, where $\boldsymbol{\mathcal{H}}(\mathbf{h}_1, \mathbf{h}_2)$ is the composite channel matrix $\boldsymbol{\mathcal{H}}$ defined in (4), and $\boldsymbol{\mathcal{H}}(\mathbf{h}_3, \mathbf{h}_4)$ is the counterpart of $\boldsymbol{\mathcal{H}}(\mathbf{h}_1, \mathbf{h}_2)$ with $(\mathbf{h}_3, \mathbf{h}_4)$ replacing $(\mathbf{h}_1, \mathbf{h}_2)$. Let us split $\mathbf{A}$ into four submatrices, each of size $K \times K$ that should satisfy the following equation:

$$\begin{bmatrix} \boldsymbol{\mathcal{D}}_3 & \boldsymbol{\mathcal{D}}_4 \\ \boldsymbol{\mathcal{D}}_4^* & -\boldsymbol{\mathcal{D}}_3^* \end{bmatrix} \boldsymbol{\Phi}_{12} = \begin{bmatrix} \boldsymbol{\mathcal{D}}_1 & \boldsymbol{\mathcal{D}}_2 \\ \boldsymbol{\mathcal{D}}_2^* & -\boldsymbol{\mathcal{D}}_1^* \end{bmatrix} \boldsymbol{\Phi}_{12} \begin{bmatrix} \mathbf{A}_1 & \mathbf{A}_2 \\ \mathbf{A}_3 & \mathbf{A}_4 \end{bmatrix}. \quad (13)$$

With $\boldsymbol{\mathcal{D}}$ defined in (4), we multiply (13) by $\boldsymbol{\mathcal{D}}^{\mathcal{H}}$ to obtain

$$(\boldsymbol{\mathcal{D}}_1^*\boldsymbol{\mathcal{D}}_3 + \boldsymbol{\mathcal{D}}_2\boldsymbol{\mathcal{D}}_4^*)\boldsymbol{\Theta} = (|\boldsymbol{\mathcal{D}}_1|^2 + |\boldsymbol{\mathcal{D}}_2|^2)\boldsymbol{\Theta}\mathbf{A}_1 \quad (14)$$

$$(\boldsymbol{\mathcal{D}}_2^*\boldsymbol{\mathcal{D}}_3 - \boldsymbol{\mathcal{D}}_1\boldsymbol{\mathcal{D}}_4^*)\boldsymbol{\Theta} = (|\boldsymbol{\mathcal{D}}_1|^2 + |\boldsymbol{\mathcal{D}}_2|^2)\boldsymbol{\Theta}\mathbf{A}_3 \quad (15)$$

$$(\boldsymbol{\mathcal{D}}_1\boldsymbol{\mathcal{D}}_3^* + \boldsymbol{\mathcal{D}}_2^*\boldsymbol{\mathcal{D}}_4)\boldsymbol{\Theta} = (|\boldsymbol{\mathcal{D}}_1|^2 + |\boldsymbol{\mathcal{D}}_2|^2)\boldsymbol{\Theta}\mathbf{A}_4 \quad (16)$$

$$(-\boldsymbol{\mathcal{D}}_2\boldsymbol{\mathcal{D}}_3^* + \boldsymbol{\mathcal{D}}_1^*\boldsymbol{\mathcal{D}}_4)\boldsymbol{\Theta} = (|\boldsymbol{\mathcal{D}}_1|^2 + |\boldsymbol{\mathcal{D}}_2|^2)\boldsymbol{\Theta}\mathbf{A}_2. \quad (17)$$

Let $D_i(u), i \in [1,2,3,4]$ denote the uth diagonal entry of $\boldsymbol{\mathcal{D}}_i$ and $\boldsymbol{\theta}_u^T$ the uth row of $\boldsymbol{\Theta}$. Since $|D_1(u)|^2 + |D_2(u)|^2$ has at most L zeros for $u = 1, \ldots, J$, we can find from (14) at least $J - L \geq K + 1$ equations satisfying

$$\frac{D_1^*(u)D_3(u) + D_2(u)D_4^*(u)}{|D_1(u)|^2 + |D_2(u)|^2}\boldsymbol{\theta}_u^T = \boldsymbol{\theta}_u^T\mathbf{A}_1. \quad (18)$$

At this point, we will make use of the following lemma.

Lemma 1: If for any $K \times K$ matrix $\mathbf{A}'$, there exist at least $K+1$ rows of $\boldsymbol{\Theta}$ [where $\boldsymbol{\Theta}$ satisfies a2)] such that $\lambda_k\boldsymbol{\theta}_k^T = \boldsymbol{\theta}_k^T\mathbf{A}'$, then $\mathbf{A}' = \lambda\mathbf{I}_K$ for some λ.

The proof of Lemma 1 can be deduced from [8, App. 2], where $\mathbf{A}'$ was unnecessarily assumed to have full rank. Applying Lemma 1 to (18), we obtain $\mathbf{A}_1 = \lambda_1\mathbf{I}_K$. Similarly, we obtain from (15)–(17) that $\mathbf{A}_3 = \lambda_3\mathbf{I}_K, \mathbf{A}_4 = \lambda_4\mathbf{I}_K$, and $\mathbf{A}_2 = \lambda_2\mathbf{I}_K$. Substituting $\mathbf{A}_1$ and $\mathbf{A}_4$ into (14) and (16), we obtain

$$(\boldsymbol{\mathcal{D}}_1^*\boldsymbol{\mathcal{D}}_3 + \boldsymbol{\mathcal{D}}_2\boldsymbol{\mathcal{D}}_4^*) = \lambda_1(|\boldsymbol{\mathcal{D}}_1|^2 + |\boldsymbol{\mathcal{D}}_2|^2) \quad (19)$$

$$(\boldsymbol{\mathcal{D}}_1\boldsymbol{\mathcal{D}}_3^* + \boldsymbol{\mathcal{D}}_2^*\boldsymbol{\mathcal{D}}_4) = \lambda_4(|\boldsymbol{\mathcal{D}}_1|^2 + |\boldsymbol{\mathcal{D}}_2|^2) \quad (20)$$

$$\mathbf{R}_{\check{\mathbf{y}}\check{\mathbf{y}}}^{(\bar{N}+1)} = \begin{cases} \mathbf{R}_{\check{\mathbf{y}}\check{\mathbf{y}}}^{(\bar{N})} + \dfrac{1}{W}\left[\check{\mathbf{y}}(\bar{N})\check{\mathbf{y}}^{\mathcal{H}}(\bar{N}) - \mathbf{R}_{\check{\mathbf{y}}\check{\mathbf{y}}}^{(1)}\right], & \bar{N} \leq W \\ \mathbf{R}_{\check{\mathbf{y}}\check{\mathbf{y}}}^{(\bar{N})} + \dfrac{1}{W}[\check{\mathbf{y}}(\bar{N})\check{\mathbf{y}}^{\mathcal{H}}(\bar{N}) - \check{\mathbf{y}}(\bar{N}-W)\check{\mathbf{y}}^{\mathcal{H}}(\bar{N}-W)], & \bar{N} > W. \end{cases} \quad (12)$$

which implies that $\lambda_4 = \lambda_1^*$. Similarly, from (15) and (17), we obtain $\lambda_3 = -\lambda_2^*$. Therefore, (13) can be re-expressed as

$$\begin{bmatrix} \mathcal{D}_3 & \mathcal{D}_4 \\ \mathcal{D}_4^* & -\mathcal{D}_3^* \end{bmatrix} \mathbf{\Phi}_{12} = \begin{bmatrix} \mathcal{D}_1\lambda_1 - \mathcal{D}_2\lambda_2^* & \mathcal{D}_1\lambda_2 + \mathcal{D}_2\lambda_1^* \\ \mathcal{D}_2^*\lambda_1 + \mathcal{D}_1^*\lambda_2^* & \mathcal{D}_2^*\lambda_2 - \mathcal{D}_1^*\lambda_1^* \end{bmatrix} \mathbf{\Phi}_{12}. \tag{21}$$

We obtain from (21)

$$\begin{bmatrix} \mathcal{D}_3 \\ \mathcal{D}_4^* \end{bmatrix} = \lambda_1 \begin{bmatrix} \mathcal{D}_1 \\ \mathcal{D}_2^* \end{bmatrix} - \lambda_2^* \begin{bmatrix} \mathcal{D}_2 \\ -\mathcal{D}_1^* \end{bmatrix}. \tag{22}$$

Therefore, both $\mathbf{h}_{12}{:=}[\mathbf{h}_1^T, \mathbf{h}_2^{\mathcal{H}}]^T$ and $\mathbf{h}_{21}{:=}[\mathbf{h}_2^T, -\mathbf{h}_1^{\mathcal{H}}]^T$, as well as their linear combinations, satisfy (10). Note that $\mathbf{h}_{12}$ and $\mathbf{h}_{21}$ are orthogonal. $\mathbf{h}_{12}^{\mathcal{H}}\mathbf{h}_{21} = 0$ and any vector satisfying (10) can be expressed as a linear combination of $\mathbf{h}_{12}$ and $\mathbf{h}_{21}$, as in (22); thus, the matrix $\mathbf{Q}$ has a left null space of dimension two, which indicates that it loses row rank by two, and the channel estimator of (10) does not necessarily yield the desired impulse response vectors uniquely. This fact is due to the inherent symmetry between antenna pairs and can be easily understood by noting that

$$\begin{bmatrix} \mathcal{D}_1 & \mathcal{D}_2 \\ \mathcal{D}_2^* & -\mathcal{D}_1^* \end{bmatrix} \begin{bmatrix} \tilde{\mathbf{s}}(2n) \\ \tilde{\mathbf{s}}(2n+1) \end{bmatrix} = \begin{bmatrix} \mathcal{D}_2 & -\mathcal{D}_1 \\ -\mathcal{D}_1^* & -\mathcal{D}_2^* \end{bmatrix} \begin{bmatrix} \tilde{\mathbf{s}}(2n+1) \\ -\tilde{\mathbf{s}}(2n) \end{bmatrix}.$$

For symmetric constellations, it is impossible to determine from the received blocks only whether the transmitter has sent $\tilde{\mathbf{s}}(2n)$ and $\tilde{\mathbf{s}}(2n+1)$ on channels $\mathbf{h}_1$ and $\mathbf{h}_2$ or $\tilde{\mathbf{s}}(2n+1)$ and $-\tilde{\mathbf{s}}(2n)$ on channels $\mathbf{h}_2$ and $-\mathbf{h}_1$, respectively (see also [17] for a related observation). ■

We have shown that the matrix $\mathbf{Q}$ in (10) loses row rank by two. Hence, it is natural to seek channel estimators based on the two available eigenvectors rather than one. However, these two eigenvectors are closely related, as we assert next.

Result 2: Denote the two eigenvectors corresponding to the left null space of $\mathbf{Q}$ by $\mathbf{h}_{34} = [\mathbf{h}_3^T, \mathbf{h}_4^{\mathcal{H}}]^T$ and $\mathbf{h}_{56} = [\mathbf{h}_5^T, \mathbf{h}_6^{\mathcal{H}}]^T$. Define the vector $\mathbf{h}_{43}{:=}[\mathbf{h}_4^T, -\mathbf{h}_3^{\mathcal{H}}]^T$. There exists a constant ϕ, such that

$$\mathbf{h}_{56} = e^{j\phi}\mathbf{h}_{43}. \tag{23}$$

Proof: Based on Result 1 and (22), we express $\mathbf{h}_{34}$ and $\mathbf{h}_{56}$ as

$$\begin{aligned} \mathbf{h}_{34} &= \lambda_1\mathbf{h}_{12} - \lambda_2^*\mathbf{h}_{21} \\ \mathbf{h}_{56} &= \gamma_1\mathbf{h}_{12} - \gamma_2^*\mathbf{h}_{21}. \end{aligned} \tag{24}$$

Because $\mathbf{h}_{12}$ and $\mathbf{h}_{21}$ are orthogonal basis vectors with equal norm, it follows that $|\lambda_1|^2+|\lambda_2|^2 = |\gamma_1|^2+|\gamma_2|^2$ since $\mathbf{h}_{34}$ and $\mathbf{h}_{56}$ are unit norm eigenvectors. Moreover, because $\mathbf{h}_{34}$ and $\mathbf{h}_{56}$ are orthogonal, we have $\lambda_1^*\gamma_1 + \lambda_2\gamma_2^* = 0$.

Letting $|\lambda_1|^2 + |\lambda_2|^2 = a^2$, the vectors $[\lambda_1, -\lambda_2^*]^T/a$ and $[\gamma_1, -\gamma_2^*]^T/a$ form an orthonormal basis for the 2-D space $\mathbb{C}^2$. Because the vectors $[\lambda_1, -\lambda_2^*]^T/a$ and $[\lambda_2, \lambda_1^*]^T/a$ also form an orthonormal basis for $\mathbb{C}^2$, we deduce that the vectors $[\gamma_1, -\gamma_2^*]^T$ and $[\lambda_2, \lambda_1^*]^T$ are parallel. Since they also have the same norm, there exists a ϕ such that

$$[\gamma_1, -\gamma_2^*]^T = e^{j\phi}[\lambda_2, \lambda_1^*]^T. \tag{25}$$

Based on (24) and (25), we obtain $\mathbf{h}_{43} = \lambda_2\mathbf{h}_{12} + \lambda_1^*\mathbf{h}_{21}$ and arrive at (23). ■

Result 2 shows that $\mathbf{h}_{56}$ can be constructed from $\mathbf{h}_{34}$ using (23). Thus, the second eigenvector does not provide additional information, and the channels cannot be identified up to one scalar ambiguity by using two eigenvectors. In the noisy case, where $\mathbf{h}_{34}$ and $\mathbf{h}_{56}$ are replaced by the corresponding estimates $\hat{\mathbf{h}}_{34}$ and $\hat{\mathbf{h}}_{56}$, (23) still holds true: $\hat{\mathbf{h}}_{56} = e^{j\phi}\hat{\mathbf{h}}_{43}$, as we prove in Appendix A. Hence, the error vectors $\hat{\mathbf{h}}_{34} - \mathbf{h}_{34}$ and $\hat{\mathbf{h}}_{56} - \mathbf{h}_{56}$ are correlated, and we cannot decrease the channel estimation error through averaging, e.g., by obtaining the estimate of $\mathbf{h}_{43}$ via $(\hat{\mathbf{h}}_{43} - e^{-j\phi}\hat{\mathbf{h}}_{56})/2$, even when ϕ is assumed to be known. The second eigenvector is basically redundant and can thus be ignored.

Based on Results 1 and 2, we summarize our findings for identical precoders in the following.

Theorem 1: Suppose a1), a2), and a3) hold true; if $\mathbf{\Theta}_1 = \mathbf{\Theta}_2 = \mathbf{\Theta}$, then the underlying channels can be identified from $[\mathbf{h}_3^T, \mathbf{h}_4^{\mathcal{H}}]\mathbf{Q} = \mathbf{0}^T$ up to two scalar ambiguities as

$$\begin{bmatrix} \mathbf{h}_1 \\ \mathbf{h}_2 \end{bmatrix} = \begin{bmatrix} \alpha_1\mathbf{I} & \alpha_2^*\mathbf{I} \\ -\alpha_2\mathbf{I} & \alpha_1^*\mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{h}_3 \\ \mathbf{h}_4 \end{bmatrix}. \tag{26}$$

Proof: From Result 2, we ignore $\mathbf{h}_{56}$ and pick up only $\mathbf{h}_{34}$ from (24). Based on Result 1 and (22), the true channel can be expressed from $\mathbf{h}_{34}$ as

$$\begin{bmatrix} \mathbf{h}_1 \\ \mathbf{h}_2 \end{bmatrix} = \frac{1}{|\lambda_1|^2 + |\lambda_2|^2} \begin{bmatrix} \lambda_1^*\mathbf{I} & \lambda_2^*\mathbf{I} \\ -\lambda_2\mathbf{I} & \lambda_1\mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{h}_3 \\ \mathbf{h}_4 \end{bmatrix}. \tag{27}$$

For brevity, we define two scalars $\alpha_1{:=}\lambda_1^*/(|\lambda_1|^2 + |\lambda_2|^2)$ and $\alpha_2{:=}\lambda_2/(|\lambda_1|^2 + |\lambda_2|^2)$ and obtain (26) from (27).

B. Distinct Precoders

To break the symmetry and identify the underlying channels uniquely, a preweighting approach that loads different power over different antennas was explored but not advocated in [17] because of the loss it incurs in bit error rate (BER) performance. Here, we break this symmetry by introducing distinct linear precoders $\mathbf{\Theta}_1$ and $\mathbf{\Theta}_2$ that can be designed to have equal powers. Channel identifiability can then be guaranteed up to one scalar ambiguity, as the following theorem asserts.

Theorem 2: Suppose a1), a2), and a3) hold true; let $\bar{\mathbf{D}}$ denote any diagonal matrix with unit amplitude diagonal entries, and let $\bar{\mathbf{\Theta}}_1$, $\bar{\mathbf{\Theta}}_2$ be formed by any $J - L$ rows of $\mathbf{\Theta}_1$, $\mathbf{\Theta}_2$, respectively. If $\bar{\mathbf{\Theta}}_1$ and $\bar{\mathbf{\Theta}}_2$ satisfy $\bar{\mathbf{D}}\bar{\mathbf{\Theta}}_1 \notin \mathcal{R}(\bar{\mathbf{\Theta}}_2)$, the solution of $[\mathbf{h}_3^T, \mathbf{h}_4^{\mathcal{H}}]\mathbf{Q} = \mathbf{0}^T$ is unique up to a constant, and thus, channel identifiability within one scalar is guaranteed:

$$\begin{bmatrix} \mathbf{h}_1 \\ \mathbf{h}_2 \end{bmatrix} = \begin{bmatrix} \alpha\mathbf{I} & \mathbf{0} \\ \mathbf{0} & \alpha^*\mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{h}_3 \\ \mathbf{h}_4 \end{bmatrix}. \tag{28}$$

Proof: With distinct $\mathbf{\Theta}_1$ and $\mathbf{\Theta}_2$, (14)–(17) turn into

$$(\mathcal{D}_1^*\mathcal{D}_3 + \mathcal{D}_2\mathcal{D}_4^*)\mathbf{\Theta}_1 = (|\mathcal{D}_1|^2 + |\mathcal{D}_2|^2)\mathbf{\Theta}_1\mathbf{A}_1 \tag{29}$$

$$(\mathcal{D}_2^*\mathcal{D}_3 - \mathcal{D}_1\mathcal{D}_4^*)\mathbf{\Theta}_1 = (|\mathcal{D}_1|^2 + |\mathcal{D}_2|^2)\mathbf{\Theta}_2\mathbf{A}_3 \tag{30}$$

$$(\mathcal{D}_1\mathcal{D}_3^* + \mathcal{D}_2^*\mathcal{D}_4)\mathbf{\Theta}_2 = (|\mathcal{D}_1|^2 + |\mathcal{D}_2|^2)\mathbf{\Theta}_2\mathbf{A}_4 \tag{31}$$

$$(-\mathcal{D}_2\mathcal{D}_3^* + \mathcal{D}_1^*\mathcal{D}_4)\mathbf{\Theta}_2 = (|\mathcal{D}_1|^2 + |\mathcal{D}_2|^2)\mathbf{\Theta}_1\mathbf{A}_2. \tag{32}$$

From (29) and (31), we obtain $\mathbf{A}_1 = \lambda_1\mathbf{I}$, $\mathbf{A}_4 = \lambda_4\mathbf{I}$, and $\lambda_4 = \lambda_1^*$ following the proof of Lemma 1.

The right-hand sides of (30) and (32) have at least $J - L$ nonzero rows each. Let us collect only $J - L$ nonzero rows

with the row indices denoted as $u_1, u_2, \ldots, u_{J-L}$. Define the $(J-L)\times(J-L)$ matrix $\bar{\mathbf{D}}'$ as the diagonal matrix with $[\bar{\mathbf{D}}']_{i,i} = [D_2^*(u_i)D_3(u_i) - D_1(u_i)D_4^*(u_i)]/[|D_1(u_i)|^2 + |D_2(u_i)|^2]$, and let $\bar{\mathbf{\Theta}}_1'$ and $\bar{\mathbf{\Theta}}_2'$ be formed by the $u_1, u_2, \ldots, u_{J-L}$ rows of $\mathbf{\Theta}_1$ and $\mathbf{\Theta}_2$, respectively. We then obtain from (30) and (32)

$$\bar{\mathbf{D}}'\bar{\mathbf{\Theta}}_1' = \bar{\mathbf{\Theta}}_2'\mathbf{A}_3, \quad -(\bar{\mathbf{D}}')^*\bar{\mathbf{\Theta}}_2' = \bar{\mathbf{\Theta}}_1'\mathbf{A}_2. \tag{33}$$

From (33), we deduce that $-|\bar{\mathbf{D}}'|^2\mathbf{\Theta}_2 = \mathbf{\Theta}_2\mathbf{A}_3\mathbf{A}_2$. Applying Lemma 1, we find that $\mathbf{A}_3\mathbf{A}_2 = -\lambda\mathbf{I}$ and $|\bar{\mathbf{D}}'|^2 = \lambda\mathbf{I}$. If $\lambda \neq 0$, then there exists a $(\bar{\mathbf{D}}'/\sqrt{\lambda}, \bar{\mathbf{\Theta}}_1', \bar{\mathbf{\Theta}}_2')$ triplet within the class of $(\bar{\mathbf{D}}, \bar{\mathbf{\Theta}}_1, \bar{\mathbf{\Theta}}_2)$ triplets specified by Theorem 2, which satisfies $\bar{\mathbf{D}}'\bar{\mathbf{\Theta}}_1' \in \mathcal{R}(\bar{\mathbf{\Theta}}_2')$. Since this is impossible under the design constraints of Theorem 2, we arrive at $\lambda = 0$, which allows only $[\mathcal{D}_3 \ \mathcal{D}_4^*] = \lambda_1[\mathcal{D}_1 \ \mathcal{D}_2^*]$. Setting $\alpha = \lambda_1^*/|\lambda_1|^2$, we obtain (28). ■

If $\mathbf{\Theta}_1 = \mathbf{\Theta}_2$ as in Section IV-A, the scalar λ can be any non-negative number, and a simple class of $\mathbf{A}_3, \mathbf{A}_2$ matrices is $\mathbf{A}_3 = \sqrt{\lambda}e^{j\varphi}\mathbf{I}$, $\mathbf{A}_2 = -\sqrt{\lambda}e^{-j\varphi}\mathbf{I}$, where φ is arbitrary. Again, channel identifiability up to one scalar is not guaranteed, in agreement with Theorem 1.

To select the appropriate $\mathbf{\Theta}_1$ and $\mathbf{\Theta}_2$ precoders, we can, for instance, construct them as Vandermonde matrices with distinct generators $[\rho_{\mu,1}, \ldots, \rho_{\mu,J}]_{\mu=1,2}$

$$\mathbf{\Theta}_\mu = \begin{bmatrix} 1 & \rho_{\mu,1}^{-1} & \cdots & \rho_{\mu,1}^{-(K-1)} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & \rho_{\mu,J}^{-1} & \cdots & \rho_{\mu,J}^{-(K-1)} \end{bmatrix} \tag{34}$$

and then check whether $\bar{\mathbf{D}}\bar{\mathbf{\Theta}}_1 \notin \mathcal{R}(\bar{\mathbf{\Theta}}_2)$ by simulation offline. Note also that Vandermonde matrices like (34) with distinct generators satisfy a2).

C. Minimally Redundant Precoders

Note that both Theorems 1 and 2 adopt a1), which requires selecting block sizes to assure redundancy of $J - K \geq L + 1$. Instead of a1), we will now focus on minimally redundant precoders with **a1')** $J = K + \tilde{L}$, where $\tilde{L} \in [1, L]$.

Suppose that the channels $\mathbf{h}_1$ and $\mathbf{h}_2$ have Z common zeros that are located on the FFT grid; thus, Z out of the J entries $\{|D_1(u)|^2 + |D_2(u)|^2\}_{u=1}^J$ will be zero. If $Z < \tilde{L}$, we can find $J - Z \geq K + 1$ equations like (18). Mimicking the steps followed to prove Theorems 1 and 2, we obtain, respectively, the following identifiability results.

Theorem 3: Suppose a1'), a2), and a3) hold true; if $\mathbf{\Theta}_1 = \mathbf{\Theta}_2 = \mathbf{\Theta}$, channel identifiability is guaranteed up to two scalars, as in (26), for those channel pairs having $Z < \tilde{L}$ common zeros located on the FFT grid.

Theorem 4: Suppose a1'), a2), and a3) hold true; let $\bar{\mathbf{D}}$ denote any diagonal matrix with unit amplitude diagonal entries, and let $\bar{\mathbf{\Theta}}_1$, $\bar{\mathbf{\Theta}}_2$ be formed by any $K+1$ rows of $\mathbf{\Theta}_1$, $\mathbf{\Theta}_2$, respectively. If $\bar{\mathbf{\Theta}}_1$ and $\bar{\mathbf{\Theta}}_2$ satisfy $\bar{\mathbf{D}}\bar{\mathbf{\Theta}}_1 \notin \mathcal{R}(\bar{\mathbf{\Theta}}_2)$, then channel identifiability is guaranteed within one scalar as in (28) for those channel pairs with $Z < \tilde{L}$ common zeros located on the FFT grid.

Under a1'), Theorems 3 and 4 show that channel identifiability now depends on the channel zero locations and, thus, on each particular channel realization of the underlying random fading channels. Theorems 3 and 4 imply that the chance of losing identifiability is at most the probability that two random channels $\mathbf{h}_1$ and $\mathbf{h}_2$ share $Z \geq \tilde{L}$ common zeros on the FFT grid. This probability is very low for uncorrelated (or even weakly correlated) channels. Therefore, the proposed channel estimator can be applied in practice, even under a1'), if the two transmit antennas are sufficiently separated.

Because a channel of order L has exactly L zeros, the case with $J = K + L$ and $Z = L$ turns out to be a special case that is not covered by Theorems 3 and 4. We will check this special case in the following. From (13), we infer that $\mathcal{D}_3\mathbf{\Theta}_1 = \mathcal{D}_1\mathbf{\Theta}_1\mathbf{A}_1 + \mathcal{D}_2\mathbf{\Theta}_2\mathbf{A}_3$, which implies that if $\mathbf{h}_1$, $\mathbf{h}_2$ have the same $Z = L$ roots located on the FFT grid indexed by $u_1, u_2, \ldots, u_L$, then we have $D_3(u_l) = 0$ (similarly $D_4(u_l) = 0$) since $D_1(u_l) = D_2(u_l) = 0$ for $l = 1, \ldots, L$. Thus, $\mathbf{h}_1, \mathbf{h}_2, \mathbf{h}_3$ and $\mathbf{h}_4$ have the same L roots and are all proportional to each other, which assures that channel identifiability is also guaranteed. However, it is shown in Appendix B that we need to resolve two scalar ambiguities of the form $\mathbf{h}_3 = \alpha_1\mathbf{h}_1$ and $\mathbf{h}_4 = \alpha_2\mathbf{h}_2$ for both identical and distinct precoders, which is different from what we obtained in Theorems 1 and 2.

Therefore, from the received data only, multiple channels can be estimated simultaneously up to one or two scalar ambiguities with linearly block precoded ST-OFDM transmissions. To enable channel equalization, we show next how to resolve these scalar ambiguities by inserting known symbols in the transmitted sequence.

V. Resolving Scalar Ambiguities

To resolve the scalar ambiguities inherent to all blind channel estimators, known symbols are needed in the transmitted sequence. In our block precoded multicarrier ST setting, the known symbols can be inserted either before or after precoding. We term the corresponding symbols as pre- and post-precoding pilots, respectively. Specifically, pre-precoding pilots correspond to known entries in the blocks $\mathbf{s}(n)$ of (1), whereas post-precoding pilots amount to inserting known symbols into the blocks $\tilde{\mathbf{s}}(n)$ of (1). Note that each post-precoding pilot rides on a single subcarrier that is often referred to as a pilot tone in the multicarrier literature [21]. On the other hand, pre-precoding pilots are inserted among the information symbols so that all subcarriers are shared by both pilots and information bearing symbols. In a single carrier system, pre-precoding pilots are often referred to as pilot symbols [3].

Before detailing how we resolve scalar ambiguities, let us first look at how these scalars impact channel equalization. From (26), we obtain the estimated channels as

$$\begin{aligned} \mathcal{D}_{34} &= \begin{bmatrix} \mathcal{D}_3 & \mathcal{D}_4 \\ \mathcal{D}_4^* & -\mathcal{D}_3^* \end{bmatrix} \\ &= \begin{bmatrix} \mathcal{D}_1 & \mathcal{D}_2 \\ \mathcal{D}_2^* & -\mathcal{D}_1^* \end{bmatrix} \begin{bmatrix} \alpha_1^*\mathbf{I}_J & \alpha_2\mathbf{I}_J \\ -\alpha_2^*\mathbf{I}_J & \alpha_1\mathbf{I}_J \end{bmatrix} \frac{1}{|\alpha_1|^2 + |\alpha_2|^2}. \end{aligned} \tag{35}$$

Furthermore, we verify that $\bar{\mathcal{D}}_{34} := \mathcal{D}_3^*\mathcal{D}_3 + \mathcal{D}_4^*\mathcal{D}_4$ equals

$$\bar{\mathcal{D}}_{34} = \frac{(\mathcal{D}_1^*\mathcal{D}_1 + \mathcal{D}_2^*\mathcal{D}_2)}{|\alpha_1|^2 + |\alpha_2|^2} = \frac{\bar{\mathcal{D}}_{12}}{|\alpha_1|^2 + |\alpha_2|^2} \tag{36}$$

where $\bar{\boldsymbol{\mathcal{D}}}_{12}$ is defined after (5). We pursue the scalar determination with identical precoders first. Multiplying $\check{\mathbf{y}}(n)$ by $\boldsymbol{\mathcal{D}}_{34}^{\mathcal{H}}$ yields $\check{\mathbf{z}}(n):=\boldsymbol{\mathcal{D}}_{34}^{\mathcal{H}}\check{\mathbf{y}}(n)$ as [cf. (5)]

$$\begin{aligned}\check{\mathbf{z}}(n) &= \begin{bmatrix}\mathbf{z}(2n)\\ \mathbf{z}(2n+1)\end{bmatrix}\\ &= \begin{bmatrix}\alpha_1\mathbf{I}_J & -\alpha_2\mathbf{I}_J\\ \alpha_2^*\mathbf{I}_J & \alpha_1^*\mathbf{I}_J\end{bmatrix}\begin{bmatrix}\bar{\boldsymbol{\mathcal{D}}}_{12}\boldsymbol{\Theta}\mathbf{s}(2n)\\ \bar{\boldsymbol{\mathcal{D}}}_{12}\boldsymbol{\Theta}\mathbf{s}(2n+1)\end{bmatrix}\frac{1}{|\alpha_1|^2+|\alpha_2|^2}\\ &= \begin{bmatrix}\alpha_1\mathbf{I}_J & -\alpha_2\mathbf{I}_J\\ \alpha_2^*\mathbf{I}_J & \alpha_1^*\mathbf{I}_J\end{bmatrix}\begin{bmatrix}\bar{\boldsymbol{\mathcal{D}}}_{34}\boldsymbol{\Theta}\mathbf{s}(2n)\\ \bar{\boldsymbol{\mathcal{D}}}_{34}\boldsymbol{\Theta}\mathbf{s}(2n+1)\end{bmatrix}. \end{aligned} \tag{37}$$

Based on $\check{\mathbf{z}}(n)$ in (37), the scalar ambiguities can now be resolved by pilots. For clarity and brevity of presentation, we omit the noise and focus on using only one pair of pilots (p_1, p_2) with p_1 and p_2 placed on the same position in two successive symbol blocks. Extension to the noisy case can be addressed similarly. With multiple pilots, the resolved scalars can be averaged to improve estimation accuracy. We will first discuss post-precoding pilots, where separation of the subcarriers for pilots and data leads to a simple derivation.

A. Post-Precoding Pilots

Suppose that only the first subcarrier $\rho_0:=e^{j0}$ is chosen as a pilot tone. Then, the overall precoder can be expressed as

$$\boldsymbol{\Theta} = \begin{bmatrix}1 & \mathbf{0}^T\\ \mathbf{0} & \boldsymbol{\Theta}'\end{bmatrix} \tag{38}$$

and (1) can then be rewritten as

$$\tilde{\mathbf{s}}(2n) = \begin{bmatrix}p(2n)\\ \boldsymbol{\Theta}'\mathbf{s}'(2n)\end{bmatrix}, \quad \tilde{\mathbf{s}}(2n+1) = \begin{bmatrix}p(2n+1)\\ \boldsymbol{\Theta}'\mathbf{s}'(2n+1)\end{bmatrix} \tag{39}$$

where $p(2n)$ and $p(2n+1)$ are known symbols, whereas $\mathbf{s}'(2n)$ and $\mathbf{s}'(2n+1)$ are successive data blocks that are not known to the receiver.

Define the positive constant c^2 to be the first diagonal entry of $\bar{\boldsymbol{\mathcal{D}}}_{34}$, i.e., $c^2:=[\bar{\boldsymbol{\mathcal{D}}}_{34}]_{1,1} = (|H_1(0)|^2 + |H_2(0)|^2)/(|\alpha_1|^2 + |\alpha_2|^2)$. Pick up two known symbols (p_1, p_2) from the $2n$th and $(2n+1)$st blocks from (39), i.e., $p_1 = p(2n)$ and $p_2 = p(2n+1)$, and define $z_1 = [\mathbf{z}(2n)]_1$ and $z_2 = [\mathbf{z}(2n+1)]_1$; we then obtain, from (37)

$$\begin{bmatrix}z_1\\ z_2\end{bmatrix} = c^2\begin{bmatrix}\alpha_1 & -\alpha_2\\ \alpha_2^* & \alpha_1^*\end{bmatrix}\begin{bmatrix}p_1\\ p_2\end{bmatrix}. \tag{40}$$

Conjugating the second row of (40), we obtain

$$\begin{bmatrix}z_1\\ z_2^*\end{bmatrix} = c^2\begin{bmatrix}p_1 & -p_2\\ p_2^* & p_1^*\end{bmatrix}\begin{bmatrix}\alpha_1\\ \alpha_2\end{bmatrix} \tag{41}$$

from which α_1 and α_2 can be found as

$$\begin{bmatrix}\alpha_1\\ \alpha_2\end{bmatrix} = \frac{1}{c^2(|p_1|^2+|p_2|^2)}\begin{bmatrix}p_1^* & p_2\\ -p_2^* & p_1\end{bmatrix}\begin{bmatrix}z_1\\ z_2^*\end{bmatrix}. \tag{42}$$

With distinct precoders, setting $\alpha_2 = 0$ and $\alpha = \alpha_1$ in (42) leads to

$$\alpha = (p_1^* z_1 + p_2 z_2^*)/[c^2(|p_1|^2 + |p_2|^2)]. \tag{43}$$

Hence, the residual scalar ambiguities can be resolved with post-precoding pilots using (42) or (43), and the symbol blocks can be estimated from $\check{\mathbf{z}}(n)$ of (37) based on the knowledge of α_1 and α_2.

However, using post-precoding pilots requires that the channels $\mathbf{h}_1$ and $\mathbf{h}_2$ do not have common zeros (or deep fades) on ρ_0; otherwise, $c^2 = 0$ (or $c^2 \approx 0$) by definition, and the scalars are either nonidentifiable, or their estimation accuracy is severely affected by the additive noise, even with known symbols continuously transmitted on that subcarrier. To overcome this limitation, we need to allocate extra subcarriers in order to transmit additional pilots. For example, if the number of pilot subcarriers $N_p > L$, then the scalars are guaranteed to be recoverable since the channels can have at most L zeros. However, for a fixed number of total system subcarriers, fewer subcarriers will then be left for data transmission, and the system's spectral efficiency will suffer accordingly.

Another way to avoid deep fades on certain subcarriers could have been to hop several pilot subcarriers over multiple successive blocks. However, this approach is not feasible here since we need to fix the positions of pilot subcarriers in order to maintain the same precoder $\boldsymbol{\Theta}$ for subspace decomposition. Fixing the positions of pilot tones imposes a structure like (38), which, in turn, constrains the design of distinct precoders required by Theorem 2.

We show next that pre-precoding pilots are more flexible for block precoded ST transmissions.

B. Pre-Precoding Pilots

Because the known symbols are now inserted in the data stream before precoding, we need to equalize the channel and compensate for the precoding first, before resolving the residual scalar ambiguities. With identical precoders, a zero-forcing equalization can be applied to $\mathbf{z}(2n)$ and $\mathbf{z}(2n+1)$ in (37) by pre-multiplying with $(\bar{\boldsymbol{\mathcal{D}}}_{34}\boldsymbol{\Theta})^\dagger$, where † stands for matrix pseudo-inverse. Based on (37), the equalizer outputs $\grave{\mathbf{s}}(2n):=(\bar{\boldsymbol{\mathcal{D}}}_{34}\boldsymbol{\Theta})^\dagger\mathbf{z}(2n)$ and $\grave{\mathbf{s}}(2n+1):=(\bar{\boldsymbol{\mathcal{D}}}_{34}\boldsymbol{\Theta})^\dagger\mathbf{z}(2n+1)$ can be written as

$$\begin{bmatrix}\grave{\mathbf{s}}(2n)\\ \grave{\mathbf{s}}(2n+1)\end{bmatrix} = \begin{bmatrix}\alpha_1\mathbf{I}_J & -\alpha_2\mathbf{I}_J\\ \alpha_2^*\mathbf{I}_J & \alpha_1^*\mathbf{I}_J\end{bmatrix}\begin{bmatrix}\mathbf{s}(2n)\\ \mathbf{s}(2n+1)\end{bmatrix}. \tag{44}$$

Suppose that two pilot symbols p_1 and p_2 are placed inside two consecutive blocks $\mathbf{s}(2m)$ and $\mathbf{s}(2m+1)$ at position k. Define $\grave{s}_1 = [\grave{\mathbf{s}}(2m)]_k$ and $\grave{s}_2 = [\grave{\mathbf{s}}(2m+1)]_k$; we then obtain from (44) that

$$\begin{bmatrix}\grave{s}_1\\ \grave{s}_2\end{bmatrix} = \begin{bmatrix}\alpha_1 & -\alpha_2\\ \alpha_2^* & \alpha_1^*\end{bmatrix}\begin{bmatrix}p_1\\ p_2\end{bmatrix}. \tag{45}$$

Following the derivation of (42), one can easily solve for α_1 and α_2 as

$$\begin{bmatrix}\alpha_1\\ \alpha_2\end{bmatrix} = \frac{1}{|p_1|^2+|p_2|^2}\begin{bmatrix}p_1^* & p_2\\ -p_2^* & p_1\end{bmatrix}\begin{bmatrix}\grave{s}_1\\ \grave{s}_2^*\end{bmatrix}. \tag{46}$$

With α_1 and α_2 resolved, the true channels can then be found from (26). However, this step is not necessary since the symbol estimates can be obtained directly from (44) as

$$\begin{bmatrix}\hat{\mathbf{s}}(2n)\\ \hat{\mathbf{s}}(2n+1)\end{bmatrix} = \frac{1}{|\alpha_1|^2+|\alpha_2|^2}\begin{bmatrix}\alpha_1^*\mathbf{I}_J & \alpha_2\mathbf{I}_J\\ -\alpha_2^*\mathbf{I}_J & \alpha_1\mathbf{I}_J\end{bmatrix} \times \begin{bmatrix}\grave{\mathbf{s}}(2n)\\ \grave{\mathbf{s}}(2n+1)\end{bmatrix}.$$

With distinct precoders, we should equalize $\mathbf{z}(2n)$ by $(\bar{\boldsymbol{\mathcal{D}}}_{34}\boldsymbol{\Theta}_1)^\dagger$, and $\mathbf{z}(2n+1)$ by $(\bar{\boldsymbol{\mathcal{D}}}_{34}\boldsymbol{\Theta}_2)^\dagger$. Substituting $\alpha=\alpha_1$ and $\alpha_2=0$ in (46), the scalar α can be figured out as

$$\alpha=(p_1^*\grave{s}_1+p_2\grave{s}_2^*)/(|p_1|^2+|p_2|^2). \tag{47}$$

Since pilot symbols are up to the designer's choice, (p_1,p_2) can be selected with equal amplitudes $|p_1|=|p_2|$. Thus, (46) and (47) can be further simplified to

$$\begin{aligned}&\alpha_1=(\grave{s}_1/p_1+\grave{s}_2^*/p_2^*)/2,\quad \alpha_2=(-\grave{s}_1/p_2+\grave{s}_2^*/p_1^*)/2\\ &\alpha=(\grave{s}_1/p_1+\grave{s}_2^*/p_2^*)/2.\end{aligned} \tag{48}$$

Compared with post-precoding pilots, pre-precoding pilots have the following two advantages. First, pre-precoding pilots are insensitive to channel zero locations. Indeed, each pilot symbol is distributed by the precoder to all the subcarriers. This guarantees its recovery, regardless of the channel zero locations, because the channels have at most L nulls, and not all $J>L$ subcarriers can be significantly attenuated. Second, we have flexibility in selecting the positions of pre-precoding pilots. Within a data burst of N blocks $\{\check{\mathbf{s}}(n)\}_{n=0}^{N-1}$, the pilots can be placed at any block, and the position within each block can also be arbitrary. The positions of pre-precoding pilots do not affect the choice of precoders and, thus, the proposed subspace based channel estimator.

A concern regarding estimation accuracy of the scalars (α_1,α_2) with pre-precoding pilots is the possible noise enhancement when zero forcing equalizers are applied. Note, however, that the matrix $\bar{\boldsymbol{\mathcal{D}}}_{34}\boldsymbol{\Theta}$ is always of full rank by construction, and thus, the noise enhancement may not be severe. On the other hand, since the noise enhancement is related to the pilot position (say, the kth position) within the symbol block, multiple pairs of pilots with different positions can be employed, and then, the estimation error can be suppressed through averaging. Different from post-precoding pilots, pre-precoding pilots do not occupy a subset of subcarriers throughout a data burst. With a small number of known symbols inserted to determine the inherent scalar ambiguities, pre-precoding pilots offer higher spectral efficiency than post-precoding pilots in general.

VI. Distinct Versus Identical Precoders

As indicated by Theorems 1 and 2, with distinct precoders $\boldsymbol{\Theta}_1$ and $\boldsymbol{\Theta}_2$, the channels can be identified up to one scalar α instead of two scalars (α_1,α_2), which must be determined with identical precoders $\boldsymbol{\Theta}_1=\boldsymbol{\Theta}_2=\boldsymbol{\Theta}$. With one pair of known symbols (p_1,p_2), the residual scalar ambiguities can be resolved by (42) and (43) for post-precoding pilots or by (46) and (47) for pre-precoding pilots. Therefore, the advantage of distinct precoders over identical precoders is not clearly justified since two scalars are also not difficult to resolve for identical precoders as in (42) and (46). To reveal the advantage of distinct over identical precoders, we resort to performance analysis. Since pre-precoding pilots are favorable over post-precoding pilots, we next detail the noise analysis for the scalar ambiguity determination of (46) and (47). The same procedure can be applied to (42) and (43) as well.

In the presence of additive noise, we can rewrite (44) as

$$\begin{bmatrix}\grave{\mathbf{s}}(2n)\\ \grave{\mathbf{s}}(2n+1)\end{bmatrix}=\begin{bmatrix}\alpha_1\mathbf{I}_J & -\alpha_2\mathbf{I}_J\\ \alpha_2^*\mathbf{I}_J & \alpha_1^*\mathbf{I}_J\end{bmatrix}\begin{bmatrix}\mathbf{s}(2n)+\mathbf{w}(2n)\\ \mathbf{s}(2n+1)+\mathbf{w}(2n+1)\end{bmatrix} \tag{49}$$

where $\mathbf{w}(2n)$ and $\mathbf{w}(2n+1)$ denote the additive noise after the zero-forcing equalization by $(\bar{\boldsymbol{\mathcal{D}}}_{34}\boldsymbol{\Theta})^\dagger$ and can thus be related to $\check{\mathbf{w}}(n)$ in (3) by [cf., (37) and (44)]

$$\begin{bmatrix}\mathbf{w}(2n)\\ \mathbf{w}(2n+1)\end{bmatrix}=\begin{bmatrix}(\bar{\boldsymbol{\mathcal{D}}}_{34}\boldsymbol{\Theta})^\dagger & \mathbf{0}\\ \mathbf{0} & (\bar{\boldsymbol{\mathcal{D}}}_{34}\boldsymbol{\Theta})^\dagger\end{bmatrix}\boldsymbol{\mathcal{D}}_{34}^{\mathcal{H}}\check{\mathbf{w}}(n). \tag{50}$$

Equation (50) reveals that $\mathbf{w}(2n)$ and $\mathbf{w}(2n+1)$ are not dependent on the unknowns (α_1,α_2). With the definitions $w_1:=[\mathbf{w}(2m)]_k$ and $w_2:=[\mathbf{w}(2m+1)]_k$, the noisy version of (45) is

$$\begin{bmatrix}\grave{s}_1\\ \grave{s}_2\end{bmatrix}=\begin{bmatrix}\alpha_1 & -\alpha_2\\ \alpha_2^* & \alpha_1^*\end{bmatrix}\begin{bmatrix}p_1+w_1\\ p_2+w_2\end{bmatrix}. \tag{51}$$

Let us denote with $(\hat{\alpha}_1,\hat{\alpha}_2)$ the solution of (51), which contains the additive noise, and evaluate the estimation error relative to the true value of (α_1,α_2), which is the solution of (45) with no additive noise. Plugging the noisy $(\grave{s}_1,\grave{s}_2)$ in (51) into (46), we can solve for $(\hat{\alpha}_1,\hat{\alpha}_2)$ and establish the following relationship:

$$\begin{aligned}\begin{bmatrix}\hat{\alpha}_1\\ \hat{\alpha}_2\end{bmatrix}&=\begin{bmatrix}\alpha_1\\ \alpha_2\end{bmatrix}+\frac{1}{|p_1|^2+|p_2|^2}\\ &\quad\times\begin{bmatrix}p_1^* & p_2\\ -p_2^* & p_1\end{bmatrix}\begin{bmatrix}w_1 & -w_2\\ w_2^* & w_1^*\end{bmatrix}\begin{bmatrix}\alpha_1\\ \alpha_2\end{bmatrix}\\ &:=\begin{bmatrix}\alpha_1\\ \alpha_2\end{bmatrix}+\begin{bmatrix}\Delta\alpha_1\\ \Delta\alpha_2\end{bmatrix}.\end{aligned} \tag{52}$$

For simplicity, we assume that w_1 and w_2 are uncorrelated (or weakly correlated) and satisfy $E\{w_iw_j^*\}=\sigma_\epsilon^2\delta(i-j), E\{w_iw_j\}=0$ for $i,j\in[1,2]$. It can be easily verified that $\Delta\alpha_1$ and $\Delta\alpha_2$ have the same variance $\sigma_\epsilon^2(|\alpha_1|^2+|\alpha_2|^2)/(|p_1|^2+|p_2|^2)$. For distinct precoders, we set $\alpha_2=0$ and $\alpha=\alpha_1$ in (52) to obtain

$$\hat{\alpha}=\alpha+\Delta\alpha \tag{53}$$

where $\Delta\alpha$ has variance $\sigma_\epsilon^2|\alpha|^2/(|p_1|^2+|p_2|^2)$.

Replacing $(\mathbf{h}_3,\mathbf{h}_4)$ in (26) by its noisy version $(\hat{\mathbf{h}}_3,\hat{\mathbf{h}}_4)$, where $\hat{\mathbf{h}}_3=\mathbf{h}_3+\Delta\mathbf{h}_3$ and $\hat{\mathbf{h}}_4=\mathbf{h}_4+\Delta\mathbf{h}_4$, we obtain the estimated channels for identical precoders as

$$\begin{aligned}\begin{bmatrix}\hat{\mathbf{h}}_1\\ \hat{\mathbf{h}}_2\end{bmatrix}&=\begin{bmatrix}\mathbf{h}_1\\ \mathbf{h}_2\end{bmatrix}+\underbrace{\begin{bmatrix}\alpha_1\mathbf{I} & \alpha_2^*\mathbf{I}\\ -\alpha_2\mathbf{I} & \alpha_1^*\mathbf{I}\end{bmatrix}\begin{bmatrix}\Delta\mathbf{h}_3\\ \Delta\mathbf{h}_4\end{bmatrix}}_{:=\mathbf{e}_{1,i}}\\ &\quad+\underbrace{\begin{bmatrix}\Delta\alpha_1\mathbf{I} & \Delta\alpha_2^*\mathbf{I}\\ -\Delta\alpha_2\mathbf{I} & \Delta\alpha_1^*\mathbf{I}\end{bmatrix}\begin{bmatrix}\hat{\mathbf{h}}_3\\ \hat{\mathbf{h}}_4\end{bmatrix}}_{:=\mathbf{e}_{2,i}}\end{aligned} \tag{54}$$

where the subscript i in $\mathbf{e}_{1,i}$ and $\mathbf{e}_{2,i}$ stands for identical precoders, as d stands for distinct precoders later on. The estimation error $\mathbf{e}_{1,i}$ is caused by the mismatch between the asymptotic correlation matrix $\mathbf{R}_{\check{\mathbf{y}}\check{\mathbf{y}}}$ and the finite sample average $\mathbf{R}_{\check{\mathbf{y}}\check{\mathbf{y}}}^{(N)}$. As N increases, $\|\mathbf{e}_{1,i}\|^2$ decreases consistently. The estimation error $\mathbf{e}_{2,i}$ occurs due to the imperfectly resolved scalar ambiguities. Notice that $\mathbf{e}_{2,i}$ is correlated with $\mathbf{e}_{1,i}$ since the scalar ambiguity

determination step employs the noisy estimate ($\hat{\mathbf{h}}_3$, $\hat{\mathbf{h}}_4$). Using (54), and based on the fact $||\hat{\mathbf{h}}_{34}||^2 = 1$, we obtain

$$\begin{aligned} ||\mathbf{e}_{1,i}||^2 &= (|\alpha_1|^2 + |\alpha_2|^2)(|\Delta\mathbf{h}_3|^2 + |\Delta\mathbf{h}_4|^2) \\ ||\mathbf{e}_{2,i}||^2 &= |\Delta\alpha_1|^2 + |\Delta\alpha_2|^2. \end{aligned} \tag{55}$$

For distinct precoders, setting $\alpha = \alpha_1$ and $\alpha_2 = 0$ in (26) yields

$$\begin{bmatrix} \hat{\mathbf{h}}_1 \\ \hat{\mathbf{h}}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{h}_1 \\ \mathbf{h}_2 \end{bmatrix} + \underbrace{\begin{bmatrix} \alpha\mathbf{I} & \mathbf{0} \\ \mathbf{0} & \alpha^*\mathbf{I} \end{bmatrix} \begin{bmatrix} \Delta\mathbf{h}_3 \\ \Delta\mathbf{h}_4 \end{bmatrix}}_{:=\mathbf{e}_{1,d}} + \underbrace{\begin{bmatrix} \Delta\alpha\mathbf{I} & \mathbf{0} \\ \mathbf{0} & \Delta\alpha^*\mathbf{I} \end{bmatrix} \begin{bmatrix} \hat{\mathbf{h}}_3 \\ \hat{\mathbf{h}}_4 \end{bmatrix}}_{:=\mathbf{e}_{2,d}}. \tag{56}$$

Subsequently, we obtain

$$\begin{aligned} ||\mathbf{e}_{1,d}||^2 &= |\alpha|^2(|\Delta\mathbf{h}_3|^2 + |\Delta\mathbf{h}_4|^2) \\ ||\mathbf{e}_{2,d}||^2 &= |\Delta\alpha|^2. \end{aligned} \tag{57}$$

Notice that $\mathbf{h}_{34}$ is an eigenvector with unit norm, whether identical or distinct precoders are used. By comparing (26) with (28), it follows that $|\alpha_1|^2 + |\alpha_2|^2 = |\alpha|^2$. Therefore, we obtain from (55) and (57) that

$$\begin{aligned} E\{||\mathbf{e}_{1,d}||^2\} &= E\{||\mathbf{e}_{1,i}||^2\} \\ E\{||\mathbf{e}_{2,d}||^2\} &= \frac{1}{2}E\{||\mathbf{e}_{2,i}||^2\} = \sigma_\epsilon^2|\alpha|^2(|p_1|^2 + |p_2|^2). \end{aligned} \tag{58}$$

Thus, distinct precoders lead to a 3–dB gain over identical precoders when it comes to suppressing the channel estimation error caused by the imperfectly resolved scalar ambiguities. To achieve the same channel estimation accuracy, identical precoders need to employ twice the number of pilots relative to distinct precoders. Hence, designing distinct precoders instead of identical precoders pays off either in terms of increasing the system efficiency by using half the number of pilots or in terms of improving the system performance with the same number of pilots, which is a feature that we also verified by simulations.

VII. Simulations

To test the proposed channel estimation algorithm, we use as figure of merit the averaged normalized mean square error (NMSE) of the channels defined as: $(1/2)\sum_{\mu=1}^{2}||\hat{\mathbf{h}}_\mu - \mathbf{h}_\mu||^2/||\mathbf{h}_\mu||^2$. We set the system parameters as $L = 8$, $K = 24$, $J = K + L = 32$. The precoder $\mathbf{\Theta}_1$ is formed by the first K columns of a $J \times J$ Walsh–Hadamard matrix $\mathbf{C}$, i.e., $\mathbf{\Theta}_1 = \mathbf{C}(:, 1:K)$, whereas the precoder $\mathbf{\Theta}_2 = \mathbf{\Theta}_1$ if identical precoders are used; otherwise, $\mathbf{\Theta}_2 = \mathbf{C}(:, 2:K+1)$ if distinct precoders are employed. The symbols are drawn from a QPSK constellation. We define the SNR to be the average received bit energy to noise ratio E_b/N_0. Equivalently, E_b/N_0 is the total transmitted energy[3] from two antennas per information bit to noise ratio by setting the total variance for all taps of each channel to unity: $E\{||\mathbf{h}_\mu||^2\} = 1, \forall\mu \in [1, 2]$. We

[3]We ignore the energy contained in the cyclic prefix for OFDM transmissions. Compensation for this energy loss can be easily made.

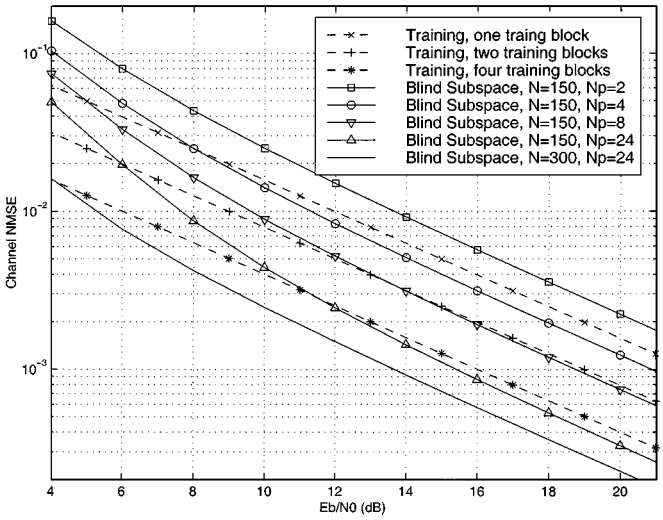

Fig. 2. Training versus blind subspace based channel estimation (static channels).

test the proposed channel estimator for static as well as slowly time-varying channels.

`Test Case 1 (static channels):` The channel taps are randomly generated with equal variance $1/(L+1)$. The simulation results are averaged over 500 channels. We first compare the proposed blind subspace-based channel estimator (with distinct precoders) against the training-based approach of [15]. We collect $N = 150$ symbol blocks to perform the subspace decomposition and use N_p pairs of pilots to resolve the remaining scalar ambiguity. We emphasize here that symbol block in this section refers to $\check{\mathbf{s}}(n)$ in (3), which consists of two OFDM symbols. For the training-based channel estimator, the training blocks are not precoded, and they are optimally designed according to ([15, Eq. (30)] to minimize the channel estimation error. Fig. 2 shows that the blind subspace method outperforms the training method with one training block when $N_p = 4$ and has comparable performance with the training method with two training blocks when $N_p = 8$. When $N_p = 24$, the subspace-based method entails identical redundancy as one block of training but achieves the performance of the training method with four training blocks. The performance of the subspace method improves further as more data blocks become available (see the curve in Fig. 2 with $N = 300$). By exploiting the inherent structure information, the subspace-based method indeed can provide accurate channel estimates.

We next compare the performance of the subspace-based method for distinct precoders and identical precoders. Fig. 3 demonstrates that identical precoders need twice the number of pilots to achieve the same performance as distinct precoders, which is in agreement with our theoretical analysis in Section VI.

`Test Case 2 (slowly time-varying channels):` The slowly time-varying FIR channels are generated according to the channel model A specified by ETSI for HIPERLAN/2 [5], where each tap varies according to Jakes' model with a maximum Doppler frequency of 52 Hz corresponding to a typical terminal speed $v = 3$ m/s and a carrier frequency of 5.2 GHz. We set the subcarrier spacing to 312.5

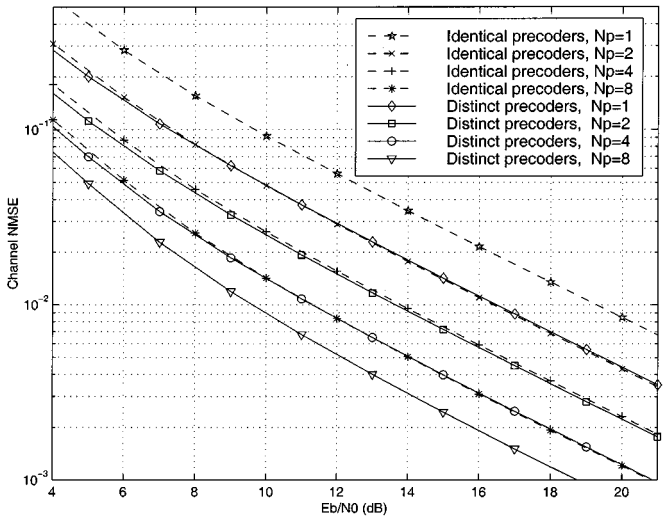

Fig. 3. Identical versus distinct precoders (static channels).

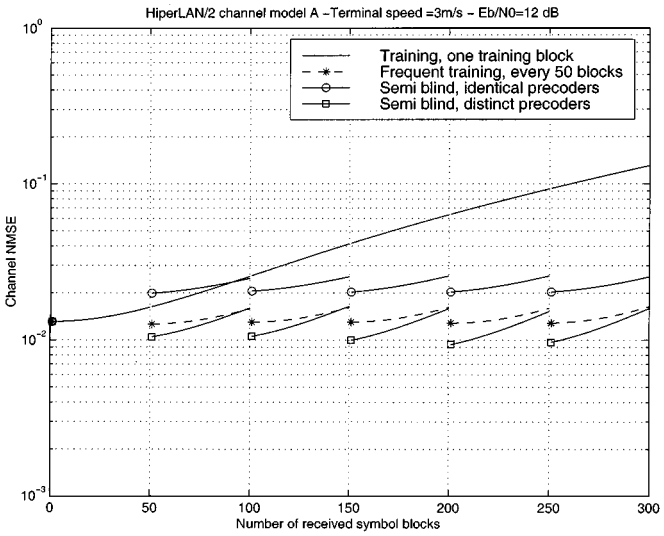

Fig. 4. Channel NMSE along the data burst $N_p = 4$.

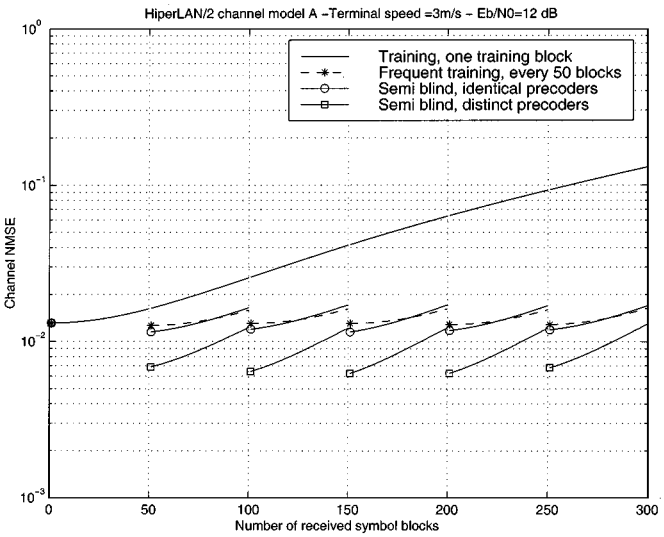

Fig. 5. Channel NMSE along the data burst $N_p = 8$.

kHz, which is the same as the broadband wireless local area network HIPERLAN/2 [6], [24]. Since we have $J = 32$ subcarriers instead of 64 subcarriers, our OFDM symbol duration is set to be half of that of HIPERLAN/2 [24]. We assume here that each data burst (a frame) has $N = 300$ symbol blocks out of which the first block is not precoded and serves as a training block with an optimally designed training sequence. The proposed semi-blind channel estimator is implemented using the rectangular sliding window in (12) with $W = 150$, and is initialized by the training-based method. Except for the first 50 symbol blocks, channel estimation is performed every time 50 new blocks become available. Therefore, only five subspace decompositions are actually performed during each frame transmission. To resolve the residual scalar ambiguities, N_p pairs of pre-precoding pilots are embedded every 50 symbol blocks. Note that the redundancy incurred by the pilots is negligible because we only place $N_p = [1, 2, 4, 8]$ pairs of known symbols in every $50(2K) = 2400$ data symbols.

We average simulation results over 500 Monte-Carlo (MC) trials. Figs. 4 and 5 show the NMSE for the proposed semi-blind channel estimator and the training-based approach of [15] at a typical SNR of $E_b/N_0 = 12$ dB, with $N_p = 4, 8$, respectively. We see that the semi-blind channel estimator tracks closely the slow channel variations, whereas the training-based channel estimates drift away. To equip the training method with channel tracking capability, one may invoke the decision directed (DD) approach. The DD approach updates the channel estimates by treating the previous demodulated data symbols as known training sequence. However, the DD approach is prone to noise propagation, and we have verified that it suffers from catastrophic "run-away" effects in our setup. Thus, we omit the DD approach here. Alternatively, training blocks can be inserted frequently to track channel variations at the expense of rate loss. We compare the proposed channel estimator against frequent retraining, which inserts one optimally designed training block every 50 data blocks. Figs. 4 and 5 show that the semi-blind subspace channel estimator (with distinct precoders) outperforms frequency retraining at SNR $= 12$ dB. Notice that the performance of the proposed semi-blind channel estimator may be further improved at the expense of computational complexity if performed more frequently. We also plot the NMSE averaged over the entire frame as a function of SNR in Fig. 6. We infer from Fig. 6 that the semi-blind subspace channel estimator (with distinct precoders and $N_p = 4, 8$) has comparable (or slightly better) performance as frequent retraining in the modest SNR range but tends to be inferior when the SNR increases to a sufficiently high level. Fig. 6 also illustrates the advantage of distinct over identical precoders in slowly time-varying channels. Compared with Fig. 3, the time-varying nature of the channels leads to an error floor, and the performance improvement by increasing the number of pilots becomes less significant.

To check the overall performance of channel estimation, equalization and ST decoding, we plot, in Fig. 7, the bit error rate (BER) averaged over the entire data burst with ZF equalizers constructed from different channel estimates. The BER curves are averaged over hundreds to thousands of MC trials with at least 100 trials containing errors. Compared with the benchmark BER performance obtained with perfect channel

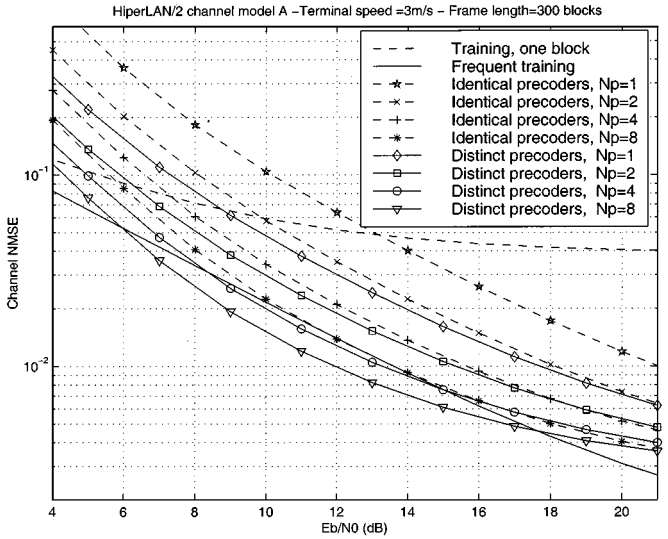

Fig. 6. Channel NMSE versus SNR (slowly time-varying channels).

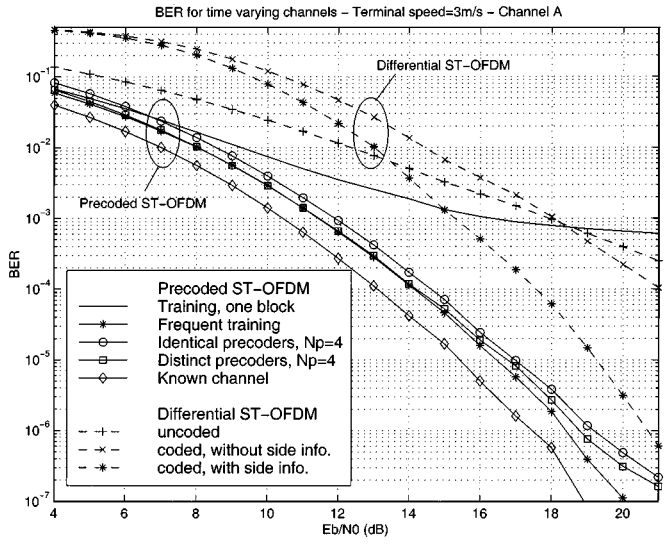

Fig. 7. BER comparisons $v = 3$ m/s $W = 150$ blocks.

knowledge at the receiver, our semi-blind channel estimator with $N_p = 4$ only incurs less than 2 dB SNR loss, and its performance is comparable with that of frequent retraining, which incurs $1/50 = 2\%$ rate loss. A high error floor is observed for the one-block training approach since the channels are time varying, and no tracking mechanism has been invoked.

To illustrate the advantage of channel acquisition and coherent detection at the receiver, we also plot in Fig. 7 the BER performance of the competing differential ST-OFDM alternative, where the differential encoding of [22] is applied on each subcarrier to dispense with channel estimation. To make up for the information rate of our linearly precoded ST-OFDM, convolutional coding with rate $3/4(= K/J)$ is also tested for differential ST-OFDM, where the convolutional encoder is the one used in the HIPERLAN/2 standard [6], [24] with rate $1/2$, punctured to rate $3/4$, and followed by interleaving. Since the differential decoder output takes binary values [22], the Viterbi decoding algorithm with hard decision is applied here. Without assuming any side channel information, the path metric for Viterbi

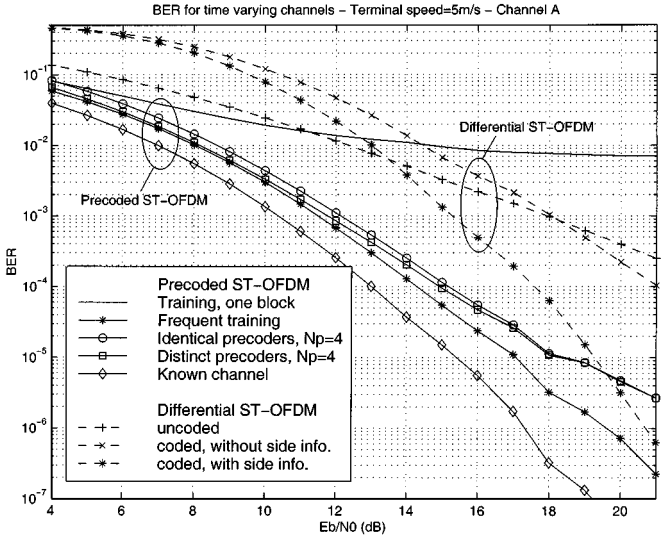

Fig. 8. BER comparisons $v = 5$ m/s, $W = 150$ blocks.

decoding is set to be the Hamming distance between the received bit stream at the output of the differential decoder (which is denoted by $\hat{c}_1, \ldots, \hat{c}_n$, where $\hat{c}_i \in [0, 1]$) and the possible codeword candidates (which are denoted by $c_1, \ldots, c_n$) [27]. Note that the accuracy of the $\hat{c}_i$ estimate varies, depending on the subcarrier on which it rides. If $\hat{c}_i$ is transmitted over the ρ_ith subcarrier, then the reliability of $\hat{c}_i$ depends on the amplitude of the effective fading coefficient on this subcarrier defined by $f_i^2 := |H_1(\rho_i)|^2 + |H_2(\rho_i)|^2$. If side information on the channel fading coefficients $\{f_i^2\}_{i=1}^n$ can be acquired at the receiver (we assume perfect knowledge in this simulation), the path metric could be modified using the weighted Hamming distance $\sum_{i=1}^n f_i^2(\hat{c}_i - c_i)^2$, which reduces to the Hamming distance (up to a scalar) if $f_1^2 = \cdots = f_n^2$. Fig. 7 demonstrates that precoded ST-OFDM equipped with our semi-blind channel estimator outperforms the differential ST-OFDM considerably for both uncoded and coded[4] transmissions at the considered SNR range. Furthermore, we must underscore that the differential encoder in [22] is only applicable to PSK signals, whereas the proposed channel estimator can be applied to arbitrary signal constellations.

Fig. 8 is the counterpart of Fig. 7 but with a higher speed $v = 5$ m/s. We observe that as speed increases, the semi-blind subspace-based method performs worse than the frequent retraining in the moderate-to-high SNR range and is also inferior to differential ST-OFDM at extremely high SNR. Notice that the window size W affects the performance of the subspace-based method. Indeed, if we reduce the window size to $W = 100$, as shown in Figs. 9 and 10, the performance of the semi-blind channel estimator improves at high SNR, whereas it deteriorates at low SNR, revealing the tradeoffs between improved channel tracking accuracy and decreased noise suppression capability. For both speeds $v = 3$ m/s and $v = 5$ m/s, the proposed channel estimator with distinct precoders and window size $W = 100$

[4]Viterbi decoding with soft decision is known to outperform its hard decision counterpart used herein, especially with fading channels. However, soft Viterbi decoding of convolutionally coded ST-OFDM transmissions relying on the differential scheme of [21] is nontrivial. Nevertheless, it is a worthwhile direction for future research, and pertinent results will be reported elsewhere.

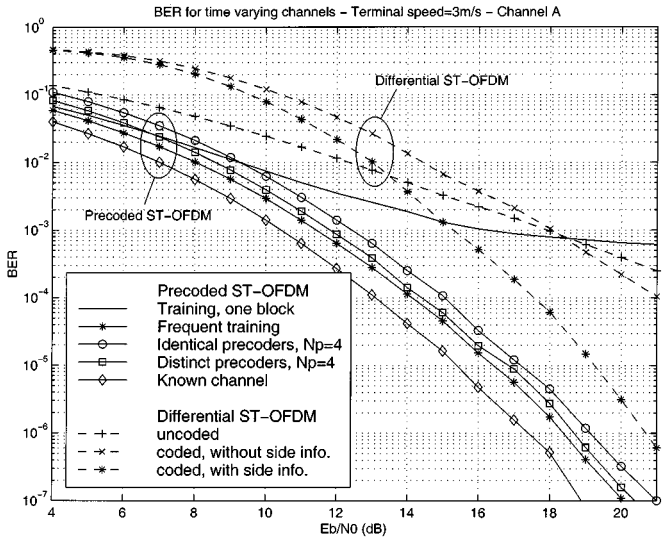

Fig. 9. BER comparisons $v = 3$ m/s, $W = 100$ blocks.

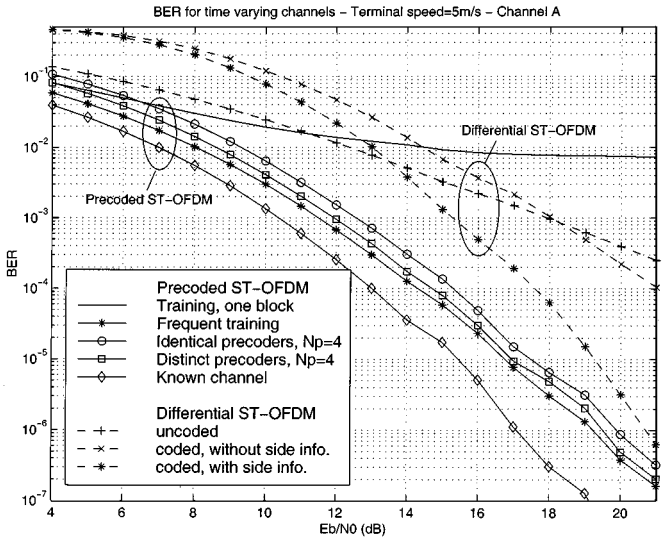

Fig. 10. BER comparisons $v = 5$ m/s, $W = 100$ blocks.

performs slightly worse (about 0.3 dB) than frequent retraining but better than differential ST-OFDM in the considered SNR range.

In a nutshell, the proposed semi-blind channel estimator is capable of tracking slow channel variations. It achieves comparable performance with frequent retraining and exhibits the improved performance of linearly precoded[5] ST-OFDM over differential alternatives through coherent detection. The large gap (especially at high SNR) between the BERs with and without channel tracking highlights the importance of accurate channel tracking in wireless space-time coded transmissions.

VIII. Conclusion

We have developed, in this paper, a subspace-based blind channel estimator for block precoded space-time OFDM transmissions over frequency-selective channels. Regardless of channel zero locations and irrespective of the underlying signal constellations, we established that blind channel identifiability is guaranteed up to one or two scalar ambiguities when distinct or identical precoders are employed for even and odd indexed blocks. Relying on known pilots inserted either before or after precoding, we have demonstrated how to resolve the residual scalar ambiguities. With the same number of pilots, distinct precoders enjoy a 3-dB gain over identical precoders when it comes to suppressing the channel estimation error caused by the imperfectly resolved scalar ambiguities. Simulations have illustrated that the proposed method is capable of tracking slow channel variations in its semi-blind implementation and has comparable performance with a frequent retraining approach. Equipped with the proposed semi-blind channel estimator, linearly precoded ST-OFDM outperforms the differential ST-OFDM without or with convolutional coding.

For space time block-coded multiple access, it has been shown recently that users can be separated deterministically using generalized multicarrier codes [16]. Following the multiuser interference elimination step, the single user (semi-) blind channel estimator proposed herein is directly applicable to multiuser scenarios.

Appendix A
Proof of $\hat{\mathbf{h}}_{56} = e^{j\phi}\hat{\mathbf{h}}_{43}$

Solving (10) in the presence of additive noise amounts to finding the eigenvector corresponding to the smallest eigenvalue of the matrix $\mathbf{Q}\mathbf{Q}^{\mathcal{H}}$. Performing SVD on $\mathbf{Q}\mathbf{Q}^{\mathcal{H}}$ yields $\mathbf{Q}\mathbf{Q}^{\mathcal{H}} = \mathbf{U}\mathbf{\Sigma}_Q\mathbf{U}^{\mathcal{H}}$, where $\mathbf{\Sigma}_Q$ is diagonal with its eigenvalues sorted in decreasing order $\sigma_{2(L+1)} \geq \ldots \geq \sigma_1 \geq 0$.

We next show that $\sigma_2 = \sigma_1$ deterministically, i.e., the smallest eigenvalue has at least multiplicity two, even when $\mathbf{Q}$ is constructed based on finite symbol blocks that are collected in the presence of additive noise. In this case, $\tilde{\mathbf{u}}_k = [\hat{\mathbf{u}}_k^T, \check{\mathbf{u}}_k^T]^T$ is obtained from the SVD of the estimated covariance matrix $\mathbf{R}_{\check{\mathbf{y}}\check{\mathbf{y}}}^{(N)}$ [cf. `Step 1` after (11)]. Let us split the matrix $\mathbf{Q}$ into four $J \times J$ submatrices as

$$\mathbf{Q} = \begin{bmatrix} \mathbf{Q}_1 & \mathbf{Q}_2 \\ \mathbf{Q}_3 & \mathbf{Q}_4 \end{bmatrix} \tag{59}$$

where $\mathbf{Q}_1, \ldots, \mathbf{Q}_4$ can be verified from (8)–(10) to be

$$\mathbf{Q}_1 = \mathbf{V}^T \sum_{k=1}^{2J-2K} \left[\mathbf{D}(\hat{\mathbf{u}}_k^*)\mathbf{\Theta}_1\mathbf{\Theta}_1^{\mathcal{H}}\mathbf{D}(\hat{\mathbf{u}}_k) + \mathbf{D}(\check{\mathbf{u}}_k)\mathbf{\Theta}_2^*\mathbf{\Theta}_2^T\mathbf{D}(\check{\mathbf{u}}_k^*)\right]\mathbf{V}^* \tag{60}$$

$$\mathbf{Q}_2 = \mathbf{V}^T \sum_{k=1}^{2J-2K} \left[\mathbf{D}(\hat{\mathbf{u}}_k^*)\mathbf{\Theta}_1\mathbf{\Theta}_1^{\mathcal{H}}\mathbf{D}(\check{\mathbf{u}}_k) - \mathbf{D}(\check{\mathbf{u}}_k)\mathbf{\Theta}_2^*\mathbf{\Theta}_2^T\mathbf{D}(\hat{\mathbf{u}}_k^*)\right]\mathbf{V} \tag{61}$$

$$\mathbf{Q}_3 = \mathbf{V}^{\mathcal{H}} \sum_{k=1}^{2J-2K} \left[\mathbf{D}(\check{\mathbf{u}}_k^*)\mathbf{\Theta}_1\mathbf{\Theta}_1^{\mathcal{H}}\mathbf{D}(\hat{\mathbf{u}}_k) - \mathbf{D}(\hat{\mathbf{u}}_k)\mathbf{\Theta}_2^*\mathbf{\Theta}_2^T\mathbf{D}(\check{\mathbf{u}}_k^*)\right]\mathbf{V}^* \tag{62}$$

$$\mathbf{Q}_4 = \mathbf{V}^{\mathcal{H}} \sum_{k=1}^{2J-2K} \left[\mathbf{D}(\check{\mathbf{u}}_k^*)\mathbf{\Theta}_1\mathbf{\Theta}_1^{\mathcal{H}}\mathbf{D}(\check{\mathbf{u}}_k) + \mathbf{D}(\hat{\mathbf{u}}_k)\mathbf{\Theta}_2^*\mathbf{\Theta}_2^T\mathbf{D}(\hat{\mathbf{u}}_k^*)\right]\mathbf{V}. \tag{63}$$

[5]Results on the benefits of linear precoding over coding for error control of OFDM transmissions through frequency-selective channels can be found in [26].

With $\mathbf{\Theta}_1 = \mathbf{\Theta}_2$, we obtain from (60)–(63) that $\mathbf{Q}_1^* = \mathbf{Q}_4$ and $\mathbf{Q}_2^* = -\mathbf{Q}_3$. If $\hat{\mathbf{h}}_{34} = [\hat{\mathbf{h}}_3^T, \hat{\mathbf{h}}_4^{\mathcal{H}}]^T$ is the eigenvector of $\mathbf{Q}\mathbf{Q}^{\mathcal{H}}$ corresponding to the eigenvalue σ_1, then so is $\hat{\mathbf{h}}_{43} = [\hat{\mathbf{h}}_4^T, -\hat{\mathbf{h}}_3^{\mathcal{H}}]^T$. Because $\hat{\mathbf{h}}_{34}$ and $\hat{\mathbf{h}}_{43}$ are orthogonal, it follows that the eigenvalue σ_1 has at least multiplicity two. More generally, any eigenvalue of matrix $\mathbf{Q}\mathbf{Q}^{\mathcal{H}}$ has even multiplicity, and hence, $\sigma_{2L+2} = \sigma_{2L+1} \geq \cdots \geq \sigma_2 = \sigma_1$.

We assume here that $\sigma_3 > \sigma_2 = \sigma_1$, which is at least true asymptotically since $\mathbf{Q}\mathbf{Q}^{\mathcal{H}}$ loses rank by two in the noiseless case. Therefore, the subspace spanned by the eigenvectors corresponding to σ_1 has dimension two. From the SVD, we find its orthonormal basis as $\hat{\mathbf{h}}_{34}$ and $\hat{\mathbf{h}}_{56}$. On the other hand, we can form another orthonormal basis as $\hat{\mathbf{h}}_{34}$ and $\hat{\mathbf{h}}_{43}$. Therefore, $\hat{\mathbf{h}}_{56}$ should be proportional to $\hat{\mathbf{h}}_{43}$, and we can write down $\hat{\mathbf{h}}_{56} = e^{j\phi}\hat{\mathbf{h}}_{43}$ because they have the same norm.

APPENDIX B
IDENTIFIABILITY WHEN $J = K + L$ AND $Z = L$

It has been shown for this case that $\mathbf{h}_1 \propto \mathbf{h}_2 \propto \mathbf{h}_3 \propto \mathbf{h}_4$. Let us define a scalar β such that $\mathbf{h}_2 = \beta\mathbf{h}_1$. The matrix $\boldsymbol{\mathcal{H}}$ in (4) is then given by

$$\boldsymbol{\mathcal{H}}(\mathbf{h}_1, \mathbf{h}_2) = \begin{bmatrix} \boldsymbol{\mathcal{D}}_1\mathbf{\Theta}_1 & \boldsymbol{\mathcal{D}}_2\mathbf{\Theta}_2 \\ \boldsymbol{\mathcal{D}}_2^*\mathbf{\Theta}_1 & -\boldsymbol{\mathcal{D}}_1^*\mathbf{\Theta}_2 \end{bmatrix} = \begin{bmatrix} \boldsymbol{\mathcal{D}}_1\mathbf{\Theta}_1 & \beta\boldsymbol{\mathcal{D}}_1\mathbf{\Theta}_2 \\ \beta^*\boldsymbol{\mathcal{D}}_1^*\mathbf{\Theta}_1 & -\boldsymbol{\mathcal{D}}_1^*\mathbf{\Theta}_2 \end{bmatrix}. \tag{64}$$

Consider the $2K \times 2K$ matrix $\bar{\boldsymbol{\mathcal{H}}}(\mathbf{h}_1, \mathbf{h}_2)$ obtained after removing the zero rows of $\boldsymbol{\mathcal{H}}(\mathbf{h}_1, \mathbf{h}_2)$ in (64) corresponding to the zero diagonal entries of $\boldsymbol{\mathcal{D}}_1$; we then obtain

$$\begin{aligned} &\bar{\boldsymbol{\mathcal{H}}}(\mathbf{h}_1, \mathbf{h}_2) = \begin{bmatrix} \bar{\boldsymbol{\mathcal{D}}}_1\bar{\mathbf{\Theta}}_1 & \beta\bar{\boldsymbol{\mathcal{D}}}_1\bar{\mathbf{\Theta}}_2 \\ \beta^*\bar{\boldsymbol{\mathcal{D}}}_1^*\bar{\mathbf{\Theta}}_1 & -\bar{\boldsymbol{\mathcal{D}}}_1^*\bar{\mathbf{\Theta}}_2 \end{bmatrix} \\ &\Rightarrow \bar{\boldsymbol{\mathcal{H}}}(\mathbf{h}_1, \mathbf{h}_2) \begin{bmatrix} \mathbf{0} & \bar{\mathbf{\Theta}}_1^{-1}\bar{\mathbf{\Theta}}_2 \\ -\bar{\mathbf{\Theta}}_2^{-1}\bar{\mathbf{\Theta}}_1 & \mathbf{0} \end{bmatrix} = \bar{\boldsymbol{\mathcal{H}}}(-\beta\mathbf{h}_1, \mathbf{h}_1) \end{aligned} \tag{65}$$

where $\bar{\boldsymbol{\mathcal{D}}}_1$, $\bar{\mathbf{\Theta}}_1$, and $\bar{\mathbf{\Theta}}_2$ are defined accordingly. Hence, the matrices $\bar{\boldsymbol{\mathcal{H}}}(\mathbf{h}_1, \mathbf{h}_2) = \bar{\boldsymbol{\mathcal{H}}}(\mathbf{h}_1, \beta\mathbf{h}_1)$ and $\bar{\boldsymbol{\mathcal{H}}}(-\beta\mathbf{h}_1, \mathbf{h}_1)$ have the same range space, which implies that $\boldsymbol{\mathcal{H}}(\mathbf{h}_1, \beta\mathbf{h}_1)$ and $\boldsymbol{\mathcal{H}}(-\beta\mathbf{h}_1, \mathbf{h}_1)$ have the same range space. Therefore, at least the solution of $\mathbf{h}_3 = -\beta\mathbf{h}_1$, $\mathbf{h}_4 = \mathbf{h}_1$ does not allow for one scalar α to satisfy $(\mathbf{h}_3, \mathbf{h}_4^*) = \alpha(\mathbf{h}_1, \mathbf{h}_2^*)$. Hence, the channels can only be identified up to two scalars in the form $\mathbf{h}_3 = \alpha_1\mathbf{h}_1$ and $\mathbf{h}_4 = \alpha_2\mathbf{h}_2$.

REFERENCES

[1] S. M. Alamouti, "A simple transmit diversity technique for wireless communications," *IEEE J. Select. Areas Commun.*, vol. 16, pp. 1451–1458, Oct. 1998.

[2] H. Bölcskei, R. W. Heath Jr., and A. J. Paulraj, "Blind channel identification and equalization in OFDM-based multi-antenna systems," *IEEE Trans. Signal Processing*, vol. 50, pp. 96–109, Jan. 2002.

[3] J. K. Cavers, "An analysis of pilot symbol assisted modulation for rayleigh fading channels," *IEEE Trans. Veh. Technol.*, vol. 40, pp. 686–693, Nov. 1991.

[4] P. Comon and G. H. Golub, "Tracking a few extreme singular values and vectors in signal processing," *Proc. IEEE*, vol. 78, pp. 1327–1343, Aug. 1990.

[5] *Channel Models for HIPERLAN/2 in Different Indoor Scenarios*, 1998.

[6] *Broadband Radio Access Networks (BRAN); HIPERLAN Type 2; Physical (PHY) Layer*.

[7] G. J. Foschini and M. J. Gans, "On limits of wireless communication in a fading environment when using multiple antennas," *Wireless Pers. Commun.*, vol. 6, no. 3, pp. 311–335, Mar. 1998.

[8] G. B. Giannakis, P. Anghel, and Z. Wang. All-digital unification and equalization of generalized multi-carrier transmissions through frequency-selective uplink channels. *IEEE Trans. Signal Processing* [Online]. Available: http://spincom.ece.umn.edu/ (submitted).

[9] G. B. Giannakis, Y. Hua, P. Stoica, and L. Tong, Eds., *Signal Processing Advances in Wireless and Mobile Communications—Volume I, Trends in Channel Estimation and Equalization*: Prentice-Hall, Oct. 2000.

[10] B. Hassibi and B. Hochwald. (2000) How much training is needed in multiple-antenna wireless links?. *IEEE Trans. Inform. Theory* [Online]. Available: http://mars.bell-labs.com

[11] B. M. Hochwald and T. L. Marzetta, "Unitary space-time modulation for multiple-antenna communications in rayleigh flat fading," *IEEE Trans. Inform. Theory*, vol. 46, pp. 543–564, Mar. 2000.

[12] B. M. Hochwald and W. Sweldens, "Differential unitary space-time modulation," *IEEE Trans. Commun.*, vol. 48, pp. 2041–2052, Dec. 2000.

[13] B. L. Hughes, "Differential space-time modulation," *IEEE Trans. Inform. Theory*, vol. 46, pp. 2567–2578, Nov. 2000.

[14] Y. Li, J. C. Chung, and N. R. Sollenberger, "Transmitter diversity for OFDM systems and its impact on high-rate data wireless networks," *IEEE J. Select. Areas Commun.*, vol. 17, pp. 1233–1243, July 1999.

[15] Y. Li, N. Seshadri, and S. Ariyavisitakul, "Channel estimation for OFDM systems with transmitter diversity in mobile wireless channels," *IEEE J. Select. Areas Commun.*, vol. 17, pp. 461–471, Mar. 1999.

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

[17] Z. Liu, G. B. Giannakis, S. Barbarossa, and A. Scaglione, "Transmit-antennae space-time block coding for generalized OFDM in the presence of unknown multipath," *IEEE J. Select. Areas Commun.*, vol. 19, pp. 1352–1364, July 2001.

[18] Z. Liu, G. B. Giannakis, B. Muquet, and S. Zhou, "Space-time coding for broadband wireless communications," in *Wireless Communications and Mobile Computing*. New York: Wiley, Jan.–Mar. 2001, vol. 1, pp. 33–53.

[19] B. Muquet, M. de Courville, P. Duhamel, and V. Buzenac, "A subspace based blind and semi-blind channel identification method for OFDM systems," *Proc. IEEE-SP Workshop Signal Process. Adv. Wireless Commun.*, pp. 170–173, May 9–12, 1999.

[20] A. F. Naguib, N. Seshadri, and R. Calderbank, "Increasing data rates over wireless channels," *IEEE Signal Processing Mag.*, vol. 17, pp. 76–92, May 2000.

[21] M. Sandell and O. Edfors, "A Comparative Study of Pilot-Based Channel Estimators for Wireless OFDM," Lulea Univ. Technol., Lulea, Sweden, Tech. Rep., Res. Rep. TULEA 1996, 1996.

[22] V. Tarokh and H. Jafarkhani, "A differential detection scheme for transmit diversity," *IEEE J. Select. Areas Commun.*, vol. 18, pp. 1169–1174, July 2000.

[23] *Tech. Spec. Group Radio Access Network*, 1999.

[24] R. D. J. van Nee, G. A. Awater, M. Morikura, H. Takanashi, M. A. Webster, and K. W. Halford, "New high-rate wireless LAN standards," *IEEE Commun. Mag.*, vol. 37, pp. 82–88, Dec. 1999.

[25] Z. Wang and G. B. Giannakis, "Wireless multicarrier communications: Where Fourier meets Shannon," *IEEE Signal Processing Mag.*, vol. 17, pp. 29–48, May 2000.

[26] ——, "Linearly precoded or coded OFDM against wireless channel fades?," *Proc. 3rd IEEE Workshop Signal Process. Adv. Wireless Commun.*, pp. 267–270, Mar. 20–23, 2001.

[27] S. B. Wicker, *Error Control Systems for Digital Communication and Storage*. Englewood Cliffs, NJ: Prentice-Hall, 1995.

[28] B. Yang, "Projection approximation subspace tracking," *IEEE Trans. Signal Processing*, vol. 43, pp. 95–107, Jan. 1995.

[29] L. Zheng and D. Tse, "Information theoretic limits for non coherent multi-antenna communications," in *Proc. Wireless Commun. Netw. Conf.*, Chicago, IL, Sept. 23–28, 2000, pp. 18–22.

Shengli Zhou (S'99) was born in Anhui, China, in 1974. He received the B.S. degree in 1995 and the M.Sc. degree in 1998 from the University of Science and Technology of China (USTC), all in electrical engineering and information science. He is now pursuing the Ph.D. degree with the Department of Electrical and Computer Engineering, University of Minnesota, Minneapolis.

His broad interests lie in the areas of communications and signal processing, including transceiver optimization, blind channel estimation and equalization algorithms, and wireless, multicarrier, space-time coded, and spread-spectrum communication systems.

Bertrand Muquet was born in France in 1973. He received the Ing. degree in electrical engineering from the Ecole Supérieure d'Electricité (Supélec), Gif-sur-Yvette, France, in 1996 and the "Agrégation" degree in applied physics from the Ecole Normale Supérieure de Cachan, Cachan, France, in 1997. He received the Ph.D., degree from Ecole Nationale Supérieure des Télécommunications, Paris, France, in 2001.

From 1998 to 2001, he was a research engineer working on OFDM systems at Motorola Labs, Paris. His area of expertise lies in signal processing and digital communications with emphasis on multicarrier and OFDM systems, blind channel estimation and equalization, and iterative and turbo algorithms. He is now with Stepmind, a company based in France involved wireless communications with interests in GSM, GPRS, EDGE, Hiperlan/2, and IEEE802.11a.

Georgios B. Giannakis (F'97) received the Diploma in electrical engineering from the National Technical University of Athens, Athens, Greece in 1981. He received the M.Sc. degree in electrical engineering in 1983, the M.Sc. degree in mathematics in 1986, and the Ph.D. degree in electrical engineering in 1986, all from the University of Southern California (USC), Los Angeles.

After lecturing for one year at USC, he joined the University of Virginia, Charlottesville, in 1987, where he became a Professor of Electrical Engineering in 1997. Since 1999 he has been a Professor with the Department of Electrical and Computer Engineering, University of Minnesota, Minneapolis, where he now holds an ADC Chair in Wireless Telecommunications. His general interests span the areas of communications and signal processing, estimation and detection theory, time-series analysis, and system identification—subjects on which he has published more than 140 journal papers, 270 conference papers, and two edited books. Current research topics focus on transmitter and receiver diversity techniques for single- and multiuser fading communication channels, precoding and space-time coding for block transmissions, and multicarrier and wideband wireless communication systems.

Dr. Giannakis is the (co-) recipient of four best paper awards from the IEEE Signal Processing (SP) Society (in 1992, 1998, 2000, and 2001). He also received the Society's Technical Achievement Award in 2000. He co-organized three IEEE-SP Workshops and guest (co-) edited four special issues. He has served as Editor in Chief for the IEEE SIGNAL PROCESSING LETTERS, as Associate Editor for the IEEE TRANSACTIONS ON SIGNAL PROCESSING and the IEEE SIGNAL PROCESSING LETTERS, as Secretary of the SP Conference Board, as Member of the SP Publications Board, as Member and Vice-Chair of the Statistical Signal and Array Processing Technical Committee, and as Chair of the SP for Communications Technical Committee. He is a member of the Editorial Board for the PROCEEDINGS OF THE IEEE, and the steering committee of the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS. He is a member of the IEEE Fellows Election Committee, the IEEE-SP Society's Board of Governors, and a frequent consultant for the telecommunications industry.